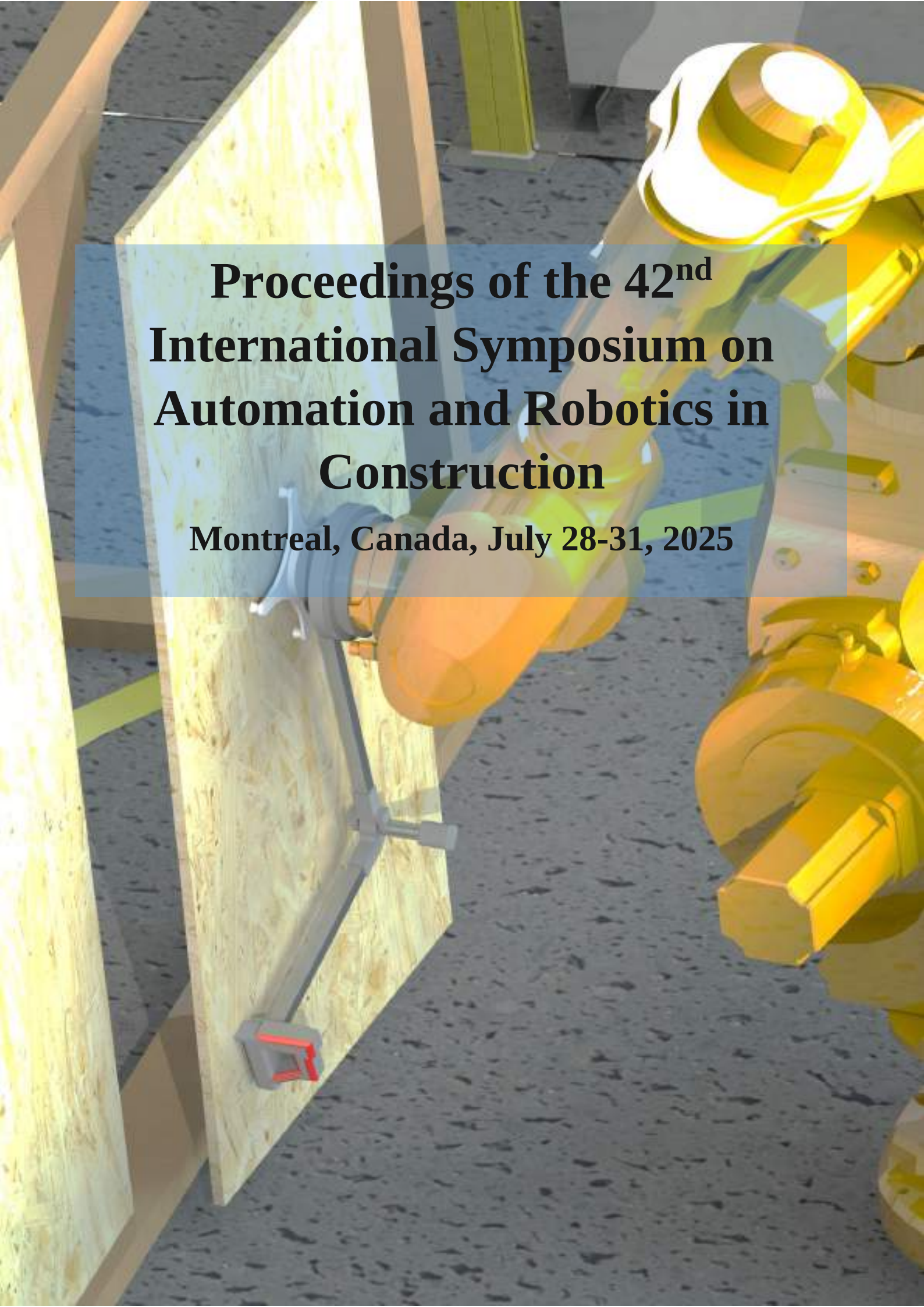


Proceedings of the 42nd International Symposium on Automation and Robotics in Construction

Montreal, Canada, July 28-31, 2025

Proceedings of the 42th International Symposium on Automation and Robotics in Construction

ISSN (for the proceedings series): 2413-5844

ISBN (for this issue of the proceeding series): 978-0-6458322-2-8.

The proceeding series is Scopus indexed.

Scopus®

All papers are available in the IAARC Website and presentations in the IAARC YouTube Channel







Copyrights reserved.

© 2025 International Association on Automation and Robotics in Construction

This work including all its parts is protected by copyright. Any use outside the narrow limits of copyright law without the consent of the individual authors is inadmissible and punishable. This applies in particular to reproductions, translations, microfilming and saving and processing in electronic systems.

The reproduction of common names, trade names, trade names etc. in this work does not justify the assumption that such names are to be regarded as free within the meaning of the trademark and trademark protection legislation and can therefore be used by everyone, even without special identification.

Cover design: Dax Pimentel, Mario Cervantes, Jiansong Zhang, Luis C Félix-Herrán; Image: Patented Robotic End-Effector for Placing and Fastening Paneling Material for Building Construction Operations Developed by the Automation and Intelligent Construction (AutoIC) Lab at Purdue University

Editorial Board

Editors in Chief

Zhang, Jiansong (Purdue University)
Chen, Qian (University of British Columbia)
Lee, Gaang (University of Alberta)
Gonzalez, Vicente A. (University of Alberta)
Kamat, Vineet R. (University of Michigan)

Editors (Area Chairs)

Abdelmegid, Mohammed (University of Leeds)
Al Balkhy, Wassim (CoLeanIT)
Besklubova, Svetlana (University of Cambridge)
Brosque, Cynthia (Stanford University)
Cai, Jiannan (University of Texas at San Antonio)
Chang, Soowon (Purdue University)
Gan, Vincent (National University of Singapore)
Gheisari, Masoud (University of Florida)
Herrera, Rodrigo F. (Pontifical Catholic University of Valparaíso)
Hu, Rongbo (Kajima Technical Research Institute Singapore)
Hu, Yuqing (Pennsylvania State University)
Isaac, Shabtai (Ben-Gurion University of the Negev)
Iturralde, Kepa (University of Stuttgart)
Jebelli, Houtan (University of Illinois Urbana-Champaign)
Kim, Jinwoo (Hanyang University)
Lee, Hoonyong (University of Sydney)
Liang, Ci-Jyun Polar (Stony Brook University)
Lin, Jiarui (Tsinghua University)
Liu, Meiyin (Rutgers University)
Ma, Jong Won (Concordia University)
Mantha, Bharadwaj (University of Sharjah)
Mei, Qipei (University of Alberta)
Mourgues, Claudio (Pontifical Catholic University of Chile)
Ng, Ming Shan (Charmaine) (Kyoto Institute of Technology)
Raphael, Benny (Indian Institute of Technology Madras)
Turkan, Yelda (Oregon State University)
Wu, Lingzi (University of Washington)
Wong Chong, Oscar (Technological University of Panama)
Zeng, Ningshuang (Southeast University)
Zhou, Yang (University of Auckland)
Zhu, Zhenhua (University of Wisconsin-Madison)
Zou, Zhengbo (Columbia University)
Zou, Yang (University of Auckland)

Local Organizing Committee

Amin Hammad, Concordia University (Chair)
Osama Moselhi, Concordia University (Honorary Chair)
Mohamed Al-Hussein, University of Alberta (Chair of MOC)
Mazdak Nik-Bakht, Concordia University (Chair of CSCE/CRC)
Ashutosh Bagchi, Concordia University
Po-Han Chen, Concordia University
Rebecca Dziedzic, Concordia University
Khaled Galal, Concordia University
Michel Guevremont, Hydro-Quebec
Sang Hyeok Han, Concordia University
Shahin Karimidorabati, Concordia University
Zhen Lei, University of New Brunswick
Jong Won Ma, Concordia University
Mohammed Mawlana, Thomas Jefferson University
Ali Motamedi, École de technologie supérieure
Fuzhan Nasiri, Concordia University
Sara Rankohi, Université du Québec à Montréal
Ahmed Soliman, Concordia University
Farid Vahdatikhaki, University of Twente
Cheng Zhang, Xi'an Jiaotong-Liverpool University
Sungkon Moon, Ajou University
Qian Chen, University of British Columbia
Gursans Guven Isin, University of Manitoba
Tadesse Zelele, University of Alberta
Fardin Bahreini, Concordia University
Golnaz Mirzaei, Concordia University
Shrouk Gharib, Concordia University
Azadeh Soofypour, Concordia University
Stéphanie Croteau, Hospitality Concordia
Joanie Boivin, Hospitality Concordia

Sponsors



Reviewers

Abdelmegid, Mohammed
Adil, Muhammad
Agrawal, Ashwin
Agrawal, Ajay Kumar
Akerele, Daniel Damilare
Al Farsi, Ghassan Ibrahim
AlBalkhy, Wassim
Albeaino, Gilles
Alduais, Mohammed
Alotaibi, Emran
Alwisy, Aladdin
Amburi, Sathwik
Anandh, K.S
Ang, Yu Qian
Arana, Maider
Ariyachandra, Mahendrini
Atencio Castillo, Edison
bae, juhyeon
Baek, Francis
Barkokebas, Beda
Bartels, Niels
Bhadaniya, Parth Kamalbhai
Bielski, Jessica (Julia)
Blankenbach, Jörg
Borjigin, Ori
Borrmann, Andre
Bosché, Frédéric
BouHatoum, Makram
Bruttini, Alessandro
Bugalia, Nikhil
Buswell, Richard
Cai, Hongjie
Cai, Jiannan
Cai, Weijia
Carbonari, Alessandro
Chang, Yuntsui
Chen, Po-Han
Chen, Shiqi
Chen, Jingdao
Chen, Zhengxing
Chen, Weiwei
Cheng, Chih-Shen
Cheng, Jiun-Yao
Chern, Wei-Chih
Cheung, Clara
Collins, Fiona
Cooperman, Alissa
Correa, Fabiano
Dan, Minghao
Devkota, Ajit
Ding, Shiqi
Ding, Yujie
Dong, Beixuan
Duan, Kangkang
DUCOULOMBIER, LAURE
Durmus, Dilan
Eiris, Ricardo
Elkabalawy, Moaaz
ElMenshawy, Mohamed Ezzat
Eskudero, Ibon
Esser, Sebastian
Fan, Chao
Fawad, Muhammad
Fleckenstein, Julia
Follini, Camilla
Forcael, Eric
Gao, Yuezhen
GAO, Maggie Y
García de Soto, Borja
Garcia-Lopez, Nelly
Gautam, Yogesh
Gholizadeh Lonbar, Ahmad
Gloser, Franz-Ferdinand
Görsch, Christopher
Graser, Konrad
Grzeszczyk, Grzegorz
Gu, Yueyan
Guo, Jingjing
Habibnezhad, Mahmoud
Hackl, Jürgen
Halder, Srijeet
Ham, Youngjib
Hammad, Amin
Han, Luxuan
Han, Wei
Hartmann, Timo
Hosny, Fatma
Hu, Rongbo
Hu, Yuqing
HU, Difeng
Huang, Lei
Huang, Yaojing
Hwang, Jeongbin
Iqbal, Imtiaz
Jäkel, Jan-Iwo
Jeelani, Idris
Jeon, Jungho
Jeong, Insoo
Jiang, Yishuo
Jipa, Andrei
Jung, Jaehoon
Jung, Wooyoung
Kalasapudi, Vamsi Sai
Kang, Kyubyung
Kastner, Fabian
Katooziani, Ali
Kentroti, Niki
Khoshkenar, Amin
Khoury, Hiam Mayez
Kim, Michael Young
Kim, Namgyun
Kim, Hayoung
Kim, Jeehoon
Kim, Jinwoo
Kim, Junghoon
Kim, Siyeon
Kim, Hongjo
Kim, Hyoungkwan
Kim, Juhyeon
Kim, Yeseul
König, Markus
Kravchenko, Ekaterina
Le, Thai-Hoa
Lee, Kyeongsuk
Lee, Meesung
Lee, Bogyeong
Lee, Dongmin
Lee, Gyueun
Lee, Jisang

Reviewers

Lee, Yongcheol
Li, Hang
Li, Kexin
Li, Shaobo
LI, Zirui
LI, Mingkai
Liang, Ci-Jyun
Liang, Houhao
lin, peng
Lin, Alexander
Lin, Jacob
Lin, Zhiyang
Liu, Canlong
Liu, Kexin
Liu, Liu
Liu, Yan
Liu, Danya
Liu, Yizhi
Liu, Meiyin
Louis, Joseph
Lozano Galant, José Antonio
Ma, Jong Won
Mao, Jiazhen
Mao, Zeyu
Martinez, Jhonattan
Mengiste, Eyob
Messi, Leonardo
Miller, Seirgei Rosario
Moon, Seonghyeon
Mostafa, Kareem
Mourgues, Claudio
Muñoz La Rivera, Felipe
Napps, Daniel
Nassereddine, Hala
Nguyen, Dac Hoang
Núñez Varillas, Christopher Joseph
Obaid, Lubna
Oh, Heung Jin
Ojha, Amit
Orlando, Vincenzo
Osorio Gómez, Cristian Camilo
Pal, Aritra
Pal, Aritra
Pan, Mi
Park, Sangyoon
Park, Somin
Park, Somin
Pauwels, Pieter
Pooladvand, Shiva
Prieto, Samuel A.
Proboste-Martinez, Mathias
Quan, Guan
Raphael, Benny
Raza, Muhammad Huzaifa
Reja, Varun Kumar
Ren, Tianyu
Reynoso, Hamlet
Sabek, Mohamed
Saffert, Anne-Sophie
Saleh, Amira
Sánchez Rivera, Omar Giovanny
Sati, Ala Saif Eldin
Senthilnathan, Shanmugaraj
Shamsollahi, Dena
Shayesteh, Shayan
Shi, Zhuoya
Shirazi, Amirpooya
Siegele, Dietmar
Sohn, John
Soman, Ranjith
Sonkor, Muammer Semih
Soonwook, Kwon
Stumm, Sven
Subrin, Kevin
Suen, Christine
Suh, Seungah
Sun, Yuan
Tan, Tan
Tariq, Muhammad Abu Bakar
tian, fenghua
Vielma, Juan
Walzer, Alexander N.
Wang, Qian
Wang, Xi
Wang, Mingzhu
Wang, Xin
Wei, Chialing
Wen, Jing
Wen, Leyang
Williams-Riquer, Francisco
Xiao, Jinwu
Xiao, Bo
Xu, Jiuyi
Xu, Ping
Xu, Xiaoxiao
Xu, Yifan
Xu, Liqun
Yang, Tao
Yang, Cheng-Hsuan
Yang, Xiaofei
Yantao, YU
YAO, DONGCHI
YE, Yang
Ye, Xuling
Yeum, Chul Min
Yeung, Timson
Yi, Wen
Yong, Gunwoo
Yu, Hongrui
Yu, Jingshuo
Zhang, Yuming
Zhang, Cheng
Zhang, Hao Xuan
Zhang, Haonan
Zhang, Haonan
Zhang, Zhen
Zhao, Shuxuan
Zhao, Xianxiang
Zhao, Xiayu
Zhou, Xiaoshan
Zhou, Shenghua
Zhou, Yuhan
Zhu, Jin
Zhu, Lei
Zhu, Aimin
Zhu, Aiyu
Zhu, Zixian
Zhuang, Dian
Zuo, Hui

Foreword

The International Association for Automation and Robotics in Construction (IAARC) and the Local Organizing Committee are pleased to present the Proceedings of the 42nd International Symposium on Automation and Robotics in Construction (ISARC 2025). This year's symposium draws inspiration from the vibrant spirit and cultural diversity of its host city—Montreal, an international hub of innovation and creativity. ISARC 2025 was held from July 28–31, 2025, proudly hosted by Concordia University.

The 42nd edition of ISARC demonstrated continued growth, with remarkable participation: a total of 272 papers were submitted, marking a 24% increase from the previous year. After a rigorous multi-step peer-review process, including a rebuttal phase, 211 papers (78% acceptance rate) were ultimately selected and included in the proceedings. In total, 1038 authors and co-authors representing 123 universities, research labs, departments, and companies from 27 countries/regions across Asia, Eurasia, the Middle East, North America, Oceania, South America, and Western Europe submitted their work.

The review process was strongly supported by an expanded and diverse group of 32 Area Chairs, representing institutions across various continents, including previously underrepresented regions such as South America and Oceania. Area Chairs actively coordinated the review efforts, providing critical insights across the eight technical areas of interest within IAARC, including "Decarbonizing the Construction Industry," a theme proposed by the Local Organizers. A total of 258 reviewers participated, providing 590 comprehensive reviews. Each submission received at least two reviews, with 77% of papers receiving two reviews, 22% receiving three reviews, and 1% undergoing four reviews.

Building upon the successful innovations introduced by the Technical Committees of previous years, this year’s Technical Committee solidified and integrated these advancements into ISARC’s curation process. The review procedure included pre-screening, two rounds of formal reviews, and a rebuttal phase, which have become fundamental elements within the ISARC review framework. This rigorous process was further structured and rubricated this year by enhanced guidance provided to Area Chairs and reviewers, including streamlined guideline documents, step-by-step cue cards, short instructional videos, systemized regular communication among editorial board members, and a tailored submission system.

The ISARC 2025 symposium featured three keynote presentations, eight plenary sessions, and 200 parallel and poster sessions. We hope you find this year's proceedings engaging and the included papers insightful and inspiring. Enjoy your reading!

Jiansong Zhang
Qian Chen
Gaang Lee
Vicente A. Gonzalez
Vineet R. Kamat

Table of Contents

K - Keynote talks

GPR-Former: Context-Aware Moisture Detection 1

Kevin Lee, Chen Feng

Dynamic Analysis of a Six-Cable Parallel Robot for Automated Panelized Building Retrofits 9

Yifang Liu, Nolan Hayes, Diana Hun, Bryan Maldonado

Semantic-based behavioral decision-making method for construction robots in worker-robot collaboration 17

Xiao Lin, Hongling Guo, Ziyang Guo

P - Plenary talks

Neural Network-Based Estimation of Rebar Deflection for Robotic Rebar Installation of Vertical Structures 26

Hongjie Cai, Rongbo Hu, Ser Tong Quek, Soungho Chae, Justin K.W. Yeoh

Mobility-added multi tool end-effector for picking and nailing applications in robotic timber assembly 34

Mauricio Arredondo-Soto, Paolo Pancho-Ramirez, Rafiq Ahmad

Learning-based 6 DOF Camera Pose Estimation Using BIM-generated Virtual Scene for Facility Management 42

Thai-Hoa Le, Ju-Chi Chang, Wei-Yi Hsu, Tzu-Yang Lin, Ting-wei Chang, Jacob J. Lin

Slate detection in orthophotos of building roof panels 49

Jiajun Li, Frédéric Bosché, Chris Xiaoxuan Lu, Lyn Wilson, Boan Tao

A Large Language Model as an Assistant for Software-Independent Structural Analytical Model Review and Editing 57

Justin S. Lee, Ghang Lee

Roadmap for Exoskeleton Adoption in the Construction Industry: Leveraging Standard Tests and Minimum Criteria 65

Malcolm Dunson-Todd, Mazdak Nik-Bakht, Amin Hammad

Feasibility of an EEG-based dynamic suboptimal cognitive monitoring for field neuroergonomics....73

Eren {M.} Shahrokhi, S. Nizam Ahmed, Gaang Lee, SangHyun Lee

Tracking algorithm for hazardous area in construction sites using PTZ cameras....81

Jaehwan Seong, Hyung-soo Kim, Hyung-Jo Jung

A - Automated/robotic machines, devices, and end-effectors

Dynamic Hand-Signal Recognition and Prediction Framework for Crane Operations Using Deep Learning....89

Zhihao Wei, Ghulam Muhammad Ali, Xinming Li

Adaptive Fractal Quadtree Approach for Efficient Drone-Based Construction Site Safety Inspection....97

Marvin Cheng, Ci-Jyun Liang, Hugo Camargo

Demonstration of a robotic system for remote control of the end-hose during concrete pumping 105

Mathias Haage, Robert Larsson

Adaptive Hazard Zone Model for Improved Safety in Highway Work Zones 113

Ayenew Yihune Demeke, Moein Younesi Heravi, Israt Sharmin Dola, Inbae Jeong, Youjin Jang

Advanced Sensor Integration for Enhanced Flight Control in UAV-Based Construction Automation....122

Tianyu Ren, Xiayu Zhao, Houtan Jebelli

BIM-Driven Robotic Disassembly for Resource Recovery in End-of-Life Buildings 130

Jeehoon Kim, Christopher Rausch

The Development of a Low Code-based Planning Interface System for Coordinated Work of Multiple Construction Machines....136

Sheng Lian, Fumiya Matsushita, Ryosuke Yajima

From BIM to Autonomous Navigation: Using BIM Models to Enable Autonomous Navigation and Localization in Construction Environments 143

Jörg Husemann, Maximilian Kunz, Marcel Suiker, Karsten Berns

Case study: Robotic Underwater Inspection of Concrete Dam Face Prior to Refurbishment 151

Frédéric Nadeau, Marion Nourry, Guillaume Boivin, Eric Boudreault, Carmin Drolet

Robotized tower crane: Simulation and high-level action planning system 155

Mohammad Reza Kolani, Stavros Nousias, André Borrmann

Senser Fusion and Hierarchical Inspection Strategy for Aerial-Ground Multi-robot System 163

Xiayu Zhao, Tianyu Ren, Yizhi Liu, Houtan Jebelli

Vision-Guided Camera Pose Estimation for Robotic Rebar Tying 171

Shaopeng Xu, Huiguang Wang, Xiaoyi Lv, Lu Deng, Bo Jin, Jingjing Guo

Reinforcement Learning with Vision-based State Estimation for Generalized Robotic Assembly.....179

Mark Trovinger, Leonel Giacobbe, Tanner Grantham, Chuangchuang Sun, Jingdao Chen

Enhancing Robotic Vision through Deep Learning Techniques: From Detection to Construction 187

Yusuf Aykin, Hans Sachs, Nikolai Gerzen

Automated Helmet Detection in Construction Sites using UWB and Machine Learning 195

Amir Shahbazi Ojghaz, Sayeh Bayat, Farnaz Sadeghpour

Advanced Vision-guided Robotic Tying with 6-DoF Rebar Pose Estimation 202

Mi Liu, Jingjing Guo, Lu Deng, Xiaoyi Lyu, Linzhen Nie

Adaptive Kinematic and Dynamic Control for Hexapod Robots in Complex Construction Environments 210

Xiayu Zhao, Houtan Jebelli

Drone-Assisted Progress and Quality Tracking for Surface Finishing in Construction.....218

Tianyu Ren, Houtan Jebelli

Path tracking of bucket edge using a hydraulic excavator with a built-in angular velocity controller 226

Daisuke Endo, Genki Yamauchi, Takeshi Hashimoto, Yoshiaki Tsutsumi, Shinjiro Yamamoto, Kenta Furuuchi, Shinya Imura

Interoceptive Robotic Agents for Shared Control and Learning in Collaborative Construction Work.....234

Xiaoshan Zhou, Carol Menassa, Vineet Kamat

Evaluation of an Autonomous Driving Dumper using a Robot Evaluation Framework.....242

Tim Bernhard, Yuan-Jen Huang, Anne Fischer, Cynthia Brosque, Johannes Fottner

LLM-PCPP: A novel robot prioritized coverage path planning approach for complex environments.....250

Zhaofeng Hu, Ci-Jyun Liang

Remote Physical Control for Upgrading Heavy Construction Equipment 258

Meta Soy, Jiaming Fu, Yifu Wu, Jiansong Zhang, Dongming Gan, Jin Wei-Kocsis, Byung-Cheol Min

Robot-to-Human Construction Tool Handover Grasp Prediction for 6-DOF Robotic Arm with Parallel Gripper 264

Yu Lun Kuo, Guan Yong Xiong, Jacob J Lin, Ci Jyun Liang

Automated Decision-Making Tool for Optimal Long-Term Scheduling of MRR Strategies: A Case Study on Bridges 272

Mohammed Alsharqawi, Saleh Abu Dabous, Tarek Zayed

Integration of a real-time orientation measurement system for a real-time evaluator (RTE) to measure the position and orientation of crane-lifted components 280

Balaji Selvakumar, Nolan W Hayes, Yifang Liu, Bryan P Maldonado, Mengjia Tang, Diana Hun

Robot-Assisted 3D Scene Reconstruction Using Fixed FOV LiDAR-Based Depth Camera 288

Vincent J.L. Gan

New collision avoidance algorithm approach based on social sciences 296

Shuntaro Tamura, Masahide Horita

Enhancing Robotic Path Planning in Industrialized Construction: A Novel Benchmarking Approach 304

Amirpooya Shirazi, Aladdin Alwisy

Exploring Extreme 3D Printed Clay Overhangs with Integrated Active Hot Air Drying System on Robotic Arm 312

Mohamed Abdelaziz, Ilija Vukorep, Deena El-Mahdy

Automation of the Pneumatic Caisson Method - Optimization of Excavation Using the Greedy Algorithm - 318

Hiroki Harada, Kiichiro Ishikawa, Taichi Terui, Yusen Inagawa, Kiyoshi Nakamura, Shinichi Tachibana, Shingo Tsugawa, Kotaro Nagahama, Koji Terada

Human-Inspired Inverse Reinforcement Learning for Autonomous Excavator Operations in Lunar Environments 326

Minguk Kim, Youngjib Ham

Safe Reinforcement Learning for Objects Manipulation in Safety-Critical Coordinated Tasks 334

Fawad Khan, Wei Feng, Zhiyong Wang, Tianlun Huang, Liu Xiao, Asad Ali Shahid, Weijun Wang

CoNav Chair: Design of a ROS-based Smart Wheelchair for Shared Control Navigation in the Built Environment 342

Yifan Xu, Qianwei Wang, Jordan Lillie, Vineet Kamat, Carol Menassa

Advancing Automated Building Inspection: A BIM and GSD Integrated Framework for UAV Applications 350

Emadaldeen Benshaaban, Po-Han Chen, Shahinuzzaman

Towards Agent: LLM-based Framework for Robotic Construction by Leveraging IFC 358

Song Du, Yan Gao, Wei Tong, Yiwei Weng

Advances in the Development of a Custom End Effector for Construction Automation 366

Dax Pimentel, Mario Cervantes, Jiansong Zhang, Luis C Félix-Herrán

Exploration of Sustainable Beam Design Using Voronoi Patterns in Concrete 3D Printing 374

Gokul Santhosh S, Benny Raphael, Manu Santhanam

Design and Development of a Remote User Interface for Multi-Robot On-site Component Inspection 381

Juan Carlos Cruz Rivera, Mohsen Navazani, Ibukun Awolusi, Ao Du, Jiannan Cai

SatAOI: Delimitating Area of Interest for Swing-Arm Troweling Robot for Construction 389

Jia-Rui Lin, Shaojie Zhou, Peng Pan, Ruijia Cai, Gang Chen

Real-time 3D Perception System of Construction Equipment for Autonomous Mobile Robots 397

Duho Chung, Seunghun Im, Yohan Kim, Hyoungkwan Kim

Dynamics Model of an Excavator in an Open Source Physical Simulator for Developing Autonomous Construction 405

Taro Abe, Yosuke Matsusaka, Daisuke Endo, Genki Yamauchi, Takeshi Hashimoto

Personalized Emotion-Adaptive Robot Control Strategy for Human-Robot Collaboration in Construction 413

Francis Baek, Leyang Wen, Gunwoo Yong, SangHyun Lee

GPU-Accelerated Collision-Free Path Planning for Multi-Axis Robots in Construction Automation 421

Ilija Vukorep, Rolf Starke, Arastoo Khajehee, Nicolas Rogeau, Yasushi Ikeda

Investigating the Evolving Robot-Building Relationship via a Past-Present-Prospect Model 428

Junjie Chen, Weisheng Lu

Autonomous Construction Robot with Telescopic Arm for Drywall Panel Installation 436

Ayshin Bagherzadeh, Vahid Dolatkhah, Amin Mirfakhar, Omid Valinezhad, Mohammad Javad Haghi

Path planning for LiDAR equipped UGV considering coverage and efficiency in building inspection 444

Yun Cao, Ying Lo, Cheng Zhang

Mixed Reality Based Teleoperation System For Prefabricated Building Assembly 452

Haoran Wang, Shunjie Gu, Zuoqing Yang, Zeren Tao, Qilei Sun

Interface Management of Robots in Construction: A Case Study of Overhead Drilling Robot 460

Zhida Zhang, Mi Pan, Sujuan Zhang, Quan Sun

Scaled-down Simulation and Analysis of On-site Robotic Fabrication and Construction.....468

Raymond Xu, Luka Morita, Parmida Khosravian, Anas Itani

A Proposed Strategy for Automating Intermeshed Steel Connection Assembly using Robotics 477

Miftahur Rahman, Samuel Adebayo, David Hester, Daniel McPolin, Karen Rafferty, Ibukun Awolusi, Debra F. Laefer

Paraphrasing-Based Data Augmentation for Responsible Personnel Classification in Crane Accidents.....485

Deokyeong Kim, Leila Kosseim, Hongjo Kim, Jong Won Ma

Structuring Motion-Based Communication as a Language for Robotics: Analysis with Dense Optical Flow 492

Nicholas Albergo, Jeongbin Hwang, Doyun Lee, Kevin Han

A Validation Study of the Four-Pick Point Design Method of a Mobile Crane Rigging Assembly with Over Six Pick Points 500

Ramy Kheir, Sang Hyeok Han, Ashutosh Bagchi

Evaluation Framework for Indoor Localization Systems in the AEC Environment.....508

Lukas Fritzsche, Rolf Paul Julian Starke, Benjamin Felbrich, Ilija Vukorep

Ultra Low Delay and Resilient Multi-channel Video Encoder for Efficient Remote Construction 516

Yasuyuki Nakajima, Yasukazu Hontama, Kazuhiko Shibairi, Yasuyuki Jitsuta, Daisuke Endo, Genki Yamauchi, Takeshi Hashimoto, Hitoshi Itoh

Multi-dimensional Mapping of Confined Areas using a Hexapod Robot with Integrated Sensor Data and SLAM.....524

Zhong Wang, Qipei Mei, Gaang Lee, Thomas Bock, Vicente A. González

Exploring Cognitive Load and Task Complexity in Dynamic Tracking Tasks: Insights for Construction Workflows.....532

Amira Eltahan, Gaang Lee, Farook Hamzeh

Adjustments and Assembly of a Novel Robotic End-effector for Construction Automation..540

Megan Rudo, Jiansong Zhang

C - Construction management techniques

Tailored Vision-Language Solutions for Comprehensive Hazard Identification on Construction Sites 548

Qihua Chen, Xianfei Yin

Performance Evaluation Model to Enhance Life Cycle Productivity in Modular and Off-Site Construction: A Process-Oriented Approach 556

Babak Manouchehri, Sanghyeok Han, Fuzhan Nasiri

Enhancing Construction Inspection with Augmented Reality: An Analysis of Accuracy and Observation of its Application in Training 564

Jeffrey Kim, Darren Olsen

Fuzzy-Chaos Framework for Analyzing and Quantifying Uncertainty in Labour Productivity in Complex Construction Environments 571

Elyar Pourrahimian, Diana Salhab, Aminah Robinson Fayek, Farook Hamzeh

Video-Based Productivity Monitoring of Worker and Large-Scale Object Interactions in Construction Sites 580

Mik Wanul Khosiin, Jacob J. Lin, Eko Andi Suryo, Kartika Puspa Negara, Ismiarta Aknuranda, Chuin-Shan Chen

Development of a Fall Detection and Safety Communication System Using Small Language Models 588

Malithi Wanniarachchi Kankanamge, Abdur Shahid, Ning Yang, Jamil Uddin, Rahul Biswas

SVR and GA Aided Lean Six Sigma Method for Planning in Modular Construction 595

Angat Bhatia, Osama Moselhi, Sang Hyeok Han

Falling Objects/Debris Detection System Using Surveillance Camera at Construction Site 603

Songyuan Zheng, Meta Soy, Jinwoong Lee, Kyubyung Kang, Jiansong Zhang, Emad Elwakil

Blockchain-based Framework for Securing Construction Worker Certification and Task Authorization 609

Jaewook Jeong, Louis Kumi, Jaemin Jeong, Hyeongjun Mun, Minji Kim, Sangho Yun, Jaehui Son, Minsang Gu, Minwoo Song, Jiwon Hwang

Utilization of Artificial Intelligence in HAZOP Studies and Reports 617

Ehab Elhosary, Osama Moselhi

A Case Study on Applying Machine Learning on Blockchain (MLOB) for Modular Integrated Construction Inspections 625

Zhiming Dong, Weisheng Lu

A Critical Review of Supervised and Reduced-Dataset Vision-Based Construction Activity Recognition Methods....633

Ali Ghelmani, Amin Hammad

Automating Weekly Construction Activity Progress Reporting: Leveraging AI-Driven Workflows....641

Eyob Mengiste, Muammer Semih Sonkor, Zihao Zheng, Samuel A. Prieto Ayllón, Borja García de Soto

How Reliable Are Large Language Models? Zero-Shot Detection of Construction Hazards 649

Nishi Chaudhary, S M Jamil Uddin, Mahzabin Tamanna, Alex Albert, Abdur Rahman Shahid

Implementation of the Cybersecurity Incident Severity Scale (CISS) to Assess Cyber Incidents in the Construction Sector 657

Dongchi Yao, Bharadwaj R. K. Mantha, Borja García de Soto

Development of a Construction-Information System by Measuring, Processing, and Utilizing 3D Data Toward Creating a Digital Twin....665

Satoshi Kubota, Aika Yamaguchi

RAG-Enhanced Safety Information Retrieval for Construction: Integration of Large Language Models with Domain-Specific Information....674

Xianxiang Zhao, Advik Mehta, Falak Sethi, Brian Gue, Qipei Mei, Lingzi Wu

Enhancing Data Management Efficiency in Precast Concrete Manufacturing Through a Custom Database Solution....683

Alaa Abu Nokta, Mohammad Darwish, Mohamed Al-Hussein

D - Decarbonizing the Construction Industry

Enhancing Circular Construction Through Structural Steel Reuse: Case Study of Champlain Bridge....690

Dina Abouhelal, Amin Hammad

Reducing Carbon Emissions in 3D Printed RCC Slabs....698

Abhishek Patel, Benny Raphael

Design for Circular Disassembly of High-Performance Facade Systems....706

Naomi Galina Grigoryan, Alexandros Loutsioli Daskalakis

Enhancing IFC Models for Circular Design: Validation of Required Information through an IDS Framework 714

Lukas Guntermann, Philipp Hagedorn, Markus König

Design Optimization of Steel Structures with Reused Components Using Generative Design 722

Golnaz Mirzaei, Amin Hammad, Faraeen Homayounfard

Advancing Circular Economy in the Built Environment: A Framework for User-Centered Software Tool Development 730

Sai Karthik Vakkanthula, Mazdak Nik-Bakht

Feasibility of Reusing Prefabricated Panels from an Embodied Energy and Economics Perspective 738

Mikael Gilbert, Khalegh Barati, Xuesong Shen

How good are the best practices in carbon reduction? UK highway sector case studies 746

Jinying Xu, Kristen MacAskill

Data Collection Challenges in Waste Diversion Assessment: A Building Demolition Case Study 754

Seungah Suh, Alyssa Roy, Sophia Lee, Sophia Shelton, Christopher Rausch

Evaluating the Carbon Neutrality of Concrete Recycling Using CO2 Capture Technologies from Cement Production 760

Svetlana Besklubova, Ekaterina Kravchenko, Muhammad Huzaifa Raza, Ray Y. Zhong

Harnessing Digital Twins for Construction Site Management 768

Ahmad AL Ayoubi, Ashtarout Ammar

A Comprehensive Framework for Evaluating Blockchain-Based Systems: Embodied Carbon Management in Construction Industry 776

Ka Lin Kuan, Helen Hoi Ling Kwok, Jack Chin Pang Cheng, Kai Lung Hui, Alexis Kai Hon Lau

A Comprehensive Approach for Automated Anomaly Detection and Enhancement of EPC Datasets for Decarbonization 784

Sahar Mirzaie, Zhihao Zhang

Transforming Construction Waste into Resources for 3D Printed Concrete 792

Ekaterina Kravchenko, Muhammad Huzaifa Raza, Svetlana Besklubova, Georgy Lazorenko

Economic Viability of 3D Concrete Printing: A Comparative Study with Traditional Construction Method 797

Muhammad Huzaifa Raza, Svetlana Besklubova, Ekaterina Kravchenko, Ray Y. Zhong

A Regional Investigation of Near-Zero Energy Buildings: Assessing the Impact of PV Integration and Building Design 805

Elsayed Salem, Emad Elwakil, Ashraf Salem

H - Human factors & human-system collaboration

Signals vs. Videos: Advancing Motion Intention Recognition for Human-Robot Collaboration in Construction 813

Charan Gajjala Chenchu, Kinam Kim, Gao Lu, Zia Ud Din

Transforming Construction Robotics Education with Virtual Reality: Analyzing Student Presence and Self-Efficacy 821

Amit Ojha, Yuming Zhang, Houtan Jebelli

A Framework for Real-Time Estimation of Human Activity Intensity in Indoor Environments 829

Moein Younesi Heravi, Youjin Jang, Inbae Jeong, Israt Sharmin Dola

Understanding Cognitive Impacts of Robots in Worker-Robot Collaboration in Modular Construction 837

Yifan Wang, Xiaoyu Hou, Bo Xiao, Shane T. Mueller

Drone-Induced Distraction on Construction Sites: A Quantitative Analysis of Flight Configurations 845

Wei Han, Liqun Xu, Mahfuza Maisha Mouri, Wei-Yin Loh, Fei Dai, Zhenhua Zhu

Cognitive Load-Aware Virtual Reality Training Platform for Safe Drone Operations in Construction 853

Yuming Zhang, Houtan Jebelli

Towards AI-based Optimization of human-centered and robot-assisted construction processes 861

Anne-Sophie Saffert, Jonas Wiederer, Simon K. Hoeng, Thomas Linner, Mathias Obergriesser, Patrick Neumann

Understanding the Impact of an Active Exoskeleton System on Vigilance Level of Workers During Common Repetitive Construction Tasks 869

Amit Ojha, Yogesh Gautam, Houtan Jebelli, Abiola Akanmu

Personalized and Adaptive Virtual Reality Training for Physically Coupled Robots in Construction 877

Yuming Zhang, Amit Ojha, Shayan Shayesteh, Houtan Jebelli

A Digital Twin-based System for the Indoor Environmental Quality Monitoring in Off-Site Construction Facilities 885

Sena Assaf, Mohamed Assaf, Xinming Li, Ahmed Bouferguene, Mohamed Al-Hussein

Development of an assessment matrix for evaluating public participation procedures 893

Jonathan Matthei, Leon Streyl, Philip Balcar, Peter Bohnenkamp, Sven Mackenbach, Katharina Klemt-Albert

Enabling human-robot collaboration in unstructured construction environment through sharing control: Classification, Challenges, and Future Trends 901

Guohao Wang, Changrui Zhu, Honghu Chu, Abdul-majeed Mahamadu, Vijay M. Pawar

Effects of Visual Fidelity on Task Performance and Mental Workload in IVEs 909

Jin-Bin Im, Rong-Lu Hong, Jae-Ho Jang, Enlian Zhang, Seung-Hye Shin, Moonboo Joo, Ki-Won Lee, Kang-Moo Lee, Kyung-Ho Lee, Seo-Young Park, Ju-Hyung Kim

Design considerations of ergonomic human-robot collaboration (EHRC) for modular construction manufacturing 917

Yonglin Fu, Weisheng Lu

I - Information modeling techniques

Unsupervised Multimodal Learning for Fault Detection of Fan Coils Units Using Building Automation System Data.....925

Chin Choy Chai, JJ McArthur

IFC Schema Extension for BIM-based Automated Code Checking of Korean Design Standards for Bridges.....933

Wonbok Lee, Youngsu Yu, Bonsang Koo, Bonghyuck Choi, Seunghwan Lee

A Mixed Reality Dataset and Benchmarking of Deep Learning Models for MEP Classification 941

Boan Tao, Jiajun Li, Frédéric Bosché

Online data acquisition and building modelling for building renovation 949

Danya Liu, Kepa Iturralde, Christoph Holst

Construction Industry Vision Alberta Dataset (CIVAD): Developing a Comprehensive Object Detection Dataset for Diverse Construction Applications.....956

Mohamed Sabek, Qipei Mei, Gaang Lee, Ali Golabchi, Vicente Gonzalez

A Two-Phase AI-Driven Approach to Automated Construction Planning Using Small Language Models for Activity Sequencing and Missing Task Prediction.....964

Akarsth Kumar Singh, Aritra Pal, Shang-Hsien Hsieh

Development of a Petrochemical Plant Pipe Counting Method Based on Improved Mask R-CNN and Transformer 972

Rong-Lu Hong, Jin-Bin Im, Sang-Jun Park, En-Lian Zhang, Nomgon Ochirsuren, Young-Gun Beak, Shin-Hyun Kang, Seong-Won Chung, Cheol-Su Lee, Seunghyeon Wang, Ju-Hyung Kim

Optimizing Structure from Motion Parameters for Volume Estimation of Construction and Demolition Waste 980

Ashwani Jaiswal, Nikhil Bugalia, Quang Phuc Ha

Transformer-based Multi-resolution Fast 3D Reconstruction for Structural Damage Detection 988

Hui Zuo, Tao Sun, Hao Xie, Xiao Ma, Nima Shirzad-Ghaleroudkhani, Qipei Mei

Crack Detection and Segmentation for Bridges Using State-of-the-Art Deep Learning Methods: Single-Stage vs. Two-Stage Detectors..... 996

Ahmed Assad, Mohamad Bo Arki, Miray Sweid, Amin Hammad

Enhancing Building Semantics Preservation in AI Model Training with Large Language Model Encodings 1004

Suhyung Jang, Ghang Lee, Jaekun Lee, Hyunjun Lee

Feature Sensitivity Analysis for Enhanced HVAC Fault Detection and Diagnosis 1012

Arash Hosseini Gourabpasi, Farzad Jalaei, Mazdak Nik-Bakht

Comparing Few-Shot Learning with LLMs for Efficient Text Classification in Road Maintenance Applications..... 1017

Varun Kumar Reja, Ching Yau {Fergus} Mok, Aritra Pal, Ioannis Brilakis

Automated Extraction of Structural Elements and Section Views from Bridge Drawings Using Deep Learning..... 1025

Hakan Bayer, Markus König

Proof-of-Concept Framework of Integrating AI-based LLM Reasoning into BIM Workflows for Design Automation 1033

Jinwu Xiao, Cansu Coskun, Soowon Chang, Kyubyung Kang, Deniz Besiktepe

Non-Productive Time Reduction in Off-Site Construction: A Predictive Analytics Approach Using Deep Learning..... 1041

Xue Chen, Weining Zeng, Runcong Liu, Ahmed Bouferguene, Mohamed Al-Hussein

Interpretable Machine Learning for Predicting the Fatigue Strength of Steel: Influence of Composition and Processing Parameters..... 1049

Soheila Kookalani, Gozde Ozturk, Hamidreza Alavi, Erika Parn, Ioannis Brilakis

Ontology-driven LLM framework for knowledge graph in smart buildings 1057

Sebastien Gerin, Laurent Joblot, Nisrine Makhoul, Frederic Merienne

Integration of Vision Platforms: Generating a Unified 3D Digital Map for Real-time Construction Site Monitoring..... 1065

Jaemin Shin, Philgu Kim, Jeongwoon Choi, Yonghoon Cho, Gitaek Lee, Seokho Chi

Developing a Digital Twin-based Framework for Construction Fire Hazard Recognition Training 1073

Kexin Liu, Mohamed Sabek, Gaang Lee, Max Kinateder, Vicente A. Gonzalez

DSCformer: A Dual-Branch Network Integrating Enhanced Dynamic Snake Convolution and SegFormer for Crack Segmentation 1081

Kaiwei Yu, I-Ming Chen, Jing Wu

A framework for computer vision-based real-time asphalt pavement property monitoring. 1089

Dongying Wang, Qinshuo Shen, Faridaddin Vahdatikhaki, Seirgei Miller, Andre Doree

Vision-based Segmentation, Measurement, and Pose Estimation in Modular Integrated Construction 1097

Yuanyang Qi, Xiao Li, Ruiqi Jiang

Less is Better: A Comparative Analysis of Guidance Methods for Long-Context Window-Based Legal Interpretation 1105

Jaekun Lee, Ghang Lee

Automating WMA Construction Planning: A Physics-aware Framework 1113

Qinshuo Shen, Faridaddin Vahdatikhaki, Seirgei Miller, Andre Doree

Integrating Spatial Proximity and Visual Feature Similarity for Crew Group Detection in Construction Site 1121

Cheng-Yun Tsai, Jacob J. Lin, Ci-Jyun Liang

Automated segmentation and analysis of point clouds of pier foundations using Pier Inspection and Evaluation Report (PIER) 1129

Amine Turki, Nolan W Hayes, Balaji Selvakumar, Bryan Maldonado, Diana Hun

A Synthetic Data Generation Pipeline for Point-Cloud-Based Rebar Segmentation..... 1137

Tao Sun, Yingtong Luo, Yi Shao

Towards the Automation of Last Planner System ®..... 1143

Michael Awe, Avleen Malhi, Nicholas Mavengere, Marcin Budka, Bhargav Dave

EA-YOLOv8: An Enhanced Crack Segmentation Algorithm for Infrastructure Maintenance 1153

Feiyu Chen, Jinghui Qin, Jianyu Yin, Xianfei Yin

Process Coordination and Automation in Modular Construction Supply Chain Using Smart Contract and RPA 1160

Ningshuang Zeng, Xuling Ye, Shiqi Chen, Yan Liu, Markus König, Qiming Li

Modeling Cooling Tower Fan Speed Using Symbolic Regression 1166

Mohamed Kandil, Gary Chang, J.J. McArthur

Developing Computer Vison-based Digital Twin for Vegetation Management Near Power Distribution Networks.....1174

Fardin Bahreini, Mazdak Nik-Bakht, Amin Hammad, Mohamed Gaha

Simulation-based Integrated Transportation Planning in Panelized Construction 1182

Ahmed Zaalouk, Mohammed Sadiq Altaf, SangHyeok Han

Building Component Segmentation-oriented Indoor Localization using BIM-based Synthetic Data Generation and Deep Learning.....1190

Haomiao Zhao, Yifeng Sun, Chi Chiu Lam, Junxiang Zhu, Mi Pan, Mun On Wong

Assessing the Efficacy of Virtual Reality Site Visits to Enhance Remote Learning Experiences for Civil and Construction Engineering Students 1198

Samira Mirhasani, Jared Bosch, Joseph Louis

Efficient Camera Pose Estimation Approach for Infrastructure Inspection 1205

Anas Adeeb Alsharo, Max Midwinter, Chul Min Yeum

Bayesian Networks for data contextualization in Digital Twins of complex civil infrastructures 1213

Leonardo Binni, Anna Brunetti, Fabrizio Gara, Berardo Naticchia

On-site Semantic Mapping and Waypoint Planning for Autonomous Aerial Bridge Monitoring 1221

Yohan Kim, Sunwoong Paik, Hyoungkwan Kim

Development of a Hand-Arm Vibration Syndrome (HAVS) Detection Tool for Construction Workers.....1228

S M Jamil Uddin, Abdur Rahman Shahid, Mohd Farhan Israk Soumik, Ziyu Jin

Deep Learning for Automated Crack Recognition: Insights, Challenges, and Future Directions 1235

Ghodsiyeh Rostami, Po-Han Chen, Mahdi S. Hosseini

Unified Network-Based Representation of BIM Models for Embedding Semantic, Spatial, and Topological Data 1245

Jin Han, Xin-Zheng Lu, Jia-Rui Lin

Vision Transformer-based Local Physical Fatigue Assessment using Electromyography (EMG) Signals for Construction Worker Health Monitoring 1252

Yogesh Gautam, Yuming Zhang, Amit Ojha, Houtan Jebelli

Automated Site Extraction for Scan-to-BIM Progress Monitoring in Road Network Construction.....1260

Ziang Jiang, Xuesong Shen, Khalegh Barati, James Linke

P2P: Point Cloud to Panel Layout Optimization 1268

Nisha Deborah Philips, Yifang Liu, Nolan W. Hayes, Diana Hun, Bryan P. Maldonado

Pixel-Based Rust Severity Classification on Steel Surfaces Using HSI Features and Machine Learning Techniques 1276

Shahinuzzaman, Po-Han Chen, Emadaldeen Benshaaban

The Artificial Intelligence (AI) Impact on Construction Project Management 1284

Kwame Amoah

Integrating Ontology and LLMs for Diagnosis and Repair of Concrete Surface Defects.... 1292

Fardin Bahreini, Amin Hammad

3D Full-body Pose Estimation of Construction Machines Using Deep Neural Network and Stereo Vision 1300

Han Luo, Mingzhu Wang, Peter Kok-Yiu Wong, Pak Him Leung, Jingyuan Tang, Jack C.P. Cheng

Deployment for Additive Construction through Digital Platforms: Gananoque Project as a Case Study 1308

Pablo Salvador Banda Pérez, Ioulios Georgiou, Patricio Emanuel Carrasco Pérez, Romanella Soria, Zayed Elbadri, Alejandro Gonzalez, Daniel Cantor, Nhung Nguyen

Indoor Comfort Assessment Based on Digital Twin Platform 1316

Chia Ying Lin, I-Chen Wu

Application of BIM Technology to Soil Layer Risk Analysis for Enhanced Accuracy and Informed Decision-Making in Geotechnical Engineering 1324

Wen-Yi Wang, I-Chen Wu

Large Language Models in Construction Bidding: Opportunities and Risks of Algorithmic Collusion 1332

Chan Heo, Moonseo Park, Changbum Ryan Ahn

Development of Auto Route System for MEP (Mechanical, Electrical and Plumbing) Pipe Design in Construction 1340

Youngseo Hwang, Kenneth Sungho Park, Sungkon Moon

Automated Generation of Inspection Reports for Construction Operations 1347

Hassan Bardareh, Osama Moselhi

Automated assembly sequence planning and scheduling in precast building projects using reinforcement learning 1355

Ajay Kumar Agrawal, Yang Zou, Hongyu Jin, Mohammed Abdelmegid, Vicente Gonzalez

Analysis and Simulation of Pile Driving Construction Equipment 1363

Laszlo Kovacs, Eric Karpman, Marek Teichmann, Jozsef Kovecses

Framework for Acoustic Comfort Analysis in Digital Twins 1371

Hossein Pouresmaeeliasiabar, Ali Motamedi

Towards Remote Construction Site Analysis: From Laser Point Clouds to Virtual Reality Applications 1379

Trevor Elliott Neece, Alessandro Fascetti

Computer Vision based Automated Timber Structural Defect Detection Framework 1387

Afia Rasool, Qipei Mei, Rafiq Ahmad

Comparative Study of Structure from Motion on Construction Site 1395

Mingyun Kang, Sangmin Lee, Sebeen Yoon, Taehoon Kim

Automatic Reconstruction of Aligned Indoor Building and Façade Elements from Point Cloud Data 1401

Mudan Wang, Yuandong Pan, Anson, T.K. Chan, Sven Auerswald, Seyedehmaedeh Aghili, Ioannis Brilakis

Microscale digital twinning of rubble stone masonry walls using robotic assembly 1408

Qianqing Wang, Jingwen Wang, Stefana Parascho, Katrin Beyer

Integrating BIM and Photogrammetry for Digital Twin Development: A Case Study in Road Infrastructure 1414

Khalil Ahmed Bin Mushtaq, Uzair Manzoor, Muhammad Usman Hassan, Muhammad Fawad, Qian Chen, Muhammad Farhan Jahangir

Are Open-Vocabulary Models Ready for Detection of MEP Elements on Construction Sites 1421

Abdalwhab Bakheet Mohamed Abdalwhab, Ali Imran, Sina Heydarian, Ivanka Iordanova, David St-Onge

Deep Point Cloud Building Envelope Segmentation (DeeP-CuBES) using Deep Learning 1425

Balaji Selvakumar, Yifang Liu, Nolan W. Hayes, Diana Hun, Bryan P. Maldonado

MAP: A Versatile Platform for Predictive Algorithms in Construction and Beyond 1433

Ramyani Sengupta, Emad Elwakil, Yi Jiang

S - Sensing systems & data infrastructures

BIM-integrated Path Planning for Semi-autonomous Robots: Remote Real-time Ambient Data Collection in Indoor Spaces 1439

Fahad Iqbal, Shayan Mirzabeigi

Automated Image Dataset Generation Using Web Scraping and Large Multimodal Models for Construction Applications 1447

Shrouk Gharib, Osama Moselhi

Next-Generation Stretchable EDA Sensors with Ag-NP Fractals for Motion Artifact-Free Physiological Sensing in Dynamic Environments 1455

Yogesh Gautam, Houtan Jebelli

Pole-Shaped Object Extraction from LiDAR Point Clouds of Roads 1463

Danesh Shokri, Shirin Malihi, Saeid Homayouni, Christian Larouche, Heidar Rastiveis, Ioannis Brilakis

Real-time Multimodal Sensing System for Additive Construction by Extrusion: Integrating thermal, depth and RGB data 1471

Yilin Cai, Julie Ann Hartell, Ashrant Aryal

Integrated Framework for Scaffold Monitoring: From Point Cloud Acquisition to Anomaly Detection 1479

Juhyeon Kim, Duho Chung, Jeehoon Kim, Hyoungkwan Kim

Evaluation of Flight Parameters in UAV-Based 3D Reconstruction for Rooftop Infrastructure Assessment 1486

Nicholas Chodura, Melissa Greeff, Joshua Woods

Algorithm development for automatic detection of progressive damage in tunnel cross-sectional geometry 1494

Christopher Joseph Nuñez Varillas, Marck Steewar Regalado Espinoza

3D Measurement System for Soil Loading by an Autonomous Backhoe using OPERA 1501

Takayoshi Hachijo, Yutaro Fukase, Takashi Yokoshima, Yuki Miyashita, Shunsuke Kimura, Masanori Suzuki, Yuichiro Kasahara, Tomoya Kouno, Koshi Shibata, Ryo Kurazume, Daisuke Endo, Genki Yamauchi, Takeshi Hashimoto

One-Shot Indoor Positioning Using 360-Degree Photos 1507

Daeyoung Gil, Ghang Lee

Human Pose Estimation Using Automated Multi-Camera Calibration 1514

Israt Sharmin Dola, Inbae Jeong, Youjin Jang, Moein Younesi Heravi

LiLoc – Defect Image Localization in Laser Scans 1522

Patrick Herbers, Lisa von Rössing, König Markus

A Robust Pipeline for Facade Geometry Estimation Using AprilTags and Digital Imaging 1530

Aravind Srinivasaragavan, Danya Liu, Kepa Iturralde, Christoph Holst

Automated Inspection Methodology of Wall Rebar Spacing and Diameters Using Laser Scanning 1536

Philgu Kim, Jungyeon Kim, Jeongwoon Choi, Seokho Chi

Automated defect detection in clay printing 1544

Patricia Peralta, Thamer Al-Zuriqat, Mahmoud Noufal, Kay Smarsly

Improved Information Extraction from Bridge Inspection Reports using Fine-tuned Generative Pre-trained Transformers 1551

Abdelhady Omar, Osama Moselhi

Practical Considerations and Case Study for Implementing Autonomous UAV Video Collection of Construction Activities 1559

Yusheng Huang, Amin Hammad

SB - Services and business applications / Industry & Short Papers

Drywall finishing with collaborative robot arm in off-site construction 1567

Philippe D'Amours, Samuel Faucher, François Ferland, Alexandre Girard

BIM-FM interoperability: integrating existing FM platform with visualization of IFC models 1571

Andressa Oliveira, José Granja, Pedro Machado, Ali Motamedi, Miguel Azenha

T - Technology management and innovation

Efficient RAG(Retrieval-Augmented Generation) for Construction in Low-Resource Language 1575

Kichang Choi, Seungwon Baek, Jongwon Ma, Hongjo Kim

Assessment Framework for 3D Concrete Printing Technology in Alignment with Construction 5.0 Criteria 1582

Imtiaz Iqbal, Svetlana Besklubova, Tala Kasim

Mixed Reality-based Digital Twinning of Building Circularity: A Co-Design Approach for Sustainable Buildings 1590

Bo Su, Muhammad Fawad, Qian Chen, Marek Salamak

Knowledge-Based Systems in the Era of Large Language Models: A Case Study in Fire Safety Management 1597

Dilan Durmus, Shabtai Isaac, Alessandro Carbonari, Alberto Giretti

Proposal of a Benchmark Indicator for Automated Construction Technology 1605

Genki Yamauchi, Taro Abe, Daisuke Endo, Takeshi Hashimoto, Keiji Nagatani

BIM-Based Public Participation: Insights and Lessons Learned form a Case Study 1609

Jonathan Matthei, Maximilian Friedhelm Heinrich Christ, Sven Mackenbach, Katharina Klemt-Albert

Web Based Security Assessment of Open Source CAD tools using ZED Attack Proxy (ZAP) scanner 1617

Abdul Shakir, Bharadwaj R. K. Mantha

GPR-Former: Context-Aware Moisture Detection

Kevin Lee[1] **and Chen Feng**[1,2]

[1]Building Diagnostic Robotics, Inc., USA
[2]Tandon School of Engineering, New York University, USA

k.lee@nyu.edu, cfeng@nyu.edu

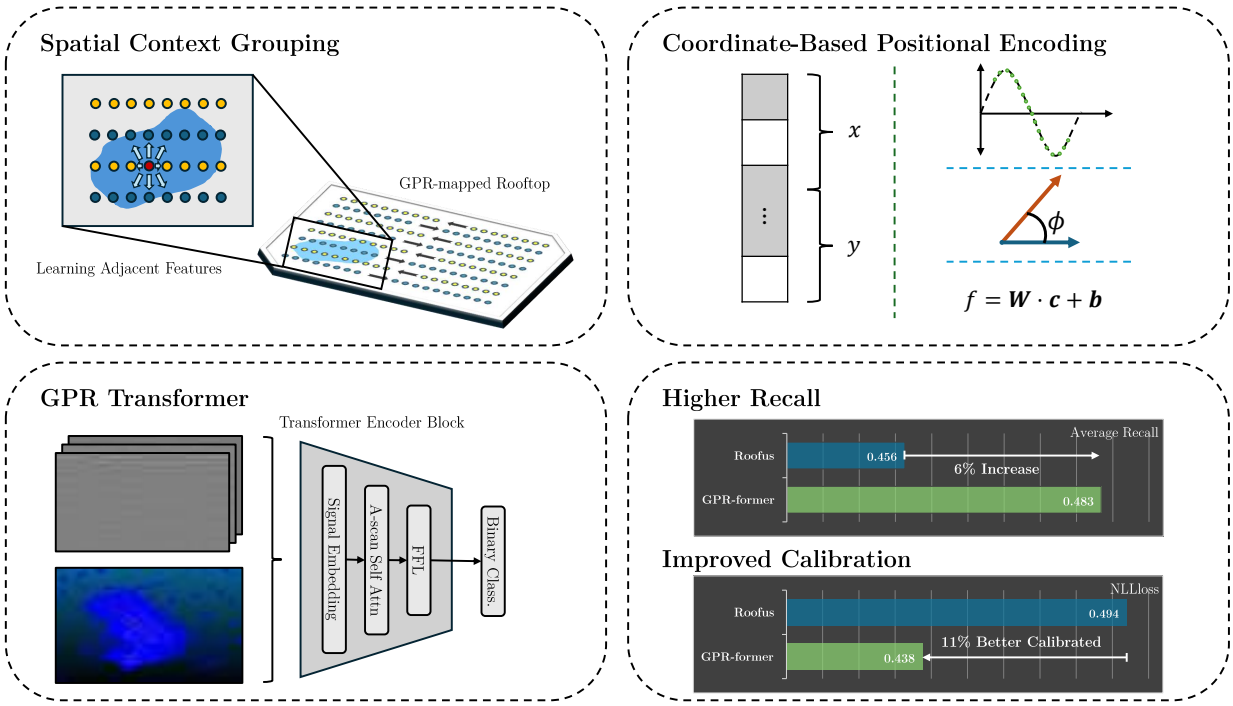


Figure 1. **Overview of GPR-former.** We take inspiration from the complex patterns in which moisture propagates and design a 2D coordinate-based spatial context grouping strategy to more effectively detect the presence of moisture damage on building rooftops. Compared to our previous method, GPR-former demonstrates superior performance in key metrics such as Average Recall as well as having better calibrated outputs.

Abstract -

Moisture damage in roofing systems poses a significant challenge to building operation and maintenance, with far-reaching implications for energy efficiency, structural integrity, health, safety, and sustainability. This study introduces *GPR-former*, a novel transformer-based architecture designed to leverage spatial context in ground-penetrating radar (GPR) data, enhancing the accuracy of moisture detection. Unlike our previous approach that analyzes isolated B-scans, GPR-former incorporates spatial groupings of A-scans with coordinate-based positional encoding, enabling the model to detect complex, nonlinear moisture patterns. Comprehensive experiments across diverse roofing materials demonstrate that GPR-former outperforms our previous method by up to 6% across performance metrics and achieves improved calibrated results. These findings highlight the potential of GPR-former as an improved, transformative tool for sustainable building maintenance, contributing to embodied carbon reduction and extending roof lifecycle management.

Keywords -

Moisture detection, ground-penetrating radar, transformers, spatial context, sustainability

1 Introduction

Moisture damage in roofing systems is a critical issue that impacts energy efficiency, structural safety, and repair costs. It also undermines sustainability efforts by increasing the embodied carbon footprint of buildings through both premature and overdue replacements. Early and ac-

curate detection of moisture damage is essential to mitigate these impacts, support lifecycle management, and enable circular economy initiatives in building renovation and maintenance.

Using GPR to detect moisture in roofing materials presents unique challenges:

- **Signal Obfuscation:** GPR signals are inherently affected by environmental noise, material heterogeneity, and structural interference, making it difficult to identify moisture-related patterns with high confidence.
- **Complex Moisture Patterns:** Moisture often propagates nonlinearly, driven by factors like capillary action and material saturation, complicating interpretation when relying on isolated scans.

Recent advancements in deep learning for GPR [1, 2, 3], particularly the use of transformers and attention mechanisms for B-scan analysis [4, 5, 6], have demonstrated the potential of neural networks in GPR data interpretation. Moreover, progress in the development of mobile robotic systems for GPR data acquisition and processing [4, 7] motivates the development and integration of these technologies. However, methods like our previous approach [4] typically operate within localized linear windows, limiting their ability to leverage spatial relationships critical for identifying moisture patterns across adjacent scans. This gap underscores the need for models that integrate spatial context and enable holistic analysis of GPR data.

To address these challenges, we propose **GPR-former**, a transformer-based architecture that introduces the following key contributions:

- **Spatial Context Grouping:** A novel grouping mechanism that aggregates A-scans based on spatial proximity, providing a richer contextual understanding of moisture propagation.
- **Coordinate-Based Positional Encoding:** A mechanism to encode spatial coordinates directly into the model, allowing GPR-former to effectively map relationships between grouped scans.
- **Comprehensive Evaluation:** Extensive experiments across diverse roofing materials validate the efficacy of the proposed approach.

2 Related Work

2.1 GPR-Based Moisture Detection

Ground-penetrating radar has been widely applied in various domains for moisture detection. For example, [8] employs a drone-mounted GPR sensor to map soil moisture across agricultural fields, demonstrating the versatility of GPR in remote sensing applications. Similarly, [9] integrates GPR with thermal imaging to detect moisture in building walls, highlighting the potential of GPR-based moisture detection specifically in the domain of building diagnostics and maintenance.

In the context of roofing systems, our previous work [4, 10] establishes a foundation for GPR-based moisture analysis. [10] explores self-supervised learning to extract semantically meaningful features from GPR A-scans, enabling downstream tasks like moisture detection with minimal reliance on labeled data. Our investigation indicates the possibility of accurately identifying and isolating GPR scans taken over moisture saturated surfaces, but fails to produce a scalable and robust classifier. Building on this, Roofus [4] applies supervised learning as well as the transformer architecture to classify moisture in real-world rooftop environments using a large-scale labeled dataset of GPR B-scans. Although effective, Roofus focuses on localized analysis within sample B-scans, failing to account for broader spatial relationships that may be critical for detecting moisture propagation across B-scans.

2.2 Transformers in Coordinate-Based Applications

Transformers [11] have gained prominence in tasks requiring spatial reasoning, with positional encoding enabling the modeling of spatial relationships in structured data. In medical imaging, for instance, [12] employs coordinate-based embeddings to enhance transformer-based models for spatially aware tasks. Similarly, CoordFormer [13] introduces a Coordinate-Aware Attention module to encode spatial-temporal coordinates, preserving dependencies critical for analyzing video and image datasets. Li et al. [14] investigates the efficacy of explicit positional encoding by concatenating 2D positional coordinates to the end of each token's embedding dimension among other strategies. Despite these advancements, the application of transformers to specifically GPR data remains underexplored, leaving untapped potential for leveraging spatial relationships in GPR-based analysis.

2.3 Limitations of Existing Approaches

Existing GPR-based methods, including Roofus [4], primarily analyze localized B-scans and overlook the spatial propagation of moisture across adjacent scans. This limitation often results in the model struggling when encountering strongly reflective objects like horizontal rebar. Similarly, while transformers with positional encoding have been effectively applied to coordinate-based tasks, their use in GPR data analysis lacks exploration. Addressing these gaps requires a model capable of capturing the unique spatial relationships inherent in GPR datasets to identify complex moisture diffusion patterns in real-world

rooftop environments.

3 Model Overview

3.1 Task Definition

The primary goal is to classify Ground Penetrating Radar (GPR) A-scans into binary categories indicating the presence or absence of moisture. Formally, given a collection of A-scans grouped by spatial proximity, denoted as $\{S_1, S_2, \ldots, S_\omega\}$, where each scan $S_i \in \mathbb{R}^{655}$ represents a feature vector derived from radar signals, the task is to output a binary prediction $P_i \in \{0, 1\}$ for each scan S_i. Here, $P_i = 1$ signifies the presence of moisture, while $P_i = 0$ indicates its absence.

Binary classification is particularly suited for the application domain, where client-facing diagnostic reports prioritize actionable outcomes. While additional insights such as moisture depth or severity could enrich the analysis, this study focuses on binary classification as a practical first step. The detection of moisture ingress typically necessitates remediation regardless of severity, aligning with the goal of actionable and interpretable results.

Unlike previous approaches that process A-scans or B-scans in isolation, GPR-former leverages spatial context by analyzing grouped A-scans across multiple B-scans. This strategy enhances sensitivity to obfuscated or diffuse moisture signals, which are often missed by single-scan models like Roofus.

3.2 Spatial Context Grouping

Intuition. Moisture in roofing materials seldom manifests as isolated points; instead, it propagates radially or irregularly due to dynamics like capillary action and absorption as illustrated in Figures 1 and 2. These patterns often form clusters or saturate adjacent regions. To capture such propagation behaviors, GPR-former organizes A-scans into spatially coherent groups using x, y coordinates obtained from LiDAR data, enabling the model to recognize subtle or diffuse moisture patterns that elude traditional methods.

Group Sampling Procedure. A straightforward K-D Tree-based algorithm is used to group A-scans according to spatial proximity. We use the K-D Tree for its simplicity and performance. The procedure is outlined as follows:

1. Construct a K-D Tree using the x, y coordinates of all A-scans from a rooftop dataset.
2. Arbitrarily select an A-scan as the seed for a group and identify its $\omega - 1$ nearest neighbors.
3. Form a group of ω A-scans and remove them from the K-D Tree.
4. Repeat steps 2–3 until the K-D Tree is empty. Discard any group with fewer than ω A-scans.
5. Ensure that all groups are formed strictly from A-scans belonging to the same rooftop to prevent erroneous mixtures of scans from different sites.

Algorithm 1: Generate Groups using K-D Tree

Input: GPR Data D, group size ω
Output: Groups G

Initialize D copy: $D_0 \leftarrow D$;
Initialize empty groups: $G \leftarrow \emptyset$;
while *size*$(D_0) \geq \omega$ **do**
  Construct K-D Tree tree $\leftarrow$ `KDTree`(D_0);
  Select the first sample as seed: seed $\leftarrow D_0[0]$;
  Query nearest neighbors:
    neighbors $\leftarrow$ `query`(tree, seed, $k = \omega - 1$);
  Form group: $G \leftarrow G \cup \{\text{seed}, \text{neighbors}\}$;
  Remove group from D_0:
    $D_0 \leftarrow$ `delete`$(D_0, \{\text{seed}, \text{neighbors}\})$;
end
return G

This procedure ensures spatial coherence by grouping A-scans strictly from the same rooftop. Groups with fewer than ω A-scans are discarded to maintain consistency. Although this procedure may not be the most computationally efficient, it is sufficiently effective for handling our large rooftop datasets in practice.

An important hyperparameter to consider is the group size ω. An ω that is too small inhibits the algorithm's capability to span the gaps between B-scans, nullifying the benefits of inter-scan context that we propose. If ω is too large, undesirable artifacts like spatial discontinuity in group membership may occur. Figure 3 provides an illustration describing such phenomenons. In our study, we explore the effects that these situations can have on performance and we investigate the optimal ω in our ablation.

3.3 Coordinate-Based Positional Encoding Strategies

Overview. Positional encodings are essential for embedding spatial relationships among grouped A-scans, enabling the model to leverage spatial context effectively. To achieve this, we explore three different strategies for encoding the positional information into each A-scan representation:

- **Sinusoidal Absolute Encoding:** Employs fixed periodic functions to encode absolute signal feature positions.
- **Rotary Encoding:** Encodes relative positional relationships within groups by applying rotational transformations to the embeddings.
- **Learned Encoding:** Learns position embeddings directly from the data via a linear layer.

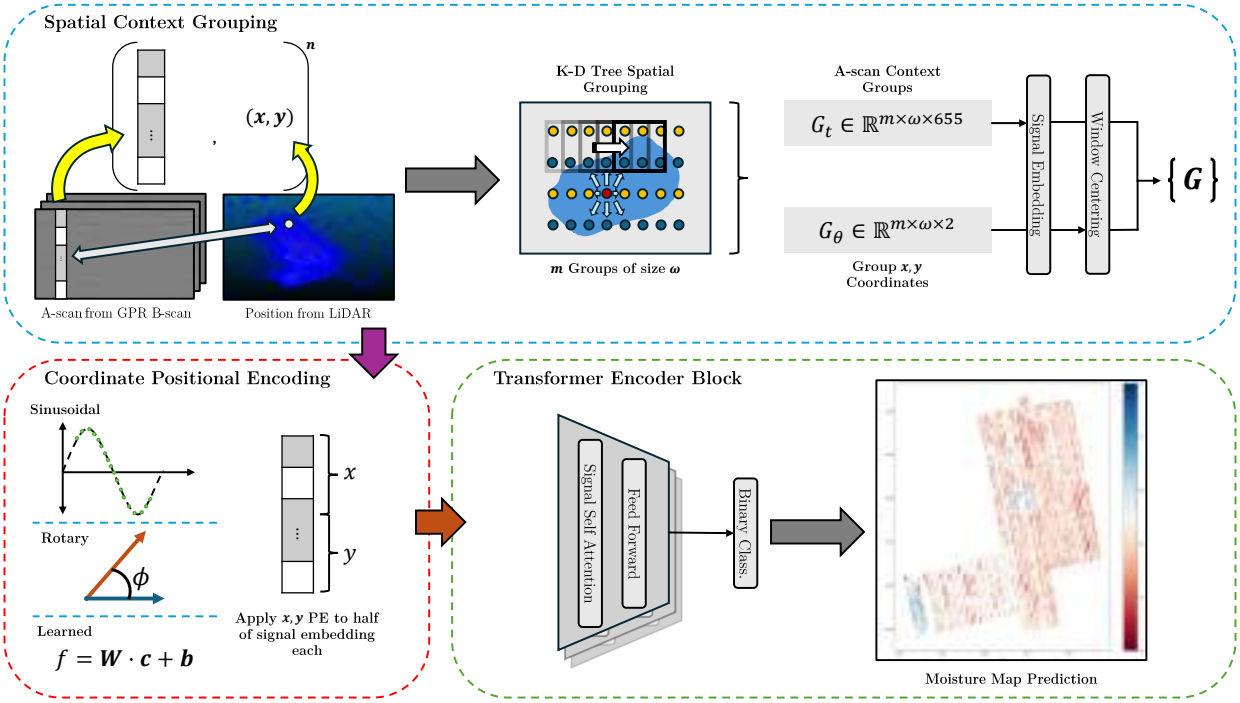


Figure 2. **GPR-former aggregates A-scans into groups of size ω using a K-D Tree sampling procedure.** These grouped scans are processed by a transformer encoder, which incorporates coordinate-based positional encodings to capture spatial relationships and contextual dependencies.

Implementation Details. To account for variability in rooftop layouts and scan alignments, the x, y coordinates are normalized relative to the group's mean position before encoding. This normalization ensures that encodings are invariant to the absolute coordinate system, allowing the model to generalize across diverse rooftop environments.

Each encoding method maps the normalized x, y coordinates to feature embeddings, with half of the embedding dimension allocated to each axis for the Sinusoidal Absolute Encoding and Rotary Encoding. The specific formulations are detailed below:

Sinusoidal Absolute Encoding. We borrow the encoding scheme used in the implementation of the Masked Autoencoder [15], where half of the feature embeddings are dedicated to the x position and the remainder to the y. This particular approach encodes x and y positions using sine and cosine functions of varying frequencies, defined as:

$$\begin{aligned} PE_x[i] &= \sin(x \cdot \beta[i]), \quad PE_x[i+1] = \cos(x \cdot \beta[i]), \\ PE_y[i] &= \sin(y \cdot \beta[i]), \quad PE_y[i+1] = \cos(y \cdot \beta[i]), \end{aligned} \tag{1}$$

where $\beta[i] = \frac{1}{10000^{2i/d}}$, d is half the embedding dimension, and i indexes the frequencies.

Rotary Encoding. We utilize a similar approach as the Sinusoidal encoder with the Rotary encoder, where the encoding transforms feature embeddings to encode relative positions using rotational transformations on x and y embeddings:

$$\begin{aligned} \mathbf{x}_{\text{rot}} &= \mathbf{x} \odot \cos(\theta) + \text{Rot}_{90}(\mathbf{x}) \odot \sin(\theta), \\ \mathbf{y}_{\text{rot}} &= \mathbf{y} \odot \cos(\theta) + \text{Rot}_{90}(\mathbf{y}) \odot \sin(\theta), \end{aligned} \tag{2}$$

where θ is derived from scaled positional values, and $\text{Rot}_{90}(\cdot)$ rotates the second half of the embedding vector by 90 degrees. This method emphasizes spatial relationships by directly modifying feature representations in embedding space.

Learned Encoding. This method learns position embeddings directly through a linear layer:

$$PE = \mathbf{W} \cdot \mathbf{c} + \mathbf{b}, \tag{3}$$

where $\mathbf{c} \in \mathbb{R}^2$ represents the input coordinates, and $\mathbf{W} \in \mathbb{R}^{d\times2}$, $\mathbf{b} \in \mathbb{R}^d$ are trainable parameters. This approach adapts the encoding to dataset-specific requirements but risks overfitting in low-data scenarios.

Practical Considerations. Unlike static encodings in image-based transformers, the dynamic nature of grouped

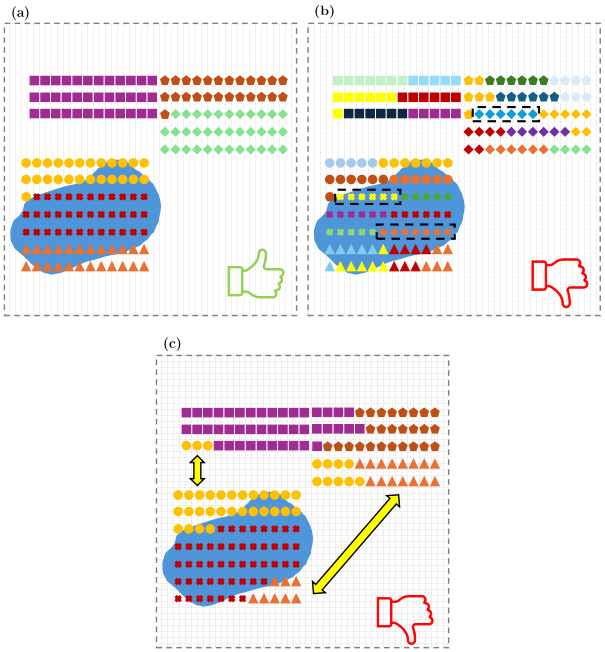


Figure 3. **Comparison between the effects that different groups sizes can have on group semantics and composition.** An ideal group size (a) has sufficiently large groups to just enough cover pertinent regions like puddles. A group size that is too small (b) has difficulty spanning the gap between B-scans, making it no different than previous methods like Roofus [4]. Groups that are too large (c) may introduce unwanted artifacts such as discontinuity and noise.

A-scans in GPR data necessitates real-time encoding computation. However, we find that the real-time computation adds minimal overhead in practice across various large-scale rooftop datasets.

While Sinusoidal Absolute Encoding efficiently captures global structure, Rotary Encoding and Learned Encoding emphasize local spatial patterns and task-specific adaptability, respectively. These differences highlight trade-offs in flexibility, interpretability, and computational requirements for each encoding scheme.

3.4 Transformer Architecture

In GPR-former, we adopt a vanilla transformer architecture similar to the design introduced by [11]. This transformer is central to our approach as it models the contextual and spatial dependencies among grouped A-scan features. Each transformer layer block is comprised of multi-head self-attention, a feed-forward network, residual connections, and layer normalization. After the transformer encoder, the resulting feature vector is passed through a final linear layer followed by a sigmoid activation to yield a binary classification probability.

Figure 2 illustrates the summarized approach of GPR-former.

4 Experiments

Implementation. GPR-former employs a transformer encoder architecture with a binary classification head as shown in Figure 2. Training is conducted using Focal Loss to address the inherent dataset imbalance (it is typical to expect moisture signals to be less frequent than dry signals), with a learning rate of 1×10^{-5} on a cosine annealing schedule. Weighted binary cross-entropy was also tested but yielded suboptimal results compared to Focal Loss, likely due to the latter's ability to emphasize difficult samples more effectively. We use the AdamW optimizer and train on an Nvidia RTX 8000 GPU with a batch size of 64.

Three versions of GPR-former are developed: **GPR-former-B** (base), **GPR-former-L** (large), and **GPR-former-H** (huge). These versions differ in the number of attention heads, layers, and representation dimensions, with GPR-former-B containing a total of 44.6M parameters, GPR-former-L containing 84.7M parameters, and lastly GPR-former-H containing 118.9M parameters. For reference, the architecture used for Roofus contains 302M parameters, making GPR-former noticeably light-weight in comparison. The exact details of each configuration is described in Figure 1. The architectures share a common classification layer, ensuring consistency across the comparison. Hyperparameters were optimized through grid search, with specific focus on dropout, learning rate, and batch size. All GPR-former models in our main evaluation use Sinusoidal positional encoding and a group size $\omega = 1024$. We provide the justification in our ablation analysis in Section 4.2.

Dataset. The dataset is sourced from the same building complex used in the Roofus investigation [4], a 46.5K m^2 retail complex with diverse roofing materials, including modified bituminous membranes, EPDM rubber, and TPO. It contains hundreds of thousands of A-scans spanning a wide range of roofing conditions and damage levels. Labels were manually annotated and cross-validated against reference moisture maps provided by a third-party partner. These reference maps were generated using established rooftop moisture detection techniques, including infrared thermography, electrical capacitance testing, and nuclear moisture gauges.

The dataset is divided into distinct sections, labeled Section 1 through Section 10, with Section 7 excluded due to scanning errors. Section 6 is reserved exclusively for testing, while the remaining sections are used for training and validation with an 85/15 split. This division ensures

Models	Hyperparameters			# Parameters
	# Heads	# Layers	Representation Dim.	
Roofus	16	24	1024	302M
GPR-former-B	4	8	768	44.6M
GPR-former-L	8	10	1024	84.7M
GPR-former-H	16	10	1280	118.9M

Table 1. **Design differences between the different sizes of GPR-former, with the tested version of Roofus included for comparison.** The different sized models vary in the number of attention heads and layers as well as their representation dimensions.

Models	Quantitative			Qualitative
	AR ↑	F2-Score ↑	AUROC ↑	NLLloss ↓
Roofus	0.4563	0.7158	0.8240	0.4940
GPR-former-B	0.4563	0.7267	0.8386	0.4465
GPR-former-L	0.4730	0.7174	0.8296	0.4443
GPR-former-H	**0.4834**	**0.7277**	**0.8436**	**0.4376**

Table 2. **Comparison of GPR-former variations and Roofus across evaluation metrics on the test set.** GPR-former-H appears to be better fit to take advantage of the additional spatial information through the increased number of encoder layers and attention heads.

that the model is evaluated on unseen data, reflecting real-world deployment scenarios.

Prior to input into the models, the GPR B-scans are pre-processed using a linear gain down the time domain. The data is then standardized by the mean and standard deviation values calculated from the training set. Additionally, we apply a random horizontal or vertical flip of the x, y coordinates with a separate probability of 0.5 for each flip during training.

Baseline. To benchmark the performance of GPR-former, we compare it with our previously developed Roofus model. Roofus also employs a transformer-based architecture but operates on GPR B-scans, treating A-scans within a window as sequential data without incorporating spatial context beyond individual B-scans. Roofus was retrained on the current dataset using the same training regimen as GPR-former, with adjustments where necessary to accommodate its architectural differences.

Evaluation Metrics. Given the critical need to identify all potentially damaged roofing areas, **recall** is selected as the key evaluation metric. While precision is valuable, prioritizing recall ensures minimal false negatives, which is essential for identifying areas that might require repair. In roofing diagnostics, conservative predictions often lead to better outcomes, as even marginally damaged areas are typically repaired along with severely damaged sections.

To capture this priority, we use metrics such as Average Recall (AR) and F2-score, with F2-score weighted more heavily toward recall. The Area Under the Receiver Operating Characteristic (AUROC) is also included to provide both consistency with our previous investigation with Roofus as well as further insights into the trade-off between true positive and false positive rates. F2-scores are calculated using the optimal threshold identified during the final validation step.

4.1 Main Results

Quantitative. Table 2 presents the performance of GPR-former variants compared to Roofus. The results show that the largest model, GPR-former-H, using $\omega = 1024$ and the Sinusoidal positional encoding, achieves the highest recall, F2-score, and AUROC, demonstrating the benefits of incorporating spatial context into the model. The slight underperformance of GPR-former-L compared to GPR-former-B in F2-Score and AUROC is an unexpected result, potentially due to overfitting or architectural inefficiencies. Further analysis of this phenomenon is left for future work.

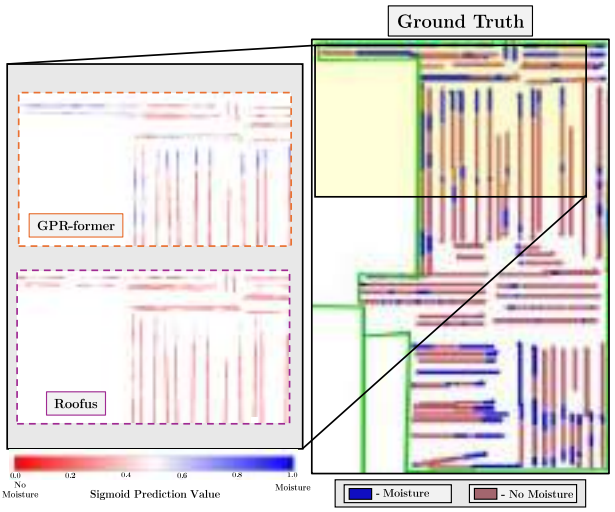
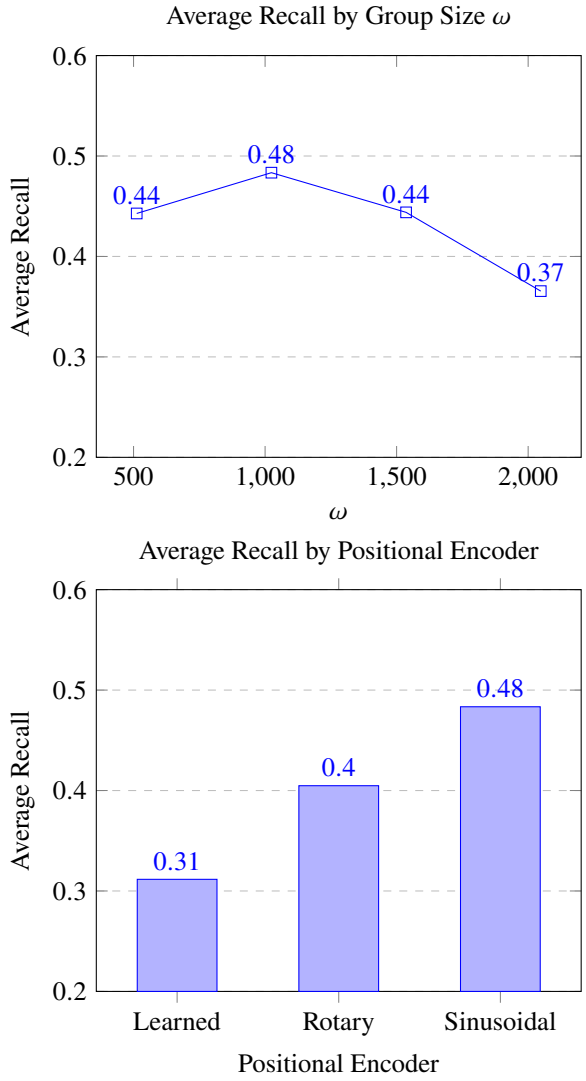


Figure 4. **Qualitative comparisons against the ground truth labels of the un-thresholded moisture predictions generated by GPR-former and Roofus [4].** The example highlighted region clearly illustrates the improvement in the calibration of GPR-former is over its predecessor.

Qualitative. Despite being hidden from client-facing reports, the un-thresholded predictions can provide additional insights into how well-calibrated the models are and the uncertainty present in them. Figure 4 illustrates the raw model outputs prior to thresholding for GPR-former and Roofus. GPR-former exhibits a wider distribution of prediction scores, indicating higher confidence in its classifications. To quantify uncertainty, we compute the negative log-likelihood loss as shown in Table 2, reveal-

ing that GPR-former produces more calibrated predictions compared to Roofus.



4.2 Ablation

Group size ω. As discussed prior, the group size ω determines the spatial context provided to the model. Larger groups introduce additional spatial information but risk incorporating irrelevant or noisy data, while smaller groups may lack sufficient context. To strike a balance, we tested group sizes of $\omega = 512, 1024, 1536, 2048$. Results indicate optimal performance at $\omega = 1024$, beyond which performance declines, likely due to grouping artifacts or interruptions in data continuity. Overlapping groups or advanced sampling methods may address these limitations in future iterations.

Positional Encoding. As previously mentioned, the three encoding schemes were evaluated: sinusoidal absolute positional encoding, a rotary positional encoder, and a learned positional encoder. Contrary to expectations, the sinusoidal encoder outperformed the rotary encoder, possibly due to the relative encoder's complexity relative to the dataset size. The poor performance of the learned encoder underscores the importance of carefully designed encoding schemes in spatial transformers, especially in low-data scenarios.

5 Discussions

Limitations. This study shares limitations with prior work [4], including the need for an expanded dataset with additional roofing material types and environmental conditions to generalize the model's applicability. Artifacts in the sampling procedure, such as large gaps between group members, can lead to discontinuities in spatial context. Furthermore, the current approach is unable to quantify the severity of damage or localize the damage beneath the surface, which somewhat limits the practical utility of the predictions.

Conclusions. The experiments validate the effectiveness of incorporating spatial context in transformer-based architectures for GPR data. GPR-former outperforms our previous model Roofus across key metrics, demonstrating the advantages of spatially aware processing.

Future Work. Directions for future research include developing advanced methods for constructing groups (e.g., overlapping groups or clustering-based sampling) and increasing the dataset size through scanning and labeling more rooftops to enhance diversity and saturate the capabilities of relative positional encoding. Investigating the applicability of self-supervised learning may also address the difficulties in large-scale data collection and annotation. Exploring the prediction of moisture depth or damage severity could further improve the practical value of the system in real-world diagnostics.

Acknowledgements. This work is supported by NSF grant #2322242.

Disclosure: Chen Feng is a founder of Building Diagnostic Robotics, Inc., a startup company that uses AI and robotics for building inspections.

References

[1] Sher B. and Feng C. Deepgpr: Learning to identify moisture defects in building envelope assemblies from ground penetrating radar. In Borja García de Soto, Vicente Gonzalez-Moret, and Ioannis Brilakis, editors, *Proceedings of the 40th International Symposium on Automation and Robotics in Construction*, pages 561–568, Chennai, India, July 2023. International Association for Automation and Robotics in Construction (IAARC). ISBN 978-0-6458322-0-4. doi:10.22260/ISARC2023/0075.

[2] Luo T. X., Zhou Y., Zheng Q., Hou F., and Lin C. Lightweight deep learning model for identifying tunnel lining defects based on gpr data. *Automation in Construction*, 165:105506, 2024. ISSN 0926-5805. doi:https://doi.org/10.1016/j.autcon.2024.105506. URL https://www.sciencedirect.com/science/article/pii/S0926580524002425.

[3] Hoang N. Q., Shim S., Kang S., and Lee J. Anomaly detection via improvement of gpr image quality using ensemble restoration networks. *Automation in Construction*, 165:105552, 2024. ISSN 0926-5805. doi:https://doi.org/10.1016/j.autcon.2024.105552. URL https://www.sciencedirect.com/science/article/pii/S0926580524002887.

[4] Lee K., Lin W., Javed T., Madhusudhan S., Sher B., and Feng C. Roofus: Learning-based robotic moisture mapping on flat rooftops with ground penetrating radar. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2024.

[5] Xu T., Yuan D., Yang G., Li B., and Fan D. Fm-gan: Forward modeling network for generating approaching-reality b-scans of gpr pipelines with transformer tuning. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–13, 2024. doi:10.1109/TGRS.2024.3359351.

[6] Chen Y., Li F., and Zhou S. New asphalt pavement dielectric constant inversion net based on improved gated attention for ground penetrating radar. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1–11, 2024. doi:10.1109/TGRS.2024.3496881.

[7] Eldemiry A., Muddassir M., and Zayed T. Autonomous data acquisition of ground penetrating radar (gpr) using lidar-based mobile robot. In Vicente Gonzalez-Moret, Jiansong Zhang, Borja García de Soto, and Ioannis Brilakis, editors, *Proceedings of the 41st International Symposium on Automation and Robotics in Construction*, pages 206–212, Lille, France, June 2024. International Association for Automation and Robotics in Construction (IAARC). ISBN 978-0-6458322-1-1. doi:10.22260/ISARC2024/0028.

[8] Wu K., Rodriguez G. A., Zajc M., Jacquemin E., Clément M., De Coster A., and Lambot S. A new drone-borne gpr for soil moisture mapping. *Remote Sensing of Environment*, 235:111456, 2019. ISSN 0034-4257. doi:https://doi.org/10.1016/j.rse.2019.111456. URL https://www.sciencedirect.com/science/article/pii/S0034425719304754.

[9] Garrido I., Solla M., Lagüela S., and Fernández N. Irt and gpr techniques for moisture detection and characterisation in buildings. *Sensors*, 20(22), 2020. ISSN 1424-8220. doi:10.3390/s20226421.

[10] Lee K., Lin W., Sher B., Javed T., Madhusudhan S., and Feng C. Exploring self-supervised gpr representation learning for building rooftop diagnostics. In Vicente Gonzalez-Moret, Jiansong Zhang, Borja García de Soto, and Ioannis Brilakis, editors, *Proceedings of the 41st International Symposium on Automation and Robotics in Construction*, pages 928–935, Lille, France, June 2024. International Association for Automation and Robotics in Construction (IAARC). ISBN 978-0-6458322-1-1. doi:10.22260/ISARC2024/0120.

[11] Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A., Kaiser L., and Polosukhin I. Attention is all you need. In Guyon I., Von Luxburg U., Bengio S., Wallach H., Fergus R., Vishwanathan S., and Garnett R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[12] Das B. K., Zhao G., Islam S., Re T. J., Comaniciu D., Gibson E., and Maier A. Co-ordinate-based positional embedding that captures resolution to enhance transformer's performance in medical image analysis. *Sci. Rep.*, 14(1):9380, April 2024.

[13] Li H., Dong H., Jia H., Huang D., Kampffmeyer M. C., Lin L., and Liang X. Coordinate transformer: Achieving single-stage multi-person mesh recovery from videos. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 8744–8753, October 2023.

[14] Li Y., Ma Z., Wang X., Wang Y., and Tan B. Vision transformer with 2d explicit position encoding. In *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7690–7694, 2024. doi:10.1109/ICASSP48485.2024.10446293.

[15] He K., Chen X., Xie S., Li Y., Dollár P., and Girshick R. Masked autoencoders are scalable vision learners. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 15979–15988, 2022. doi:10.1109/CVPR52688.2022.01553.

Dynamic Analysis of a Six-Cable Parallel Robot for Automated Panelized Building Retrofits

Yifang Liu, Nolan W. Hayes, Diana Hun, and Bryan P. Maldonado

Buildings and Transportation Science Division, Oak Ridge National Laboratory*, United States of America

liuy5@ornl.gov, hayesnw@ornl.gov, hunde@ornl.gov, maldonadopbp@ornl.gov

Abstract -

Cable-Driven Parallel Robots (CDPRs) are highly suitable for automated panelized building retrofits, thanks to their compact footprint and high payload-to-weight ratio. A common CDPR configuration featuring eight cables, where the anchors form a rectangular prism in front of the building facade, offers a large wrench feasible workspace and good control versatility. However, installing upper anchor points requires additional support structures, such as towers or beams, increasing setup complexity and posing logistical challenges in construction settings. To mitigate these challenges, we propose a six-cable CDPR model specifically designed for automated panelized building retrofits. Although the feasible workspace is limited, our analysis shows that the proposed CDPR adequately covers the critical areas required for panel installation. To validate that the six-cable system can effectively transport the end effector to the desired installation pose, we calculated optimal trajectories based on a constrained dynamic model. The simulation results of the six-cable CDPR demonstrate promising potential for automated panelized building retrofits, effectively balancing simplicity, cost-effectiveness, and functionality.

Keywords -

CDPR; Panelized Envelope; Construction Robotics; Trajectory Optimization

1 Introduction

Building retrofits using panelized prefabricated envelopes substantially enhance the energy efficiency of a building. Traditional methods for installing panels in multi-story buildings typically rely on manual labor or crane hoisting, with panels being placed by hand [1]. However, these approaches are often hindered by labor shortages, high costs, and panel installation errors. Automating this process with robotics offers a promising solution to enhance efficiency, improve quality, and reduce the overall installation cost. Cable-Driven Parallel Robots (CDPRs) are well-suited for handling heavy payloads while maintaining a compact footprint. These attributes make them particularly effective for automated building retrofits using prefabricated panelized systems.

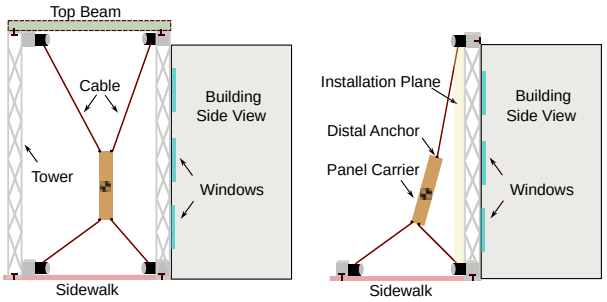


Figure 1. Side views of eight-cable (left) and six-cable (right) CDPRs for panelized building retrofits

CDPRs are highly versatile robotic systems used in various applications, including 3D printing [2], material handling [3], and prefabricated panel installation [4]. Many CDPR designs utilize an eight-cable configuration supported by rigid external elements surrounding the workspace [5]. This setup, with four anchor points on the ground and four positioned mid-air, offers a large wrench feasible workspace and high control versatility. However, installing the upper anchor points requires additional support structures, such as towers, cantilever beams, or bracing (left plot of Figure 1), which significantly increases setup complexity and presents logistical challenges in construction environments.

To address these issues, this paper introduces a novel six-cable CDPR model designed specifically for panelized building retrofits. As shown in the right plot of Figure 1, the proposed system uses six cables, with four anchors on the ground and two near the roof. By removing the anchor points that need to be suspended mid-air, this configuration simplifies setup process and mitigates logistical complexities. When the end-effector is away from the facade, the

*Notice: This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (https://www.energy.gov/doe-public-access-plan).

system operates as a spatial (three-dimensional) CDPR. However, as the center of mass of the end effector aligns with the installation plane, the bottom two cables furthest from the facade begin to sag, transitioning the system into a planar (two-dimensional) configuration. A key focus of this study is evaluating if the model is capable to transport the end effector from a starting feasible pose away from the facade (e.g., at the end of the sidewalk) to a desired pose at the installation plane near the facade. Using a dynamic model of the six-cable CDPR, a trajectory optimization problem was formulated to demonstrate the feasibility of the proposed design. Simulation results showed that the six-cable CDPR can successfully pick up a panel at a given pose and manipulate it to achieve the desired plumb and level panel position on the installation plane. To the best of our knowledge, this is the first six-cable configuration that is used for panelized building retrofits. This streamlined configuration aims to enhance feasibility and efficiency while maintaining sufficient operational control for panelized building retrofits.

The remaining sections are arranged as follows. Section 2 presents the state of the art in CDPRs for construction applications. Section 3 describes the benefits and challenges of a six-cables CDPR system. Section 4 defines the dynamic model used and the trajectory optimization problem to be solved. Section 5 demonstrates the simulation results. Section 6 concludes the paper and outlines directions for future work.

2 CDPRs in Construction Applications

Recently, CDPRs have gained attention across various fields and applications. Due to their desirable characteristics, such as scalability, reconfigurability, compact design, and high payload-to-weight ratio. They are widely investigated in construction applications such as bricklaying [6], 3D printing [7, 8], and solar power plant assembly [9]. Murphy *et al.* developed a planar CDPR for cooperative modular manipulation, enabling CDPRs to assist in construction tasks and gap-crossing scenarios [10]. For building envelope applications, Izard *et al.* explored the use of eight-cable CDPRs for inspecting building facades [11]. In addition, Iturralde *et al.* designed and implemented an eight-cable CDPR model for modular curtain wall installation in real-world applications [5]. Their work examined whether CDPRs could achieve sufficient accuracy for installing curtain wall modules and reduce manual installation time. In the context of automated panelized building retrofits, our previous work investigated the wrench feasible workspace for various cable configurations, with a primary focus on four-cable planar models and eight-cable 3D models [4]. CDPRs designed for building envelope applications have predominantly focused on eight-cable models, leaving the potential of a six-cable CDPR configuration unexplored.

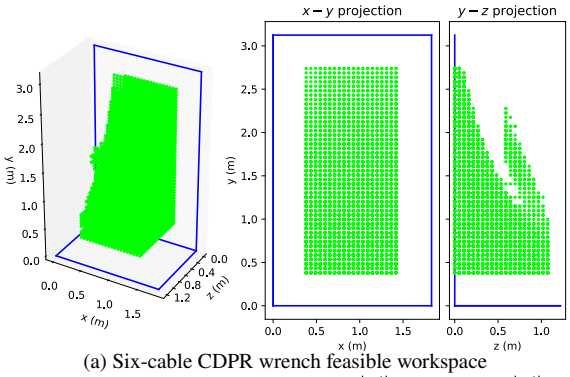
3 Six-Cable CDPR for Panelized Systems

To achieve control over the six degrees of freedom (DOF) of the end effector, an eight-cable configuration is commonly used, especially when a large workspace is required. Among the eight proximal anchor points, two need to be placed away from the building facade and away from the ground. This setup usually requires additional support structures, such as vertical towers and/or top beams, as illustrated in Figure 1. Support towers provide a simplified setup by providing the necessary rigidity for the system while minimizing dependency on the building. In this case, if a CDPR with eight cables is needed, then four towers are required to attach all anchor points. Although top beams or bracing are not strictly mandatory, they become essential when a highly rigid support structure is needed. However, constructing four towers with beams and bracing presents logistical challenges and adds considerable expenses.

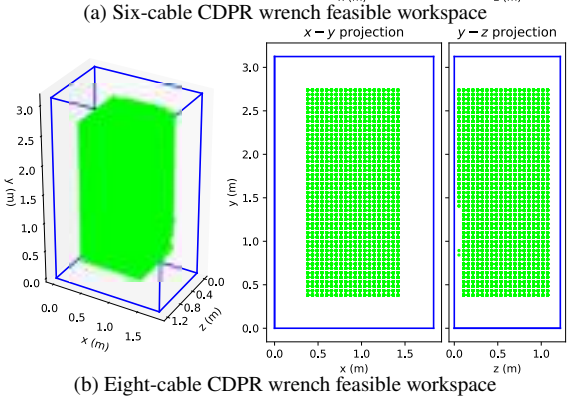
In contrast, some systems, such as Hephaestus [5], use only two welded-steel cantilever structural beams fastened to the roof of the building, with each beam element supporting two anchor points. While this approach reduces setup complexity, it is not universally applicable, particularly for older buildings where additional load-bearing capacity of the structure may be limited. The feasibility of relying solely on roof-mounted structures depends heavily on the building's age, structural integrity, and availability of attachment points. For such cases, a structural assessment is necessary to ensure that the building can support the CDPR system, which adds additional complexity and cost for retrofitting. To address these limitations and improve system robustness, we propose avoiding roof-mounted structures entirely, so that the CDPR design can be largely independent of the building's architecture.

Therefore, the tower-based setups, i.e. the combination of towers, top beams, and bracing, ensures a rigid and reliable system while minimizing dependence on the building structure itself to support the robot. Many existing multifamily residential buildings have sufficient space adjacent to the facade to accommodate the tower-based setups, which provides a suitable condition for retrofitting projects using CDPRs.

To further simplify the installation process and reduce costs, this paper investigates the feasibility of using a CDPR with only six cables for panelized building retrofits. This approach aims to maintain functionality while significantly minimizing the structural and logistical requirements associated with the traditional eight-cable configurations. The six-cable configuration offers several notable advantages. As illustrated in Figure 1, unlike the eight-cable setup that requires four top anchors, the six-cable



(a) Six-cable CDPR wrench feasible workspace



(b) Eight-cable CDPR wrench feasible workspace

Figure 2. Wrench feasible workspace of the six-cable CDPR and an eight-cable CDPR. Green dots represent feasible positions of the end effector's center of mass. Blue lines represent the CDPR frame. Left: 3D view; Middle: $x - y$ projection; Right: $y - z$ projection. Further details on the calculations can be found in Section 5.2.

system requires only two anchors located near the top of the building. This reduces the number of towers needed from four to two. Additionally, a six-cable CDPR system uses only six actuator systems, further lowering costs. The reduction in cables and towers minimizes material usage and overall expenses. These adjustments simplify the setup process and reduce installation complexity. However, this approach also introduces technical challenges. Unlike the eight-cable system, the six-cable configuration cannot provide sufficient control to achieve the same level of mobility for the end effector. This limitation reduces the workspace (see Figure 2), causing more than 30% of the wrench feasible area in the eight-cable CDPR to become wrench-infeasible. Additionally, it increases sensitivity to external disturbances and adds complexity to control and trajectory optimization. For example, in the eight-cable CDPR, when the end effector is at the pickup location, a wide range of initial orientations are wrench-feasible. However, for the six-cable CDPR, most of these initial orientations are no longer feasible. Additionally, determining

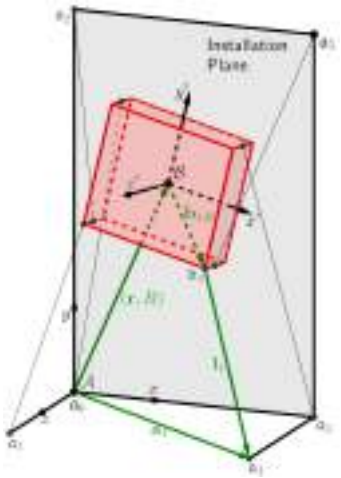


Figure 3. Geometric and dimensional definition of the six-cable CDPR model for building retrofits.

a feasible initial trajectory between a given starting position and target pose presents a significant challenge. This paper addresses these issues by proposing solutions to enhance the practicality and reliability of a six-cable CDPR. By overcoming these challenges, the six-cable configuration can serve as a promising alternative for certain retrofit applications, offering a balance between simplicity, cost-effectiveness, and functionality.

4 Modeling and Trajectory Optimization

We will start by describing the CDPR dynamic model and then formulating the trajectory optimization problem for bringing a panel from a starting pose to a desired pose.

4.1 CDPR Dynamics

The six-cable CDPR model studied in this paper is illustrated in Fig. 3. Note that in this setup, the installation plane is parallel to, but distinct from the building facade. The building facade is located on the negative side of the z axis. The fixed (black) frame of size $\boldsymbol{F} = \begin{bmatrix} F_x & F_y & F_z \end{bmatrix}^T$ sets the dimensions of the CDPR in the world coordinate system $\mathcal{A}$. The CDPR manipulates the (red) end effector of size $\boldsymbol{P} = \begin{bmatrix} P_{x'} & P_{y'} & P_{z'} \end{bmatrix}^T$, which defines the dimensions of the panel in the rotating coordinate system $\mathcal{B}$. Note that in real-world panel installation using CDPRs, the end effector serves as the panel carrier and may be bigger than the panel itself. Finally, $n = 6$ massless cables connect the panel to the frame. The cable end connected to the frame runs through pulleys and cable drums to facilitate the winding and unwinding of the cable. However, in this study, we will not model the geometry or dynamics of

the winding mechanism. Consider the following vectors for $i \in \{1, \ldots, n\}$, according to Figure 3:

- The constant vectors $\mathbf{a}_i$ denote the proximal anchor points $\mathfrak{a}_i$ in the frame with respect to $\mathcal{A}$.
- The constant vectors $\mathbf{b}_{i,\mathcal{B}}$, denote the distal anchor points $\mathfrak{b}_i$ in the panel with respect to $\mathcal{B}$.
- The pose $(\boldsymbol{r}, R)$ is defined by the location of the panel's center of mass $\boldsymbol{r} = \begin{bmatrix} r_x & r_y & r_z \end{bmatrix}^T$ relative to $\mathcal{A}$, and the orientation of the panel's frame $\mathcal{B}$ relative to $\mathcal{A}$, denoted by the rotation matrix $R \in SO_3$.
- The cable vectors $\boldsymbol{l}_i = \mathbf{a}_i - (\boldsymbol{r} + R\mathbf{b}_{i,\mathcal{B}})$ which are functions of the panel pose calculated relative to $\mathcal{A}$.
- The cable forces $\boldsymbol{f}_i = f_i \boldsymbol{l}_i / \|\boldsymbol{l}_i\|$ with $f_i \geq 0$ tension acting on the cables ($\|\cdot\|$ denotes Euclidean norm).

The translational motion of the panel is defined in terms of the acceleration of its center of mass, which is measured from the inertial frame of reference $\mathcal{A}$. Therefore, the Newton equations of translational motion are [12]:

$$\frac{d}{dt}\boldsymbol{r} = \boldsymbol{v} \tag{1}$$

$$\frac{d}{dt}\boldsymbol{v} = \boldsymbol{g} + \frac{1}{m}\sum_{i=1}^{n} f_i \frac{\boldsymbol{l}_i}{\|\boldsymbol{l}_i\|} - \frac{\mu_{\boldsymbol{v}}}{m}\boldsymbol{v}. \tag{2}$$

Here, $\boldsymbol{v}$ is the velocity of the panel's center of mass, $\boldsymbol{g}$ is the acceleration due to gravity, m is the panel's mass, and $\mu_{\boldsymbol{v}}$ is a linear friction coefficient to provide damping. Since the coordinate system $\mathcal{B}$ moves with the panel, the rate of change in orientation and Euler equations of motion written with respect to the coordinate system $\mathcal{B}$ are [12]:

$$\frac{d}{dt}\boldsymbol{\Phi} = T\boldsymbol{\omega}_{\mathcal{B}}. \tag{3}$$

$$\frac{d}{dt}\boldsymbol{\omega}_{\mathcal{B}} = -I_{\mathcal{B}}^{-1}\left(\boldsymbol{\omega}_{\mathcal{B}} \times I_{\mathcal{B}}\boldsymbol{\omega}_{\mathcal{B}}\right) + I_{\mathcal{B}}^{-1}\sum_{i=1}^{n} f_i \left(\mathbf{b}_{i,\mathcal{B}} \times \frac{R^T\boldsymbol{l}_i}{\|\boldsymbol{l}_i\|}\right) - I_{\mathcal{B}}^{-1}\mu_{\omega}\boldsymbol{\omega}_{\mathcal{B}}. \tag{4}$$

Here, $\boldsymbol{\Phi} = \begin{bmatrix} \varphi & \theta & \psi \end{bmatrix}^T$ denotes the Euler angles of the panel. The composition order $R = R_z(\psi)R_y(\theta)R_x(\varphi)$ was adopted for the elemental rotations to compute the rotation matrix. Then, the transformation matrix is [13]:

$$T = \begin{bmatrix} 1 & \sin\varphi\tan\theta & \cos\varphi\tan\theta \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi\sec\theta & \cos\varphi\sec\theta \end{bmatrix}. \tag{5}$$

In addition, $I_{\mathcal{B}}$ is the panel's inertia tensor (solid rectangular cuboid of dimensions $\boldsymbol{P}$), $\boldsymbol{\omega}_{\mathcal{B}}$ is its angular velocity observed from the rotating reference frame $\mathcal{B}$, and μ_{ω} is another friction coefficient. Equations (1)–(4) define the CDPR dynamics under gravitational, friction, and cable tensile forces. Let $x = \begin{bmatrix} \boldsymbol{r}^T & \boldsymbol{\Phi}^T & \boldsymbol{v}^T & \boldsymbol{\omega}_{\mathcal{B}}^T \end{bmatrix}^T$ be the combined state and $u = \begin{bmatrix} f_1 & \cdots & f_n \end{bmatrix}^T$ the control input. Then, the nonlinear model can be written as $\frac{d}{dt}x = h(x, u)$.

4.2 Trajectory Optimization

A general trajectory optimization problem can be formulated as minimizing a cost function while satisfying a set of constraints, including dynamics constraints, path constraints, as well as initial and terminal conditions. Consider the following cost function, formulated as the sum of the control effort and two regularization terms that ensure the smoothness of the generated motion:

$$J = \int_0^{t_T} \frac{\alpha}{N}\|\ddot{x}\|_{W_x}^2 + \frac{\beta}{N}\|\ddot{u}\|^2 \, dt. \tag{6}$$

Here, the coefficients α and β are nonnegative, N is a scaling constant equal to the number of samples observed during the trajectory, $W_x = \mathrm{diag}\begin{pmatrix} 0, & 0, & \mathbb{I}, & \mathbb{I} \end{pmatrix}$ is used for the weighted squared norm, and t_T is the terminal time. The constraints for the optimal control problem $\min_u J$ are:

Dynamic Constraints: Equations of motion

$$\frac{d}{dt}x = h(x, u), \text{ for } t \in (0, t_T) \tag{7}$$

Force Bounds: Lower bound ensures the cables only pull (sagging is allowed), while the upper bound maintains the cable tension below the rupture limit.

$$0 \leq f_i \leq f_{\max}, \forall i \in \{1, \ldots, n\} \tag{8}$$

Path Constraints: Maintain the panel inside the CDPR frame and limit both the linear and angular velocities.

$$0 \leq \boldsymbol{r} + R\mathbf{b}_{i,\mathcal{B}} \leq \boldsymbol{F}, \forall i \in \{1, \ldots, n\} \tag{9}$$

$$-\Phi_{\lim} \leq \boldsymbol{\Phi} \leq \Phi_{\lim} \tag{10}$$

$$-v_{\lim} \leq \boldsymbol{v} \leq v_{\lim} \tag{11}$$

$$-\omega_{\lim} \leq \boldsymbol{\omega}_{\mathcal{B}} \leq \omega_{\lim} \tag{12}$$

Initial Conditions: Equality constraints to ensure that the trajectory starts from $(\boldsymbol{r}_0, R_0)$ at a stationary state $\boldsymbol{v} = \boldsymbol{\omega}_{\mathcal{B}} = 0$. Let $\boldsymbol{l}_{i,0}$ be the cable vectors at $t = 0$. Then, stationary of the Newton-Euler equations (2) and (4) at the initial pickup pose becomes:

$$x_0 = \begin{bmatrix} \boldsymbol{r}_0^T & \boldsymbol{\Phi}_0^T & 0 & 0 \end{bmatrix}^T \tag{13}$$

$$0 = \begin{bmatrix} \dfrac{\boldsymbol{l}_{1,0}}{\|\boldsymbol{l}_{1,0}\|} & \cdots & \dfrac{\boldsymbol{l}_{n,0}}{\|\boldsymbol{l}_{n,0}\|} \\ \mathbf{b}_{1,\mathcal{B}} \times \dfrac{R_0^T\boldsymbol{l}_{1,0}}{\|\boldsymbol{l}_{1,0}\|} & \cdots & \mathbf{b}_{n,\mathcal{B}} \times \dfrac{R_0^T\boldsymbol{l}_{n,0}}{\|\boldsymbol{l}_{n,0}\|} \end{bmatrix} u_0 + \begin{bmatrix} m\boldsymbol{g} \\ 0 \end{bmatrix} \tag{14}$$

where x_0 and u_0 are the initial state and control. Eqn. (14) can be written in a compact matrix-vector form as $0 = A_0^T u_0 + w_\text{p}$, where A^T is referred to as the structure matrix and w_p is the stationary wrench [14].

Terminal Conditions: Equality constraints to ensure that the robot comes to a complete stop at the target pose $(\boldsymbol{r}_T, R_T)$ at time t_T. Static balance guarantees that when the robot ceases movement, the motors experience minimal stress during the stopping process:

$$x_T = \begin{bmatrix} \boldsymbol{r}_T^T & \boldsymbol{\Phi}_T^T & 0 & 0 \end{bmatrix}^T \tag{15}$$

$$0 = A_T^T u_T + w_\text{p}. \tag{16}$$

Here, x_T, u_T, A_T^T are the terminal state, terminal control, and structure matrix evaluated at t_T, respectively.

5 Simulation Results

In this section, we begin by assigning numerical values to the model parameters for simulation. Next, we present the six-cable wrench feasible workspace. Finally, we discuss the optimized trajectories for transporting the panel from a starting position to the desired final pose.

5.1 Simulation Parameters

We used a 1/3 scaled model of a three-story residential building for our simulations. The simulated cable configuration is illustrated in Figure 3. The initial and terminal poses were defined based on the expected real-world scenario. The panel was picked up from a lower height of the workspace $\boldsymbol{r}_0 = \begin{bmatrix} 0.9 & 1.0 & 0.6 \end{bmatrix}^T$ m with a pre-calculated initial orientation and zero angular and linear velocities. The panel was then held stationary at its final position at the top of the facade, fully plumb and level ($\boldsymbol{\Phi}_T = 0$). Two final positions were tested: $\boldsymbol{r}_T = \begin{bmatrix} 0.8 & 2.2 & 0 \end{bmatrix}^T$ m, and $\boldsymbol{r}_T = \begin{bmatrix} 0.8 & 2.3 & 0 \end{bmatrix}^T$ m. Here, we assume that the robot has sufficient cable length, and the cable length remains positive at all times. Other parameters and constraints are listed in Table 1. CasADi [15] was used as the numerical optimizer with a sampling time of $\Delta t = t_T / N$.

5.2 Wrench Feasible Workspace

To estimate the wrench feasible workspace $\mathcal{W}$, a set of M positions $\mathbf{r}_{j \in \{1,\ldots,M\}}$ was selected from a discretized grid. The following optimization problem was formulated to determine the optimal orientation $R_j(\boldsymbol{\Phi}_j)$ that minimizes the force vector u_j at each corresponding position:

$$\begin{aligned} \min_{\boldsymbol{\Phi}_j} & \|u_j\|^2 \quad \forall j \in \{1, \ldots, M\} \\ \text{s.t. } & 0 = A_j^T u_j + w_\text{p}, \\ & \text{Equations (8) to (10).} \end{aligned} \tag{17}$$

Table 1. Model parameters and constraints

Variable	Value	Description
m	13.6 kg	Panel mass
$\boldsymbol{F}$	$[1.83, 3.12, 1.22]^T$ m	Frame dimensions
$\boldsymbol{P}$	$[0.76, 0.76, 0.25]^T$ m	Panel dimensions
$\mu_{\boldsymbol{v}}$	100 kg/s	Linear friction coefficient
$\mu_{\boldsymbol{\omega}}$	3500 kg · m^2/s	Angular friction coefficient
f_max	1200 N	Cable maximum force
$\boldsymbol{\Phi}_\text{lim}$	90 deg	Euler angle limit
$\boldsymbol{v}_\text{lim}$	0.2 m/s	Linear velocity limit
$\boldsymbol{\omega}_\text{lim}$	2 deg/s	Angular velocity limit
t_T	20 s	Terminal time
N	2000	Samples per trajectory

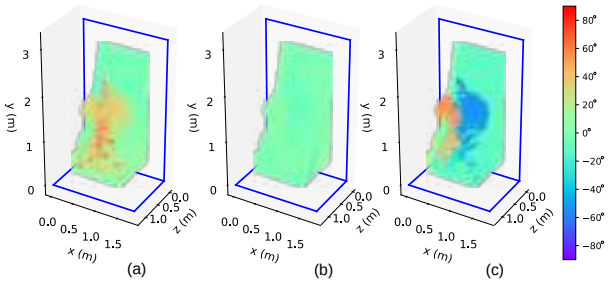


Figure 4. Plots (a), (b), and (c) represent the optimal Euler angles $\boldsymbol{\Phi}_j = [\varphi_j \quad \theta_j \quad \psi_j]^T$, respectively.

If the optimizer converges to a solution, the pose $(\mathbf{r}_j, R_j)$ is wrench-feasible. Figure 4 illustrates the optimized Euler angles $\boldsymbol{\Phi}_j$. The main region of $\mathcal{W}$ resembles a wedge shape and the majority of angles fall within [-20°, 20°]. Note that $\mathcal{W}$ connects the facade installation plane with the assumed pickup panel locations (away from the facade and close to the ground). Thus, there should exist feasible continuous trajectories that can transport the end effector from its starting pose to the installation plane within $\mathcal{W}$. Moreover, we can find trajectories that satisfy certain optimal criteria to minimize installation time and effort.

5.3 Optimized Trajectories

Given the start position $\mathbf{r}_0$ and the final desired pose $(\mathbf{r}_T, R_T)$, we use the trajectory optimization method described in Section 4.2 to compute the optimal force trajectory for the CDPR to follow in simulation. In trajectory optimization, finding a suitable initial trajectory is critical for accelerating the convergence of the optimizer. For the six-cable configuration, finding a suitable initial trajectory is particularly challenging. Firstly, achieving an initial plumb and level orientation $\boldsymbol{\Phi}_0 = 0$ is generally infeasible for most points in the workspace. In contrast, for the eight-cable CDPR, the end effector can be assumed to start in a plumb and level position or with small random rotations.

Consequently, determining an appropriate initial orientation for a given starting position is nontrivial and requires a dedicated optimization process. Selecting an improper orientation—even if wrench-feasible—can lead to convergence failure during optimization. Secondly, generating a practically feasible initial trajectory from the given starting pose to a position near the goal is another significant challenge. The trajectory should ideally satisfy feasibility criteria while avoiding abrupt discontinuities in inputs or states. Unlike in the eight-cable CDPR, where the initial trajectory can be directly obtained using inverse dynamics, the six-cable system requires a more intricate approach to ensure smooth and feasible motion. We address these two challenges in the following paragraphs.

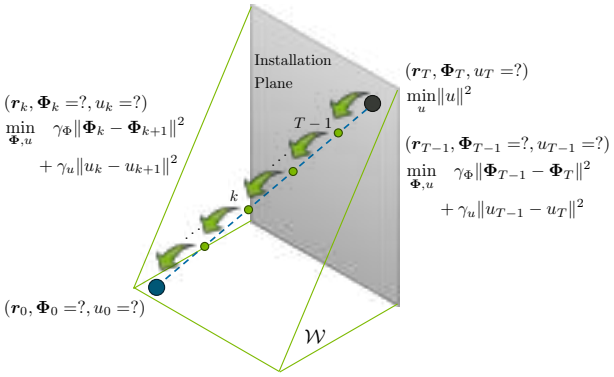


Figure 5. Backwards optimization for finding initial orientation $\boldsymbol{\Phi}_0^*$ and force trajectory $\{u_0^\dagger, \ldots, u_T^\dagger\}$

5.3.1 Determining the orientation of the starting pose

To determine the orientation of the starting pose, we first generate a linear static trajectory connecting the initial and terminal positions: $\{\boldsymbol{r}_0, \boldsymbol{r}_1^\dagger, \ldots, \boldsymbol{r}_k^\dagger, \ldots, \boldsymbol{r}_{T-1}^\dagger, \boldsymbol{r}_T\} \in \mathcal{W}$. At the goal pose $(\boldsymbol{r}_T, \boldsymbol{\Phi}_T)$, an optimization routine similar to Eqn. (17) can be done to find the optimal input force $u_T^\dagger$ by solving $\min_u \|u_T\|^2$. Then, starting from the goal pose and moving backwards along the position trajectory, the optimal orientation $\boldsymbol{\Phi}_k^*$ and input force $u_k^\dagger$ at the k-th ($0 \leq k < T = N-1$) pose is calculated using the constraints in Eqn. (17) and the following cost function:

$$\min_{\boldsymbol{\Phi},u} \gamma_\Phi \|\boldsymbol{\Phi}_k - \boldsymbol{\Phi}_{k+1}\|^2 + \gamma_u \|u_k - u_{k+1}\|^2. \tag{18}$$

Here, γ_Φ and γ_u are weighting factors. The cost function was designed to ensure continuity in both orientation and control input. This optimization is performed position by position, progressing backward until the starting position is reached, as depicted in Figure 5. The resulting optimal continuous quasi-static force trajectory $\{u_0^\dagger, \ldots, u_T^\dagger\}$ is shown at the top plot of Figure 6. Note that the final force values of two cables are zero (cables 1 and 4 start sagging), signaling the transition from a six-cable 3D CDPR system to a four-cable 2D CDPR system. The optimized $\boldsymbol{\Phi}_0^*$ is then set as the orientation for the starting pose.

5.3.2 Generating initial trajectories for the optimizer

Having a well-defined initial trajectory is crucial for accelerating the convergence of the optimizer. Using the starting pose $(\boldsymbol{r}_0, \boldsymbol{\Phi}_0^*)$, static conditions $\boldsymbol{v} = \boldsymbol{\omega}_\mathcal{B} = 0$, and quasi-static force trajectory $\{u_0^\dagger, \ldots, u_T^\dagger\}$ as inputs, we applied the system dynamics in Eqn. (7) to compute the dynamic state trajectory. The bottom plots in Figure 6 compare the dynamic trajectory (solid line) with the quasi-static trajectory (dashed line) used in the backwards optimization. Note that the final pose of the dynamic state trajectory does not coincide with the desired target pose. This occurs because the backward optimization was performed using static instead of dynamic constraints. This simplification impacts the results accuracy, as it fails to account for changes in velocity and acceleration. This dynamic state trajectory was used to initialize the optimizer.

In practice, two routing strategies can be used to generate the initial trajectory. The first method, directly connecting the start and end positions, was described in Section 5.3.1. However, this approach does not guarantee that Eqn. 18 will converge at every point along the trajectory. To address this issue, we propose a second approach that divides the process into two phases. In the first phase, the end effector is moved to a position on the installation plane, which can be selected near the initial position (similar heights). Once the end effector reaches the plane, cables 1 and 4 are slackened, transforming the system into a 2D planar configuration. Each phase's trajectory is optimized separately, and the two optimized segments are concatenated to form the initial trajectory. This trajectory is then used by the optimizer to find the final optimal trajectory from the true starting position to the desired pose.

5.3.3 Final optimized trajectories

Figure 7 illustrates two optimal trajectories originating from the same starting position but with different initial orientations. In the left plot, the trajectory used to initialize the optimizer directly connects the start and end positions. In contrast, the right plot demonstrates an initial trajectory that first guides the end effector to the installation plane before moving upwards along the plane. The reason for splitting the trajectory in the second example is that the optimizer failed when attempting to conduct the backwards optimization to find the optimal orientation for the initial position. As a result, the initial trajectory for the second example first brings the end effector to the position $\begin{bmatrix} 0.8 & 1.2 & 0 \end{bmatrix}^T$ m, before continuing to the final goal

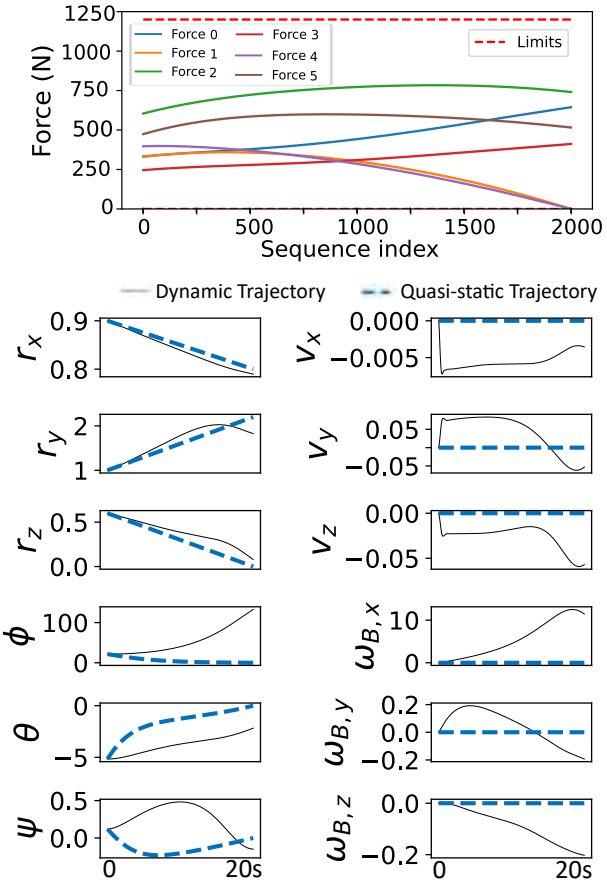


Figure 6. Example of calculated initial trajectory. Top: quasi-static force trajectory calculated using backward optimization. Bottom: comparison between quasi-static and dynamic trajectories. Positions are measured in meters and angles in degrees.

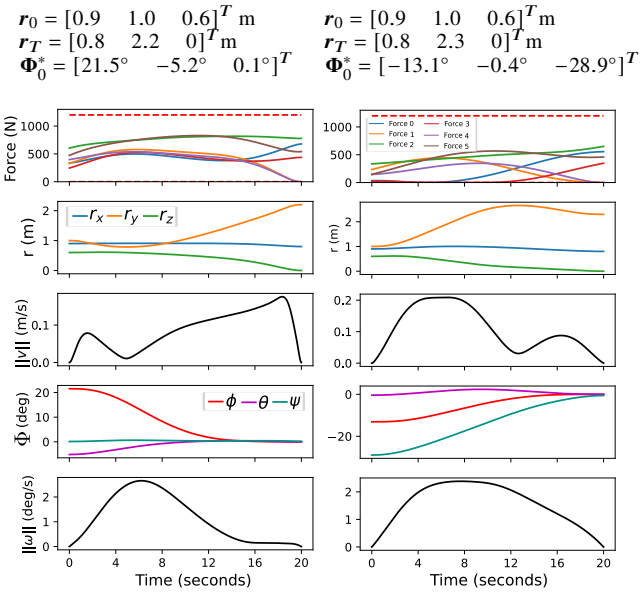


Figure 7. Two optimized trajectories starting from the same position with different initial orientations.

pose.

Although the left plot in Fig. 7 shows a shorter Euclidean distance between $\boldsymbol{r}_0$ and $\boldsymbol{r}_T$ compared to the right plot, the CasADi optimizer achieved only an acceptable level of convergence in the left case, while it successfully reached the optimal solution in the right case. This may explain why the linear velocity profile in the right plot is smoother than that in the left plot. After the optimizer converges, both methods lead to smooth force input trajectories $\{u_0^*, \ldots, u_T^*\}$. The top plots of Fig. 7 show the optimal input force. In method one (left plot), all six cables are used until the end effector reaches the installation plane, then only four cables are used. On the other hand, method two (right plot) achieves the same goal with a lower average input force, as it tends to sag at least one cable throughout most of the trajectory. Although splitting the trajectory into two phases requires running the trajectory optimization process three times: first to bring the end effector to the installation plane, then to travel within the plane, and finally to optimize the trajectory for the final pose, the process proves more practically feasible as it is more robust and converges faster.

6 Conclusion and Future Work

This paper investigated the feasibility of employing a six-cable CDPR for panelized building retrofits, with the goal of simplifying the robot installation and setup process, reducing costs, and simultaneously maintaining the essential functionalities required for building retrofits. Through simulation, we examined one specific cable arrangement, evaluated its wrench feasible workspace, and optimized trajectories within this workspace to transport the end effector from a starting position to the desired goal pose. Given the initial panel pickup position $\boldsymbol{r}_0$, desired panel position $\boldsymbol{r}_T$, and assuming that the panel will be installed plumb and level $\boldsymbol{\Phi}_T = 0$, the proposed method determines the optimal pickup orientation $\boldsymbol{\Phi}_0^*$ and computes the optimal continuous input force trajectory $\{u_0^*, \ldots, u_T^*\}$ that minimizes cable force, maintains the system within safe boundaries, and achieves the desired terminal pose.

This work primarily focuses on studying the feasibility of the six-cable CDPR in a construction setting, specifically investigating whether a dynamically feasible path exists to transport the panel from the bottom to the top. However, we do not address other real-world challenges, such as partial observability or external disturbances, which can significantly impact system performance. These aspects require further exploration and will be our future work.

Additionally, this study did not consider cable-to-cable collisions. Future research will address this limitation by exploring methods to adjust the positions of the anchors in order to avoid such collisions. Moreover, in the six-cable model, two cables must sag as the end effector nears the installation plane. Moving forward, we aim to investigate the sagging mechanism to better understand its effects on system stability and performance. Furthermore, We also plan to explore how the system can be adapted to handle larger-scale applications and more complex scenarios.

ACKNOWLEDGMENT

This research was supported by the DOE Office of Energy Efficiency and Renewable Energy (EERE), Building Technologies Office, under the guidance of Sven Mumme, and used resources at the Building Technologies Research and Integration Center, a DOE-EERE User Facility at Oak Ridge National Laboratory. Research also sponsored by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the U. S. Department of Energy.

References

[1] Christian Mascheck. Serielles sanieren: So funktioniert der "energiesprong". Online: `https://cradle-mag.de/artikel/serielles-sanieren-energiesprong.html`, Accessed: 12/19/2024.

[2] Jean-Baptiste Izard, Alexandre Dubor, et al. Large-scale 3d printing with cable-driven parallel robots. *Construction Robotics*, 1:69–76, 2017.

[3] Tobias Bruckmann, Wildan Lalo, et al. Development of a storage retrieval machine for high racks using a wire robot. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, volume 45035, pages 771–780. American Society of Mechanical Engineers, 2012.

[4] Yifang Liu, Rui Zhang, et al. Cable-driven parallel robot (cdpr) for panelized envelope retrofits: Feasible workspace analysis. In *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, volume 41, pages 1081–1088, 2024.

[5] Kepa Iturralde, Malte Feucht, et al. Cable-driven parallel robot for curtain wall module installation. *Automation in Construction*, 138:104235, 2022.

[6] Tobias Bruckmann, Christopher Reichert, et al. Concept studies of automated construction using cable-driven parallel robots. In Clément Gosselin, Philippe Cardou, Tobias Bruckmann, and Andreas Pott, editors, *Cable-Driven Parallel Robots*, pages 364–375, Cham, 2018. Springer International Publishing. ISBN 978-3-319-61431-1.

[7] Paul Bosscher, Robert L. Williams, et al. Cable-suspended robotic contour crafting system. *Automation in Construction*, 17(1):45–55, 2007. ISSN 0926-5805.

[8] Sen Qian, Changqi Li, et al. Modeling and experimental validation of a cable-driven robot for three-dimensional printing construction. *Journal of Mechanisms and Robotics*, 17(6), 2025.

[9] Andreas Pott, Christian Meyer, and Alexander Verl. Large-scale assembly of solar power plants with parallel cable robots. In *ISR 2010 (41st International Symposium on Robotics) and ROBOTIK 2010 (6th German Conference on Robotics)*, pages 1–6. VDE, 2010.

[10] Kevin Murphy, Joao CV Soares, et al. Cooperative modular manipulation with numerous cable-driven robots for assistive construction and gap crossing. *arXiv preprint arXiv:2403.13124*, 2024.

[11] Jean-Baptiste Izard, Marc Gouttefarde, et al. *Integration of a Parallel Cable-Driven Robot on an Existing Building Façade*, pages 149–164. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013. ISBN 978-3-642-31988-4.

[12] R. C. Hibbeler. *Engineering Mechanics Dynamics*. Pearson, New Jersey, 11th edition, 2007. ISBN 0-13-221504-7.

[13] Robert McGhee, Eric Robert Bachmann, et al. Rigid body dynamics, inertial reference frames, and graphics coordinate systems: A resolution of conflicting conventions and terminology. 2000.

[14] Andreas Pott and Tobias Bruckmann. *Cable-driven parallel robots*. Springer, 2013.

[15] Joel A E Andersson, Joris Gillis, et al. CasADi – A software framework for nonlinear optimization and optimal control. *Mathematical Programming Computation*, 11(1):1–36, 2019.

Semantic-based behavioral decision-making method for construction robots in worker-robot collaboration

Xiao Lin[1], Hongling Guo[1,*], Ziyang Guo[1]

[1]Department of Construction Management, Tsinghua University, Beijing, China
linxiaoshaw@foxmail.com, hlguo@tsinghua.edu.cn, guo-zy24@mails.tsinghua.edu.cn

Abstract –

Worker-robot collaboration (WRC) can enhance the efficiency and performance of construction robots in complex tasks and dynamic environments. The behavioral decision-making capabilities of robots are critical for them to effectively interact and collaborate with workers. This research presents a behavioral decision-making method for automatic construction robots. First, a semantic world model is introduced to represent the on-site WRC state with graph data structures. Then, a Learning from Demonstration (LfD) method is utilized for robots to acquire behavioral policy through worker demonstrations. Finally, behavioral decision-making for construction robots is achieved using graph-matching algorithms. The proposed method was tested and validated in a virtual reality environment through panel installation tasks. The results show a 90% success rate in extracting feasible and reproducible robot behavioral policies on a single demonstration. The proposed method is promising for enhancing the intelligence of construction robots and improving the safety and efficiency of on-site WRC.

Keywords –

Worker-robot collaboration; Construction robot; Behavioral decision-making; Semantic understanding

1 Introduction

Construction robots can address urgent issues in the construction industry, including low productivity, worker shortages, and occupational health concerns. However, the ever-changing nature of construction sites presents significant challenges for the effective deployment of construction robots on-site [1]. Worker-robot collaboration (WRC) can enable robots to adapt to dynamic construction environments by introducing workers' flexibility to help robots overcome the complexity and unpredictability of construction sites [2].

Recent studies are focusing on on-site WRC with automated construction robots collaborating with workers as teammates. These collaboration tasks include common activities such as welding [3], wall painting [4], and wooden structure assembling [5]. In WRC, construction robot needs to decide what to do and when to act to ensure both safety and the smooth progress of the collaboration. However, most construction robots in existing research rely on direct commands from workers and lack the flexibility and intelligence to boost productivity [6].

To enhance the efficiency and safety of WRC, it is essential for construction robots to possess behavioral decision-making capabilities. This involves selecting appropriate actions from their built-in skills based on the situation they perceive themselves to be in, which is crucial in dynamic environments [7]. Nevertheless, several challenges remain unaddressed for construction robots: first, how to establish a structured representation of the WRC situation; and second, how to map the WRC situation to appropriate behaviours.

The objective of this research is to propose a behavioral decision-making method for construction robots in on-site WRC. A semantic world model is proposed for WRC's scene understanding. Based on the model, behavioral policies can be learned from demonstrations and embedded into construction robots' knowledge base, ultimately supporting their behavioral decision-making. The remainder of this paper is structured as follows. A literature review is first made in Section 2. Then, the proposed method is introduced in Section 3, and experiments and results are discussed in Section 4. Finally, a conclusion is drawn in Section 5.

2 Literature Review

2.1 World model of robots in WRC

A world model refers to the information the robot has about the world around it, and that needs to be shared between several activities [8]. A well-defined world model is not only crucial for the robot to understand the scene and make behavioral decisions but also enhances its explainability [9].

Syntax-based world model interprets a scene as a sequence of events driven by the behaviors of agents. For instance, Pan et al. [10] developed a scene understanding method for construction robots utilizing syntax to outline

construction progress such as “Base setup” and “Framework setup.” Markov models are used to describe state transitions. Cunha et al. [11] adopted predicate symbolic representations to describe scenes and used a short-term memory layer to model state transitions.

A semantic world model analyzes a scene by leveraging the semantic relationships among entities to deduce the context. Niccolò et al. [12] modeled the assembly task with semantic descriptions, where assembly tasks are defined by the spatial relationships among components. This model was expanded into an enriched event chain model to describe state transition processes[13]. A semantic world model can also be used to guide robots in performing safe operations [14]. Markov processes and other stochastic models are used to describe state transitions within a semantic world model [15]. The semantic world model significantly outperforms the syntax-based world model in terms of reusability and scalability. However, existing semantic world models cannot characterize elements unique to the on-site WRC environment. Moreover, the transition path representations in current world models fail to account for the modeling of the concurrent operations of multiple agents, making them unsuitable for supporting construction robots' decision-making in WRC.

2.2 Behavioral policy for robots in WRC

A policy is a set of rules mapping the optimal behavior to the current state [15]. If an explicit world model is not provided, reinforcement learning (RL) is applied, allowing the robot to learn optimal policy through repeated interactions with the environment, guided by a defined reward function [16]. Nevertheless, designing a reward function for WRC is challenging, as it must account for factors related to both efficiency and safety. Additionally, RL-based decision-making is difficult to explain and may disrupt workers in WRC.

In scenarios where the decision-making process cannot easily be scripted or optimized, robots can learn the decision-making process from human demonstration [17]. The policy learned from the demonstration can be either probabilistic or deterministic. Probabilistic policy learning methods, such as Markov networks [18] and Bayesian networks, output a distribution of behaviors rather than deterministic actions. However, when demonstration data is limited, robots can only learn a deterministic behavioral policy. In these cases, the robot's policy can be represented using a finite state machine or a decision tree [19]. However, previous research rarely considers robot policies in WRC. On one hand, most world models are not designed for WRC scenarios that involve concurrent behaviors from multiple agents, making them unsuitable for supporting the robot's decision-making in on-site WRC. The other challenge is how to acquire demonstration data and learn behavioral policies from limited demonstration data for construction robots. These are the two central questions that need to be addressed in this research.

3 Method

The proposed robot’s behavioral decision-making method begins with the introduction of a semantic world model to represent the on-site WRC process. The world model abstracts unstructured WRC scenes and processes into structured data, enabling robots to establish semantic-based scene understanding and policy learning foundations. To facilitate behavioral policy learning from demonstrations, a virtual reality (VR) platform is developed to simulate the WRC environment, where workers demonstrate the robot’s behaviors through intuitive interaction. The behavioral policy is subsequently extracted from the demonstration data structured by the world model. Using the policy learned from the demonstration, construction robots can make behavioral decisions by matching their observation of the current WRC scene to the WRC state stored in the policy, thereby reasoning and determining their behaviors accordingly. The framework of the method is illustrated in Figure 1.

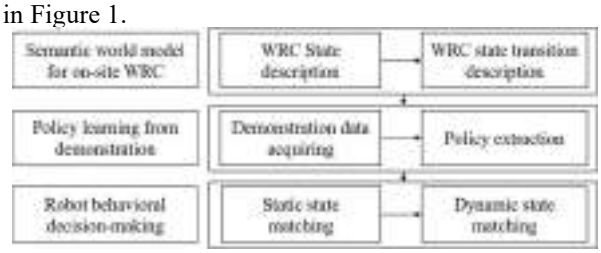


Figure 1. The framework of the proposed method.

3.1 Semantic world model for on-site WRC

3.1.1 Description of WRC state

The state of a static WRC scene is semantically described by the entities and their relationships within the environment. The entities involved in on-site WRC are identified and classified into three categories:

- **Agents**: Agents can proactively interact with other entities and drive the progress of on-site WRC, including workers and construction robots.
- **Movable objects**: Movable objects are entities that can be manipulated and relocated, including construction materials, tools, containers, etc.
- **Static Objects**: Static objects are entities that cannot be moved by agents or other objects, such as walls and floors.

Spatial relationships and kinematic relationships are defined to describe interactions between entities. A spatial relationship exists between entities that are in

spatial contact, whereas a kinematic relationship denotes movement constraints between entities (e.g., a tool being held by a worker). Since multiple semantic relationships can coexist between the same entities, a multi-graph is employed to model the static scene as $G_s = (V_s, E_s)$, where the nodes V_s represent the entities and the edges E_s represent the semantic relationships. An example of the semantic state description of a WRC scene is showcased in Figure 2, where the robot is holding a tube while the worker is touching it.

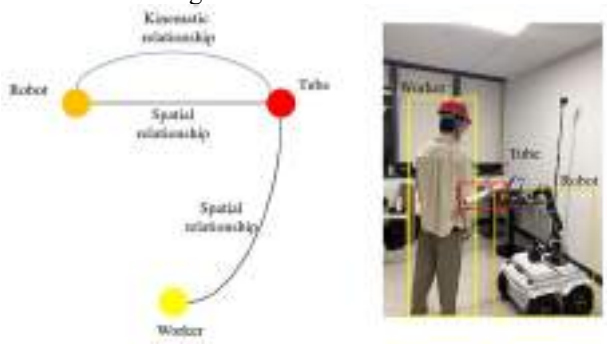


Figure 2. The semantic description of a WRC scene.

3.1.2 Description of WRC state transition

The description of WRC state transitions defines the evolution rules from one state to the next in the WRC process. The extended semantic event chain (ESEC) is introduced to describe the WRC process with concurrent agents' behaviors. ESEC models a WRC process as a quatrain (S, B, V, T), where:

- S is the set of semantic graphs representing WRC states.
- B is the set of agents' behaviors.
- V is the set of variables to be monitored during the WRC process.
- T is the set of transitions among WRC states within the WRC process.

A transition is defined by four parameters, denoted as (s, b, g, s'), where:

- s is the semantic graph before the transition.
- b is the behavior involved in the transition.
- g is the guardian condition, which determines whether the behavior can be executed.
- s' is the semantic graph after the transition.

To model a WRC process using ESEC, the WRC states in the process are serialized with an initial state and a final state identified as s_s and s_e respectively. Except for the final state, each state should have at least one associated transition to the next state. The end of a WRC process is signified by reaching the final state.

3.2 Behavioral policy learning from demonstration

A robot's policy is a mapping function that projects the state to the behavior, denoted as $s \rightarrow b$. This mapping relationship is explicitly defined in the ESEC model as a transition (s, b, g, s'). To learn the behavioral policy $s \rightarrow b$ from collaboration demonstration samples, the demonstration process is first abstracted using ESEC model to significantly reduces the state space of WRC. The behavioral policy is then extracted from the structured demonstration data.

3.2.1 Demonstration data acquiring

In the demonstration, the construction robot is controlled by a human operator, who invokes the robot's built-in behavior interfaces to collaborate with other workers. Essentially, the robot needs to replicate the behaviour policy adopted by the operator.

VR is considered a valuable tool for acquiring demonstration data due to its advantages in safety, cost-efficiency, and customization. This research creates a VR-based demonstration platform. The Unity engine is utilized to create a WRC environment with realistic physical effects. The OpenXR framework enables workers to move and interact with objects freely using controllers in the VR environment. The virtual construction robot is controlled by the Robot Operating System (ROS). Movable objects and static objects are modelled with necessary interaction functions to respond to the agents' behaviors. Semantic relationships are modeled according to their definition: a contact between two bounding boxes indicates a spatial relationship between the corresponding entities, while kinematic relationships are modeled with fixed joints between objects. The creation and deletion of the semantic relationships are controlled by specific events triggered by agents' behaviors.

During the collaboration demonstration, the following data will be recorded frame by frame: 1) entities and their relationships in the VR environment, 2) monitored variables, and 3) the virtual robot's behavior invocation events. The data will be stored along with timestamps in JSON format.

3.2.2 Policy extraction

To extract the policy from the demonstration data, the demonstration data is first processed into serialized key semantic graphs. Initially, entities and relationships in each frame are processed into semantic graphs. In two consecutive frames, if the semantic map of the latter frame undergoes any changes, it is extracted as a key semantic map. It should be noted that during the demonstration, the WRC state may undergo a rollback: for instance, a worker might pick up an object and immediately put it back down. The rollback state will

then be integrated with the historical state. Finally, the key semantic graphs are serialized into a tree structure based on their temporal order, as shown in Algorithm 1.

Algorithm 1 Key semantic graphs serializing

Input: $list_G = [G_0, G_1, G_2, \dots, G_n]$
  Initiate: $node_c = G_0$
  foreach G in $list_g$ **do**
    if $G = node_c$ **then**
      continue
    else if $G = node_c.parent$ **then**
      $node_c = node_c.parent$
    else
      if $\exists child = G, G \in node_c.children$ **then**
        $node_c = child$
      else
        add G to $node_c.children$
        $node_c = G$
      end if
    end if
  end if
Output: Serialize key semantic graphs

Furthermore, the transitions are defined based on the invoke records of the robot's behaviors to create an ESEC model. For each robot behavior b, the timestamp at which the behavior begins is used to find the preceding key semantic graph s and the succeeding key semantic graph s'. If multiple robot behaviours correspond to the same preceding and succeeding key semantic graph, this indicates that these behaviours are adjacent and can be considered as one combined behaviour. Lastly, guardian conditions are manually set, typically considering the WRC safety requirements [20]. The collection of all transitions $\{t_r\}$ that include robot behaviors constitutes the robot's behavioral policy.

3.3 Robot behavioral decision-making

With the policy learned from the demonstration, the construction robot needs to recognize the WRC state to map its behavior. Due to sensor limitations and the dynamic movement of agents and movable objects, entities and their semantic relationships may not be fully recognized. To address this issue, this section develops static and dynamic state-matching algorithms to match incomplete observation results into the WRC state.

3.3.1 Static state matching

To simplify the discussion, several assumptions about the robot's observations are made as follows:

- The entities presented in the ESEC model are perceptible.
- The observation result for an entity is binary: the entity is either observed or not.
- Some semantic relationships between entities are perceptible (e.g., all spatial relationships), referred to as R_p; the remaining semantic relationships are undetectable due to the limitations of the robot's perception method, denoted as R_u.
- The observation result for a semantic relationship is binary: the relationship is either observed or not.

The above assumptions are made based on the current state of robotic perception technologies and methods, including computer vision, LiDAR, force sensors, etc. For a given WRC state s, its semantic graph is denoted as $G_s = (V_s, E_s)$, and the corresponding observation result is represented as $G_o = (V_o, E_o)$. Based on these assumptions, the observed graph is reasoned to be a subgraph of the actual WRC state semantic graph, i.e. $V_o \in V_s, E_o \in E_s$. The observed graph is then matched to all candidate WRC state graphs from the ESEC model according to Algorithm 2.

Algorithm 2 Static graph matching

Input: $G_0 = (V_o, E_o)$, $\{S\}$
  Initiate: $\{S_m\} = \emptyset$
  foreach $s \in S$ **do**
    $G_s = (V_s, E_s)$
    match_flag = True
    foreach $v_o^i \in V_o, v_o^j \in V_o, e^{ij} = (v_o^i, v_o^j)$ **do**
      if ($e^{ij} \in E_s$ **and** $e^{ij} \in E_o$) **or**
      ($e^{ij} \notin E_s$ **and** $e^{ij} \notin E_o$) **or**
      ($e^{ij} \in E_s$ **and** $e^{ij} \notin E_o$ **and** $e^{ij} \in R_u$) **then**
        continue
      else
        match_flag = False
        break
      end if
    end for
    if match_flag = True **then**
      add s to S_m
    else
      pass
    endif
  end for
Output: Candidate matched states $\{S_m\}$

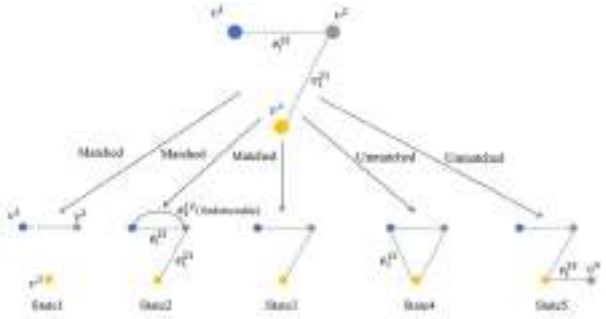


Figure 3. The static state matching process.

3.3.2 Dynamic state matching and behavior decision-making

The construction robot can reason from its historical observations to further narrow down the candidate WRC states. For two semantic graphs G_1 and G_2, their semantic difference is defined as the symmetry difference of their edge sets, i.e. $D(G_1, G_2) = E_1 \triangle E_2$ Consider two adjusted observed semantic graphs $G_o^{i-1} = (V_o^{i-1}, E_o^{i-1})$ and $G_o^i = (V_o^i, E_o^i)$, their static state match results are S_m^{i-1} and S_m^i. The following conditions must be satisfied for matching a state pair s^{i-1}, s^i to the observation pair G_o^{i-1}, G_o^i:

- $s^{i-1} \in S_m^{i-1}, s^i \in S_m^i$;
- $D(G_o^{i-1}, G_o^i) = D\left(G(s^{i-1}), G(s^i)\right)$;
- A transition exists from s^{i-1} to s^i.

Figure 4 depicts a dynamic state match example, where the observed semantic graph G_o^i is statically matched to s_2 and s_5, and G_o^{i-1} is statically matched to s_3 and s_4. According to the dynamic match rules, G_o^i is only dynamically matched to s_5 since no transition exists from s_3 or s_4 to s_2. If the dynamic matching results filter out all static matching results, dynamic matching will not be performed in the next keyframe.

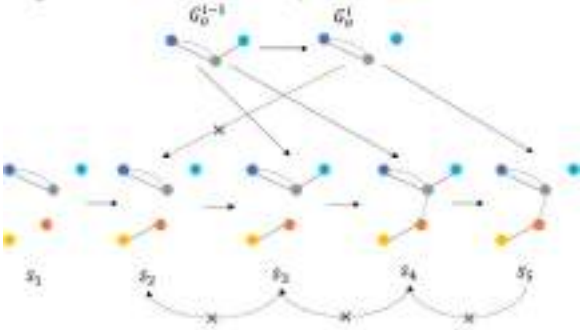


Figure 4. The dynamic state matching process.

Finally, the construction robot decides its behavior based on the matched possible WRC states and its policy. If only one WRC state is matched, or if multiple WRC states are matched but mapped to the same behavior, that behavior will be the decision-making result. Otherwise, if the candidate WRC states map to multiple behaviors, the default waiting behavior is adopted to avoid any risks.

4 Experiment and Results

An experiment was conducted to validate the feasibility of the proposed method. A typical WRC scenario was simulated in a VR environment, replicating a wall panel installation task. Wall panel installation is a common and representative task in decoration construction, which requires the worker to move the panel to the designated place on the wall and fix it using nails. Adopting a WRC method, the construction robot acts as a collaborative teammate, responsible for moving and positioning the panel to reduce manual labour and mitigate safety risks.

The experiment scene is illustrated in Figure 5. The panels and a nail gun are placed on a worktable. Workers can pick up these items and use the tools using VR controllers. The construction robot employed in the experiment is configured as a classic mobile manipulator with a suction cup serving as its end-effector. The construction robot has three built-in skills: 1) controlling the suction cup to attach or release the panel, 2) moving the robotic arm to a designated configuration, and 3) driving the chassis to a designated location. The installation place is assumed to be known by the robot.



Figure 5. The experimental setup.

The experiment consists of two phases. In the first phase, a demonstration experiment is conducted, where the construction robot's behavior is controlled by an operator. In the second phase, the construction robot collaborates automatically with the worker using the behavioral policy learned from the demonstration to perform an identical task.

10 subjects were invited to participate in both the demonstration experiment and the collaboration experiment, while an additional 20 subjects were invited to participate in the collaboration experiment. None of them had prior experience in collaborating with construction robots. Before the formal experiment, the subjects were given the opportunity to practice in the VR environment to ensure they understood the collaboration objectives and could use the VR devices correctly. After the experiment, the participants were interviewed to gather their feedback on the demonstration process and their experience in collaboration with the robot.

4.1 Demonstration results

In the demonstration experiment, each subject was required to perform two demonstrations to ensure that sufficient demonstration data was collected. Collectively, the 10 subjects contributed to 20 demonstration processes. The general WRC workflow is as follows: the subject

first grabs the panel and hands it over to the robot. The robot then moves the panel to the designated installation spot on the wall. Once in place, the subject uses a nail gun to secure the panel with four nails, one at each corner.

Even though the process appears intuitive, subjects still collaborated with the robot in a variety of ways, leading to significant diversity in the number of states within the ESEC models. The ESEC models of these demonstrations contained up to 28 WRC states and at least 14 WRC states. For example, some subjects chose to grasp both the panel and the nail gun simultaneously to avoid repeatedly picking up the nail gun. During the fixing stage, some subjects nailed all four nails before commanding the robot to release the panel, while others nailed two nails first and secured the remaining two after the robot released the panel and moved away. While either approach is feasible to complete the task, this flexibility poses a challenge for the construction robot to understand the variant WRC process.

In contrast to the diversity observed in collaboration, the construction robot consistently exhibits four distinct behaviors across various collaborative demonstrations. These four behaviors are defined as grasp, position, release, and reposition. Using the ESEC model, these behaviors and their corresponding prerequisite WRC states can be extracted. While the strategies derived from different demonstration data vary, a representative example is analyzed below and depicted in Figure 6.

- Grasp: Grasping creates a kinematics relationship between the robot and the panel. The pre-state of grasp requires the spatial relationship between worker and panel, robot and panel, and a kinematics relationship between worker and panel, which means that the worker is holding the panel while it is touching the robot.
- Position: Positioning includes moving the robotic arm to the installation posture and moving the chassis to the installation position. The pre-state of positioning requires a kinematic relationship between the robot and the panel, but no spatial or kinematic relationship exists between the worker and the panel.
- Release: Releasing removes the kinematic relationship between the robot and the panel. In most demonstrations, the pre-state of releasing requires four nails establishing a kinematic relationship with the panel and the work surface.
- Reposition: Repositioning includes moving the chassis to the receiving position and moving the robotic arm to the receiving posture. The repositioning behavior typically follows adjacently the releasing behavior.

The feasibility of demonstrated behavioral policies will be further validated in the collaboration experiment.

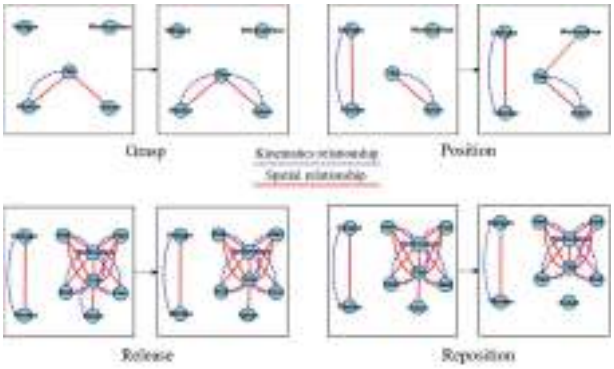


Figure 6. The behavioral policy of the construction robot learned from the demonstration.

4.2 Collaboration results

The collaboration tasks in the experiments were scaled from installing one single panel to four panels to test the robustness of the demonstrated behavioral policy. In this setup, the robot autonomously collaborates with subjects based on its observations and behavioral policies demonstrated. To simulate the construction robot's observation, a virtual camera is deployed. The camera's view follows the robot's end-effector, as illustrated in Figure 7. Entities not within the field of view along with their associated relationships will be ignored. A simulated force sensor provides information on whether a panel is attached to the robot's end-effector, indicating the kinematic relationship between the robot and the panel.

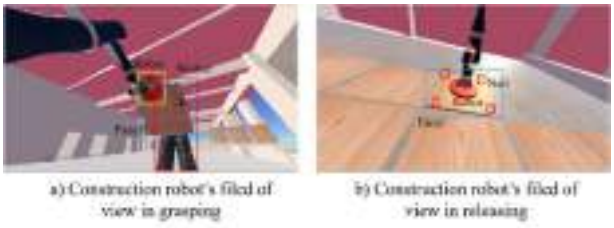


Figure 7. Construction robot's field of view in collaboration.

The 10 subjects who participated in the demonstration experiments were once again invited to participate in the collaboration experiment to validate the feasibility of the policies they demonstrated. The results of the collaboration experiment revealed that 18 out of the 20 policies were feasible, enabling the robot to successfully collaborate with the workers. However, in the remaining two unfeasible policies, the prerequisite states for some of the robot's behaviors were not correctly represented in the semantic graph. For instance, a subject demonstrated the positioning behavior while remaining close to the robot, resulting in an erroneous spatial relationship

captured in the semantic graph as the prerequisite state for the behaivor. This inconsistency hindered the robot's ability to align its observations with the appropriate WRC state during the collaboration experiments.

Excluding the two unfeasible policies, among the remaining 18 collaborations with feasible policies, the tasks were successfully completed on the first attempt in 16 instances, with the two failures resulting from human error. From these feasible policies, a representative behavior policy was adopted by the robot to collaborate with the additional 20 subjects. The success rate on the first attempt was 70%, lower than the 89% observed among the 10 subjects who had participated in the demonstration experiments. Through repeated trials, all subjects ultimately completed the tasks successfully.

Since the ESEC model is explicit, the robot's behavioral decision-making process is transparent. While the construction robot's observation is often incomplete due to limitations in its field of view and the lack of detection methods for most kinematic relationships, static and dynamic matching enable the robot to match its observation to the correct WRC states. Static matching first matches observed semantics with candidate WRC states, while dynamic matching further filters the matched states. Figure 8 illustrates the behavioral decision-making process during a collaboration.

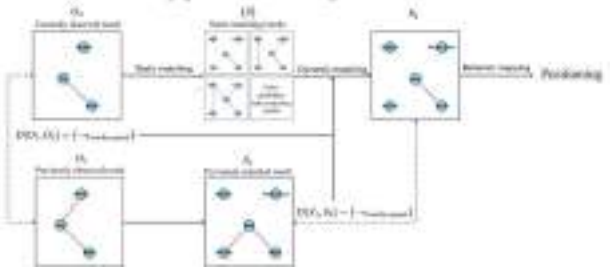


Figure 8. The behavior decision-making process in the collaboration experiment.

4.3 Discussion

The experiments utilize VR environments in both demonstration and collaboration experiments, simulating a complex and dynamic WRC task, wall panel installation. This task involves both spatial and temporal interactions between the worker and the robot, with the workflow being highly diverse and dependent on the strategies employed by the collaborating worker.

To validate the effectiveness of the proposed method, a series of measures have been implemented to ensure that the VR environments replicate real-world WRC conditions as closely as possible, including 1) utilizing the same control logic for the virtual robot as its real-world counterparts, 2) immersive VR perspectives and physical interaction logic for the participants, and 3) virtual sensors that emulate the functions and limitations of real sensors. The primary distinction from the real world lies in the simplification of the robot's perception process. In reality, the robot's detection methods (e.g. computer vision) may not always be accurate, potentially leading to mismatches between observations and correct states. In the VR environment, however, the robot's observations are always accurate. While it is reasonable to expect that the robot's observations will be accurate in most cases, establishing error recovery and correction mechanisms for behavioral decision-making may be essential for real-world construction robots.

The experiments also emphasize the flexibility of the proposed method in on-site WRC applications. In most cases, the participants need only perform a single successful demonstration to construct the robot's behavioral policy. Although the behavioral policy learned from demonstrations is task- and robot-specific, as workflows and built-in skills may vary across WRC tasks and robots, the ESEC model itself is task- and robot-agnostic. This enables the demonstration and learning process to be generalized to various construction works (such as bricklaying and welding) and construction robots (humanoid robots, etc.), allowing for the rapid establishment of corresponding behavioral policies in WRC.

The demonstration itself also provides additional benefits to on-site WRC. First, all subjects reported that the VR-based demonstration was intuitive and user-friendly. Second, subjects who engaged in the demonstration indicated that it helped them better understand the robot's behavior and intentions, thereby enhancing their sense of safety and performance in the collaboration experiments. This suggests that the proposed method holds promise for improving both safety perception and performance when applying on-site WRC.

5 Conclusion

This research proposes a behavioral decision-making method for constriction robots in on-site WRC. The proposed method comprises three main components: a semantic world model for on-site WRC, behavioral policy learning from demonstration, and a robot's behavior decision-making algorithm. The validation of the proposed framework was conducted in a virtual reality environment, exemplified by a panel installation task. The experimental results demonstrate that the framework effectively supports the robot's comprehension of complex and dynamic WRC processes, enabling accurate behavioral decision-making during autonomous collaboration with workers. Additionally, the results show that the demonstration process enhances workers' understanding of the robot's behavior and increases their sense of safety during collaboration.

The limitations of this research are summarized as follows. Firstly, the proposed method is only validated in a VR environment with simulated robots and sensors, where the impact of perception on robot decision-making may have been underestimated. Future research will focus on testing the decision-making method in real-world settings to assess the influence of environmental disturbances. Secondly, the framework currently lacks support for collaborative recovery in the event of unforeseen and erroneous WRC states. Future research can aim to integrate recovery mechanisms into the robot's policy. Finally, the experiments only involve a single worker and one robot. While the framework theoretically supports collaboration among multiple agents, the complexity of collaboration increases exponentially with the number of entities involved in the WRC process. Therefore, it is vital to validate the framework's accuracy and operational efficiency in on-site WRC scenarios involving more workers or robots.

Acknowledgement

We would like to thank the National Natural Science Foundation of China (Grant No. 52278310) for supporting this research.

References

C.-J. Liang, V.R. Kamat, C.C. Menassa. Teaching robots to perform quasi-repetitive construction tasks through human demonstration. *Automation In Construction*, 120: 103370, 2020. doi:10.1016/j.autcon.2020.103370.

[2] D. Mukherjee, K. Gupta, L.H. Chang, H. Najjaran. A survey of robot learning strategies for human-robot collaboration in industrial settings. *Robotics and Computer-Integrated Manufacturing*, 73: 102231, 2022. doi:10.1016/j.rcim.2021.102231.

[3] P. Tavares, C.M. Costa, L. Rocha, P. Malaca, P. Costa, A.P. Moreira, A. Sousa, G. Veiga. Collaborative welding system using BIM for robotic reprogramming and spatial augmented reality. *Automation in Construction*, 106: 102825, 2019. doi:10.1016/j.autcon.2019.04.020.

[4] E. Asadi, B. Li, I.-M. Chen. Pictobot: A cooperative painting robot for interior finishing of industrial developments. *IEEE Robotics & Automation Magazine*, 25 (2): 82–94, 2018. doi:10.1109/MRA.2018.2816972.

[5] D. Mitterberger, L. Atanasova, K. Dörfler, F. Gramazio, M. Kohler. Tie a knot: human–robot cooperative workflow for assembling wooden structures using rope joints. *Construction Robotics*, 6 (3): 277–292, 2022. doi:10.1007/s41693-022-00083-2.

[6] M. Zhang, R. Xu, H. Wu, J. Pan, X. Luo. Human–robot collaboration for on-site construction. *Automation in Construction*, 150: 104812, 2023. doi:10.1016/j.autcon.2023.104812.

[7] S.J. Baltrusch, F. Krause, A.W. de Vries, W. van Dijk, M.P. de Looze. What about the human in human robot collaboration?: A literature review on HRC's effects on aspects of job quality. *Ergonomics*, 65 (5): 719–740, 2022. doi:10.1080/00140139.2021.1984585.

[8] H. Bruyninckx. Building blocks for the design of complicated systems featuring situational awareness — Herman Bruyninckx, KU Leuven – TU Eindhoven. 2022. . On-line: https://robmosys.pages.gitlab.kuleuven.be/, Accessed: 12/20/2024.

[9] O. Watkins, S. Huang, J. Frost, K. Bhatia, E. Weiner, P. Abbeel, T. Darrell, B. Plummer, K. Saenko, A. Dragan. Explaining robot policies. *Applied AI Letters*, 2 (4): e52, 2021. doi:10.1002/ail2.52.

[10] Z. Pan, Y. Yu. Learning multi-granular worker intentions from incomplete visual observations for worker-robot collaboration in construction. *Automation in Construction*, 158: 105184, 2024. doi:10.1016/j.autcon.2023.105184.

[11] A. Cunha, F. Ferreira, E. Sousa, L. Louro, P. Vicente, S. Monteiro, W. Erlhagen, E. Bicho. Towards collaborative robots as intelligent co-workers in human-robot joint tasks: what to do and who does it? In *ISR 2020; 52th International Symposium on Robotics*, pages 1–8, 2020.

[12] N. Lucci, A. Monguzzi, A.M. Zanchettin, P. Rocco. Workflow modelling for human–robot collaborative assembly operations. *Robotics and Computer-Integrated Manufacturing*, 78: 102384, 2022. doi:10.1016/j.rcim.2022.102384.

[13] Recognition and prediction of manipulation actions using Enriched Semantic Event Chains. *Robotics and Autonomous Systems*, 110: 173–188, 2018. doi:10.1016/j.robot.2018.10.005.

[14] L. Brunke, Y. Zhang, R. Römer, J. Naimer, N. Staykov, S. Zhou, A.P. Schoellig. Semantically safe robot manipulation: from semantic scene understanding to motion safeguards. *arXiv preprint*, 2024. doi:10.48550/arXiv.2410.15185.

[15] M.J.-Y. Chung, A.L. Friesen, D. Fox, A.N. Meltzoff, R.P.N. Rao. A bayesian developmental

approach to robotic goal-based imitation learning. *Plos One*, 10 (11): e0141965, 2015. doi:10.1371/journal.pone.0141965.

[16] N. Akalin, A. Loutfi. Reinforcement learning approaches in social robotics. *Sensors*, 21 (4): 1292, 2021. doi:10.3390/s21041292.

[17] H. Wu, H. Li, X. Fang, X. Luo. A survey on teaching workplace skills to construction robots. *Expert Systems With Applications*, 205: 2022. doi:10.1016/j.eswa.2022.117658.

[18] M. Toussaint, T. Munzer, Y. Mollard, Li Yang Wu, Ngo Anh Vien, M. Lopes. Relational activity processes for modeling concurrent cooperation. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5505–5511, Stockholm, Sweden, 2016. doi:10.1109/ICRA.2016.7487765.

[19] K. French, S. Wu, T. Pan, Z. Zhou, O.C. Jenkins. Learning behavior trees from demonstration. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 7791–7797, 2019. doi:10.1109/ICRA.2019.8794104.

[20] International Organization for Standardization. Robots and robotic devices — Collaborative robots. 2016. . On-line: https://www.iso.org/standard/62996.html, Accessed: 11/01/2023.

Neural Network-Based Estimation of Rebar Deflection for Robotic Rebar Installation of Vertical Structures

Hongjie Cai[1,2], Rongbo Hu[2], Ser Tong Quek[1], Soungho Chae[2], and Justin K.W. Yeoh[1]

[1]Dept. of Civil and Environmental Engineering, National University of Singapore, Singapore
[2]Kajima Technical Research Institute Singapore, Kajima Corp, Singapore

h.cai@u.nus.edu , r.hu@kajima.com.sg, ceeqst@nus.edu.sg, s.chae@kajima.com.sg, justinyeoh@nus.edu.sg

Abstract -

In construction, most existing research on rebar placement focuses on guardrail system for large-scale planar slabs or multi robot arm system for off-site rebar cage assembly, frequently utilizing gantry systems. However, their substantial setup time and spatial requirements greatly reduce their suitability for smaller-scale or more spatially constrained settings. This limitation is particularly pronounced in on-site application where the rebar installation of vertical structural elements occurs in fragmented and restricted workspaces, limiting the applicability of existing guardrails or multiple robotic systems. To address this gap, this study investigates the feasibility of employing a single arm robotic system to grasp and position rebar for vertical structural components. One critical challenge in this context is to account for rebar deflection during rebar placement operations. Moreover, existing analytical equations do not provide sufficiently accurate deflection estimates. To address this problem, a neural network model was developed to predict rebar deflection across a range of sizes and lengths, allowing the robot to adapt its position dynamically according to the target location. The proposed model achieved an R^2 accuracy of 0.9714 and outperformed other models in both the ±0.01m and ±0.092m thresholds, demonstrating its effectiveness in providing precise deflection estimates. This level of precision facilitates efficient robotic rebar placement using a single-arm system.

Keywords -

Construction robotics; Rebar Installation; Rebar Deflection; Neural Networks

1 Introduction

The construction industry has long been recognized as a labor intensive and inherently hazardous field, characterized by physically demanding tasks, challenging work environments, and a high risk of injuries [1]. Over the years, researchers and innovators have directed significant efforts toward addressing these challenges through the development and deployment of construction robots. Such technologies promise to mitigate labor shortages, enhance workplace safety, and boost operational efficiency.

However, despite these advancements, the widespread integration of construction robots remains limited, particularly in structural work, an essential component often found on the project's critical path. Reinforcement bar (rebar) placement, specifically, is a crucial structural activity directly influencing the structural integrity, durability, and safety of reinforced concrete structures. Manual rebar placement can be labor-intensive, prone to errors, and subject to variable quality. Automating rebar placement presents significant opportunities for improving industry-wide efficiency, precision, and safety, marking it as a high-priority area for innovation.

Most existing research has focused on horizontal structures such as planar slabs and precast panels, while vertical structures have received far less attention. Yet vertical components which are on-off or low-repetition require in-situ casting, such as soil-retaining walls and water tanks that common in housing and infrastructure project. For infrastructure projects in particular, most of vertical elements are also cast in situ, because most of them are designed as one-off components. Moreover, Many existing robotic solutions for rebar placement are complex and space-intensive (e.g. gantry with multi-arm), requiring extensive setup and wide, unobstructed work areas that typical construction sites cannot provide. To address these shortcomings, this study investigates the feasibility of a single-arm robotic system for rebar placement in vertical structures.

In horizontal assemblies, rebar deflection is limited to the lifting phase and is naturally resolved upon placement. However, for vertical structures, deflection remains a critical factor that must be addressed throughout the placement process. Existing research primarily focuses on deflection in reinforced concrete or steel beams, with no attention given to the deflection behavior of standalone rebars. Furthermore, existing formulas are inadequate for accurately predicting rebar deflection due to the irregular shape of rebar. Therefore, the primary objective of this study is to identify an effective method to estimate deflection under cantilever conditions, providing a foundation for single-arm robotic rebar placement in vertical structural components.

2 Literature Review

2.1 Robotic or automation for rebar placement

Some rebar palcement system have been explored and developed over the past few decades, for example the IronBot [2] was used for bridge deck rebar installation and the multi-manipulator gantry system described by Mahdi Mo-

meni et al. [3] for prefabricated rebar cages, are effective in controlled factory or open-site environments. However, their extensive setup time, large footprint, and need for clear working envelopes limit their practicality on most housing and many infrastructure projects, where sites are fragmented and space constrained

Additionally, Zhang et al. [4] use a off-the-shelf kuka robot arm for tasks like rebar picking, transportation, and placement. However, the study used short rebars for demonstration, allowing deflection to be ignored, and was limited to planar mesh construction. Dörfler et al.[5] presented the Mesh Mould system, an innovative mobile robotic setup capable of fabricating vertical freeform structural wall on site. However, the system relied on short steel wire segments rather than the long rebars conventionally employed in typical construction practices.

In summary, existing automated rebar placement systems primarily focus on planar slab construction, often using short rebars that do not account for deflection. Currently, no research explores the use of a single-arm robotic system for long rebar placement in vertical structures.

2.2 Estimation of deflection of beam

The Euler–Bernoulli beam theory is a fundamental method for analyzing beam deflection. It assumes linear elastic behavior, small deflections, and negligible shear deformation, making it suitable for long, slender beams under relatively small loads [6]. The relationship between bending moment and curvature is expressed in Equation 1:

$$M = EI\frac{d^2w}{dx^2} \tag{1}$$

where M is the bending moment, E is Young's modulus, I is the moment of inertia, and w is the deflection.

Elementary beam theory neglects the square of the first derivative in the curvature formula, making it unsuitable for scenarios involving large deflections. In contrast, large deflection theory considers both initial and deformed configurations of the beam axis [7], capturing phenomena like significant rotational effects in flexible structures. Due to the mathematical complexity of these nonlinear equations, analytical solutions are often impractical, and numerical methods—such as Finite Element Analysis (FEA) tools like ANSYS and ABAQUS—are widely used. For instance, Yuan et al. [8] applied numerical methods to analyze steel beams with octagon-web designs, achieving improved accuracy in deflection predictions.

While beam deflection theories are well-studied in structural engineering, most research focuses on composite structures like reinforced concrete or steel beams with larger moments of inertia. However, there is a notable lack of investigation into the deflection behavior of rebar as an individual structural element. This gap highlights the need for focused research in this area.

3 Methods

3.1 One-arm robot for rebar placement

In this study, a single robotic arm is used for rebar placement, with an RGB-D camera detecting the position and intersection of the rebar. Figure 1 shows two cases of rebar gripping and placement: in the top diagram, the robotic gripper holds the rebar at its center, causing the ends to bend downward; in the bottom diagram, the gripper holds the rebar at one end, allowing the opposite end to deflect freely. The dashed red line indicates the target position, and the solid green line represents the actual deformation when gripped.

The workflow is outlined in Figure 2. After receiving the rebar's target size and location, the robot retrieves the rebar and moves it to the intended position. Using vision-based detection, the system calculates the error between the actual and target positions. If the error exceeds 10 mm, feedback is sent to adjust the rebar's position until it falls within the fluorescence zone, ensuring accurate placement.

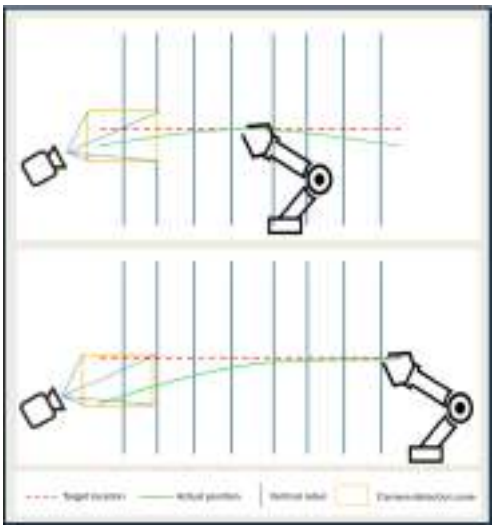


Figure 1. Two scenarios of one robot arm rebar placement system

Accurate deflection estimation is crucial for two reasons: 1) keeping the intersection point within the camera's detection range ensures proper operation, and 2) improved estimation increases the likelihood of errors being within the 10 mm threshold, reducing repositioning.

The Intel RealSense D435 camera, with a horizontal FOV of 69° and vertical FOV of 42° is used [9]. The detection zone can be calculated using Equation 2 [10]:

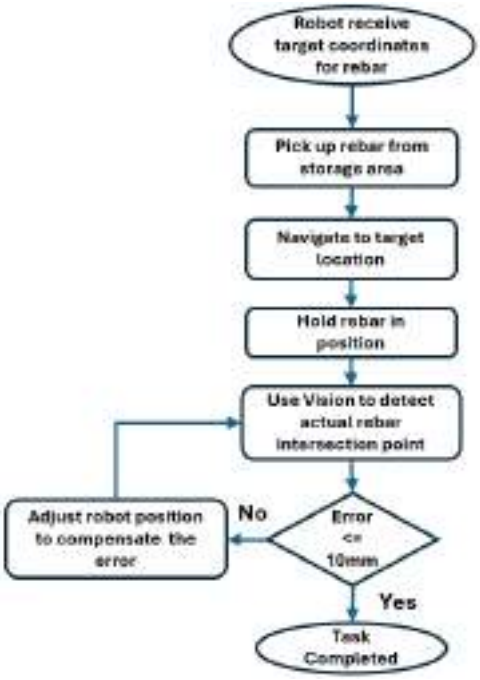


Figure 2. Flow chart for rebar placement

$$\text{Dimension} = 2 \times \text{Distance} \times \tan\left(\frac{\text{FOV}}{2}\right) \quad (2)$$

Where dimension is either Width (using Horizontal FOV) or Height (using Vertical FOV).

For example, as shown in Figure 3, positioning the camera 0.30 m from the target yields a viewing window of roughly 0.41 m × 0.23 m. After restricting analysis to the central 80 % region of interest (ROI)[9], permitting detection of deflections up to about 0.092 m below the target binding point; larger sag will lie outside this effective window. Moving the camera farther back can enlarge the coverage area but at the cost of potential loss of image fidelity [11].

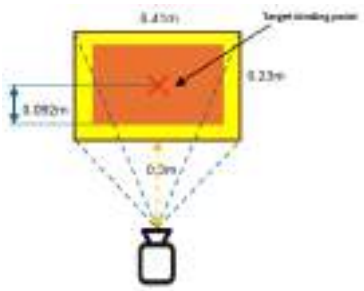


Figure 3. Illustration of camera detection zone

While no specific tolerance for rebar spacing is stated in Eurocode or Singapore Standards, common practice in Singapore allows a tolerance of ±10mm from the designed spacing. In the United States, according to ACI 117-10 [12], the tolerance for wall thicknesses between 4 to 12 inches is ±3/8 inch (approximately 10mm). Thus, ensuring the rebar deflection estimation remains within 10mm of the target location could significantly minimize the need for repeated repositioning of the robot gripper and improve operational efficiency.

3.2 Estimation of rebar deflection

In this study, four different models were developed to predict deflection behavior: two theoretical models—the Euler-Bernoulli equation and large deflection theorem conducted in ANSYS, and two data-driven models—a nonlinear regression model and a neural network model—built using empirical data from experimental testing.

Model accuracy was judged by the coefficient of determination (R^2), which measures the proportion of variance in the dependent variable explained by the independent variables, and the mean absolute error (MAE), the average magnitude of prediction errors. Additionally, two threshold - ±0.092 (camera detetction zone) and ±0.01 m (construction tolerance) were also applied to verify the performance of the predictions stayed within practical limits.

For the purposes of this study, it was assumed that the rebar is firmly grasped by the robotic gripper, simulating a cantilever boundary condition. The estimation process utilized two input variables: the length and diameter of the rebar, with the output being the deflection of the rebar. Common material properties were applied across all techniques, including an elasticity modulus of 200 GPa, a density of 7850kg/m3, and a gravitational acceleration of 9.81m/s2.

3.2.1 Theoretical and numerical estimation

Initial deflection estimations were conducted using two primary approaches: the Euler–Bernoulli beam theory and the finite element method (FEM) performed in ANSYS. The Euler–Bernoulli beam theory was first employed, providing a theoretical framework based on assumptions of linear elasticity and small deflection behavior. The maximum deflection $\delta_{\max}$ under self-weight was calculated using the equation 3:

$$\delta_{\max} = \frac{\omega L^4}{8EI} \quad (3)$$

where ω is the distributed load, L is the length of the rebar, E is the modulus of elasticity, I is the moment of inertia.

To address potential geometric non-linearity, the large deflection theorem was applied. This approach considers

large displacements and provides more accurate predictions for long rebars. The large deflection analysis were performed using ANSYS to supplement the theoretical models and provide a computational perspective on deflection behavior (Figure 4).

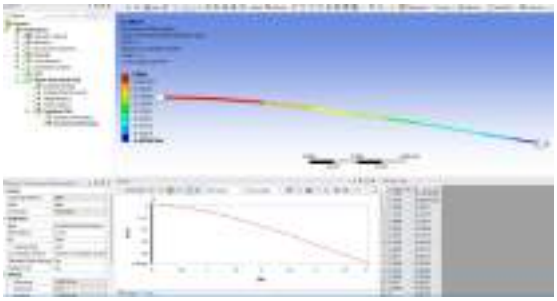


Figure 4. ANSYS analysis of rebar deflection

3.2.2 Experimental data collection

To validate the theoretical predictions and provide data for subsequent nonlinear regression analysis and neural network training, an experimental study was conducted. Rebars of varying lengths and diameters were tested under self-weight in a cantilever setup, where one end was securely fixed to a workbench using two clamps, and the other end was left free to deflect, as illustrated in Figures 5. Deflections at the free end were measured using laser displacement sensors (Leica DISTOTM X4).



Figure 5. Rebar deflection experiment set up

The experiment involved a total of 27 rebars, with diameters ranging from 10mm to 32mm and lengths ranging from 2.5m to 4m. Each rebar was tested multiple times (4-6 trials) at different cantilever lengths, resulting in a total of 154 data points. These measurements served as the foundation for evaluating the accuracy of the theoretical models and for developing data-driven predictive methods.

3.2.3 Training of Nonlinear Regression Model

To predict rebar deflection, a nonlinear regression model was developed using experimental data. The relationship between the key inputs—rebar length (L) and diameter (d) of the rebar—and the deflection $\delta_{\max}$ was modelled using the "`curve_fit`" function from Python's SciPy library. This regression analysis aimed to find the best-fit equation by minimizing the residual error between the predicted and observed values, producing a model that accurately reflects real-world behavior. The dataset was randomly split into 70% for training and 15% for testing using the "`train_test_split`" function from Python's scikit-learn library. The training set and testing set are the same for both nonlinear regression model and neural network model. The remaining 15% is validation set, which is only used in neural network model training.

The equation for maximum deflection is expressed in the general form:

$$\delta_{max} = aL^b d^c \tag{4}$$

where a, b, and c are coefficients of the regression.

Considering the mechanics of deflection, it is influenced by the rebar's self-weight (proportional to its density) and length, while being inversely proportional to material stiffness and cross-sectional properties. Incorporating these relationships, the equation is refined as:

$$\delta_{\max} = e\frac{\rho g L^f}{E d^h} \tag{5}$$

Where ρ is density of the material, g is the gravitational acceleration, E is the modulus of the elasticity. and e, f, and h are coefficients to be determined through regression.

This refined equation links deflection to fundamental physical parameters, ensuring the model aligns with theoretical expectations. By determining the values of e, f, and h through regression analysis, providing a practical equation for estimating rebar deflection under self-weight.

3.2.4 Training of Neural Network Model

The predictive model for estimating deflection employs a neural network architecture designed to capture the relationship between two primary input features: Diameter and Length. As shown in Figure 6, the network consists of a single hidden layer with 256 neurons, utilizing the ReLU activation function to learn complex, non-linear relationships. The final output layer contains a single neuron without an activation function, suitable for regression tasks.

The hyper-parameter tuning process was conducted to optimize the neural network architecture for deflection prediction. Various configurations were evaluated, including single-layer and multi-layer architectures, with a single hidden layer selected for its simplicity and performance. The number of neurons was varied from 64 to 512, with 256 neurons chosen as the optimal configuration. Regularization methods such as L2 regularization (0.01 to 0.0001) and dropout (0.01 to 0.5) were tested but excluded in the final model as they did not improve performance. The Adam optimizer and mean squared error (MSE) loss function ensured efficient convergence.

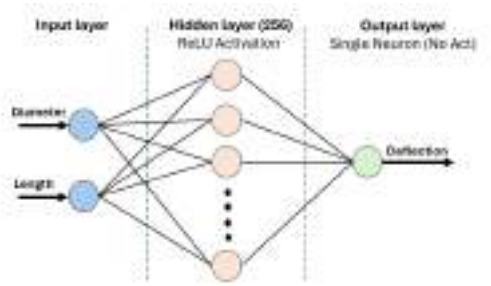


Figure 6. Neural Network Architecture

4 Results

4.1 Results of different models

The graph in Figure 7 illustrates the deflection of seven different rebar diameters under self-weight in a cantilever boundary condition, as predicted by the Euler–Bernoulli equation, FEA (large deflection), non-linear regression model, neural network model, and the actual experimental measurements. For the non-linear regression and neural network models, data from only the first trial run are presented in the graph. Trend lines are included for each model to improve visualization and facilitate a clearer comparison of accuracy among the four predictive models.

In addition, comprehensive evaluations were conducted across various diameters and length ranges using all available data points. Metrics including R^2, MAE, and the percentages of predictions falling within two accuracy thresholds (±0.092 m and ±0.01 m) were computed for each method—the Euler-Bernoulli equation, the non-linear regression model, and the neural network model. These results are summarized in Tables 1 and 2, with the best-performing method in each category highlighted in bold. However, the FEA was excluded from this detailed analysis due to the significant computational time required to run simulations. Running simulations for all 154 combinations of diameters and lengths to calculate comparable values was deemed impractical.

For both the non-linear regression and neural network models, training, validation, and testing datasets were randomly assigned. To ensure robustness, additional five trial runs were conducted for each model, using identical datasets across both models in each run. This approach also helps to verify that if the result in Table 1 was representative of the overall models' performance. The data presented here is derived from the testing dataset, providing a more reliable basis for comparison between the non-linear regression and neural network models. A detailed summary of these results is provided in Table 3.

4.2 Regression equation of non-linear model

To develop the regression model, parameters e, f, and h were estimated for each trial, and the ressult were listed on Table 4. These parameters define the model's ability to relate the input variables to the target output. The average values of these parameters were used to construct the regression equation, which is expressed as Equation 6.

$$\delta_{\max} = 10\frac{\rho g L^{3.61}}{E d^{1.75}} \tag{6}$$

4.3 Model fitting of neural network

Figure 8 shows the training and validation loss curves, which are closely aligned, indicating that the model generalizes well to unseen data. Both loss curves drop sharply

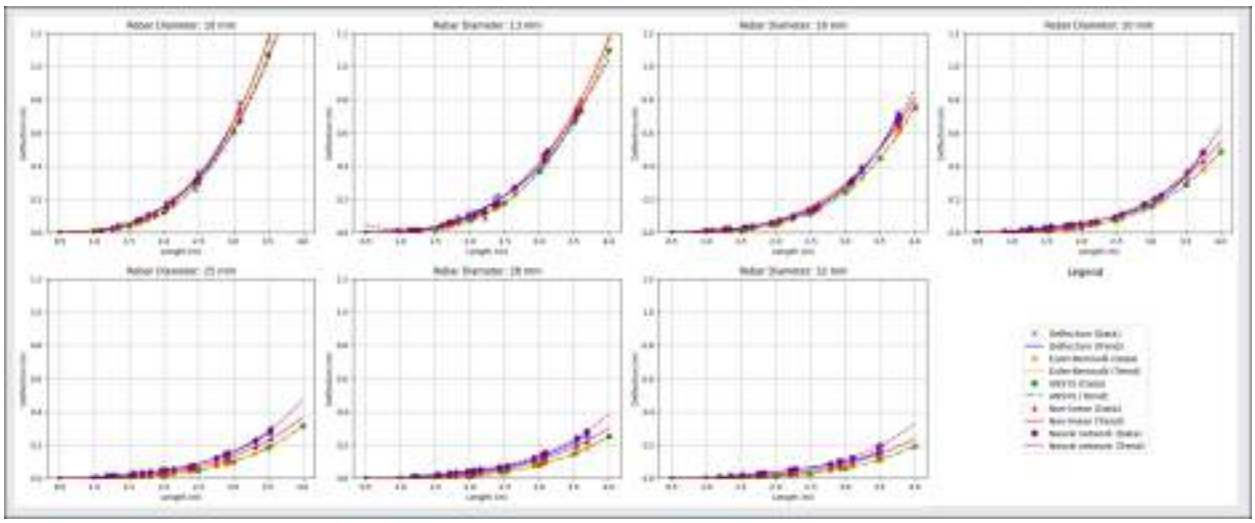


Figure 7. Trend-line comparison between measured deflection and predictions from four estimation methods

Table 1. Comparison among different models for different diameters.

Diameter (mm)	R^2			MAE (m)			±0.01 m(%)			±0.092 m(%)		
	Euler-Bernoulli	Non-linear	Neural-network	Euler-Bernoulli	Non-linear	Neural-network	Euler-Bernoulli	Non-linear	Neural-network	Euler-Bernoulli	Non-linear	Neural-network
10	0.9779	**0.9895**	0.9805	0.0252	**0.0164**	0.0208	31.25	**56.25**	**56.25**	93.75	**100**	93.75
13	0.9843	0.9850	**0.9874**	0.0264	0.0238	**0.0231**	21.74	**26.09**	**26.09**	**100**	**100**	**100**
16	0.9647	0.9911	**0.9966**	0.0294	0.0147	**0.0099**	18.18	63.64	**68.18**	86.36	**100**	**100**
20	0.9285	0.9810	**0.9834**	0.0261	0.0142	**0.0124**	18.18	**54.55**	45.45	90.91	**100**	**100**
25	0.7630	0.9417	**0.9894**	0.0318	0.0168	**0.0066**	13.64	22.73	**77.27**	90.91	**100**	**100**
28	0.7847	**0.9629**	0.9613	0.0289	0.0130	**0.0125**	9.09	40.91	**45.45**	**100**	**100**	**100**
32	0.6772	**0.9341**	0.8789	0.0260	**0.0113**	0.0135	22.22	50.00	**61.11**	93.26	**100**	**100**

Table 2. Comparison among different models for different lengths.

Length (m)	R^2			MAE (m)			±0.01 m(%)			±0.092 m(%)		
	Euler-Bernoulli	Non-linear	Neural-network	Euler-Bernoulli	Non-linear	Neural-network	Euler-Bernoulli	Non-linear	Neural-network	Euler-Bernoulli	Non-linear	Neural-network
0.5m-1.5m	-0.3522	0.1260	**0.4395**	0.0082	0.0065	**0.0052**	54.05	75.68	**83.78**	**100**	**100**	**100**
1.5m-2m	0.3510	0.7459	**0.8991**	0.0202	0.0128	**0.0071**	15.15	39.39	**75.76**	**100**	**100**	**100**
2m-2.5m	0.8045	0.9225	**0.9304**	0.0308	0.0166	**0.0142**	0.00	40.00	**50.00**	**100**	**100**	**100**
2.5m-3m	0.3228	0.8406	**0.9673**	0.0354	0.0162	**0.0070**	7.69	30.77	**84.62**	**100**	**100**	**100**
3m-3.5m	0.9597	**0.9852**	0.9572	0.0353	**0.0195**	0.0330	6.67	**26.67**	13.33	93.33	**100**	93.33
3.5m-4m	0.8884	0.9644	**0.9833**	0.0678	0.0379	**0.0251**	0.00	**17.65**	**17.65**	47.07	**100**	**100**

Table 3. Five trial run results of non-linear and neural-network models.

Trial Run	TestingR^2		Testing MAE (m)		±0.01 m (%)		±0.092 m (%)	
	Non-linear	Neural network	Non-linear	Neural network	Non-linear	Neural network	Non-linear	Neural network
1	0.9434	**0.9739**	0.0158	**0.0119**	33.33	**66.67**	**100**	**100**
2	**0.9675**	0.9638	0.0196	**0.0168**	20.83	**58.33**	**100**	**100**
3	0.9784	**0.9943**	0.0154	**0.0082**	33.33	**62.50**	**100**	**100**
4	0.9812	**0.9815**	0.0248	**0.0195**	25.00	**41.67**	95.83	**95.83**
5	**0.9789**	0.9433	0.0168	**0.0138**	29.17	**41.67**	95.83	**100**
Average	0.96988	**0.97136**	0.01848	**0.01404**	28.33	**54.17**	98.33	**99.17**

Table 4. Values of non-linear regression parameters

	e	f	h
1	10.0	3.6779	1.7355
2	10.0	3.5625	1.7632
3	10.0	3.6695	1.7386
4	10.0	3.6038	1.7548
5	10.0	3.5535	1.7692
Average	10.0	3.61344	1.75226

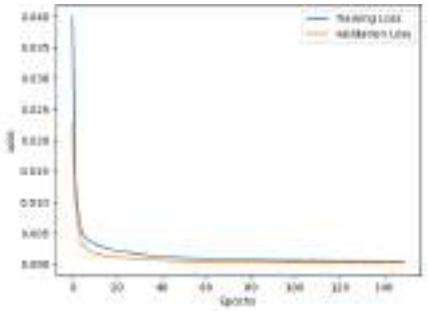


Figure 8. training vs validation loss

in the initial epochs and subsequently plateau, demonstrating that the network has effectively learned the underlying data patterns without exhibiting signs of overfitting or underfitting.

5 Discussion

5.1 Analysis and comparison of model results

Figure 7 shows that both the Euler–Bernoulli and large deflection models closely match the actual deflection values for shorter rebar lengths. However, as the rebar length increases, these models increasingly underestimate deflections, resulting in conservative predictions.

In contrast, the non-linear regression and neural-network models both sustain high predictive quality over the full span of rebar diameters and lengths: their trendlines track the measured deflections closely, confirming their superior accuracy and robustness in modelling deflection behaviour for a wide range of geometries.

Detailed numerical comparisons are presented in Tables 1 and 2. Both the neural network and non-linear regression models generally achieved higher R^2 values than the Euler–Bernoulli equation across every diameters and lengths. Diameter-wise, the non-linear regression model

has a slight edge over the neural network for several sizes, whereas length-wise the neural network leads in all categories except the 3–3.5 m interval.

In terms of MAE, the neural network consistently produced the lowest values across both diameter and length ranges, , except for the 3-3.5 m, indicating its superior absolute accuracy. The non-linear regression model also consistently performed better than the Euler–Bernoulli equation, highlighting the inherent limitations of linear-elastic assumptions for predicting complex deflection behaviors.

Most importantly, when evaluating performance based on practical accuracy thresholds (±0.01 m and ±0.092 m)—a key focus of this study—the neural network and non-linear regression models clearly outperformed the Euler–Bernoulli equation. Under the stringent ±0.01 m criterion, the neural network delivers the highest within-tolerance rate for almost all lengths and diameters. With the more lenient ±0.092 m threshold, both data-driven models achieve near-perfect accuracy, far surpassing Euler–Bernoulli—particularly for the longer bars.

A notable discrepancy is observed for parameters of 10 mm diameter and 3–3.5 m length, where the neural network model underperforms relative to the nonlinear regression model, despite its otherwise comparable or superior performance. This exception is likely a training-specific artefact: the particular network realisation appears to have missed one or two points in that band. To test whether this weakness is systematic, we retrained both models five times on identical, randomly drawn training–testing splits and compared with non-linear regression model with same testing set in each run.

As Table 3 shows, across the five runs, each model sustained an average testing R^2 values of approximately 0.97. The neural network further lowered the MAE by about 4.5 mm relative to the non-linear regression. For the broader ±0.092 m threshold, both models achieved similarly high accuracy, with the neural network slightly ahead (99.17%) compared to the non-linear regression model (98.33%). However, for the stricter ±0.01 m threshold, the neural network significantly outperformed the non-linear regression model, achieving nearly double the percentage of accurate predictions (54.17% vs. 28.33%). This substantial difference highlights the neural network model's greater capability in capturing subtle variations in rebar deflection, and the earlier 3–3.5 m anomaly is a model-realisation outlier rather than an inherent limitation.

In summary, both the non-linear regression and neural network models demonstrate superior robustness in predicting rebar deflection and consistently outperform the numerical models in all metrics. However, the neural network model excels in fine-grained precision, making it the most effective approach for accurate rebar deflection estimation.

5.2 Factors influencing rebar deflection

One possible explanation for the increased deflection is the fact that rebar is not perfectly round. The rebar used in the test complies with the BS4449:2005 standard (Grade: B500B) and features transverse and longitudinal ribs on its surface, which are designed to enhance the bonding between the concrete and the rebar [13]. However, these ribs are not accounted for in the estimated mass or cross-sectional area when using the nominal diameter, as the actual core diameter is slightly smaller than the nominal diameter. This discrepancy is illustrated in Figure 9 [14]. The irregular shape is likely a primary factor contributing to the larger deflection observed compared to theoretical predictions based on simulations or calculations assuming a regular, perfectly round steel bar. The irregularities in the shape lead to a non-uniform distribution of material. The effective moment of inertia will likely be lower than that of the nominal circular shape due to the reduced symmetry and possible loss of material at certain points.



Figure 9. Actual cross section of rebar

Additionally, Orientation or gripping point may also influence deflection, but consistent orientation is impractical on site, so bars were randomly oriented for testing. The measured deflection data, presented in Figure 7, were obtained from rebars of varying sizes and orientations. Although some variation exists, the data points consistently follow a similar trend, indicating that orientation is not a critical factor influencing deflection.

In conclusion, irregularities in rebar geometry pose significantly challenges for accurate prediction of deflection using numerical models, as indicated by the consistently large MAE observed across different rebar sizes. These results underscore the necessity for developing more advanced predictive models that effectively account for geometric irregularities.

5.3 Impact to the robot rebar palcement

With the improved accuracy in rebar deflection estimation, the use of a single robotic arm for rebar placement

becomes a viable solution. The higher prediction accuracy increases the likelihood of maintaining deflection within the required tolerance, minimizing the need for repeated intersection detection and position adjustments. This enhancement streamlines construction workflows, improving both efficiency and precision.

Additionally, accurate prediction of rebar deflection plays a critical role in other stages of rebar handling. For example, the motion planning of the robotic arm must consider predicted deflection to avoid collisions with objects during movement. Deflection estimation is also essential when determining the optimal gripping points, ensuring effective handling of the rebar.

6 Conclusion

This study emphasizes that managing rebar deflection during placement is a significant challenge when utilizing single robotic arm systems for rebar placement in vertical structural elements.

It also underscores the limitations of existing theoretical models, which fail to accurately predict standalone rebar deflection under real-world conditions. By integrating experimental data, the development of a neural network model effectively addresses these gaps, achieving nearly 100% likehood of fall within the camera's field of view and over a 50% likelihood of placing the rebar within 10mm of the target location. This advancement provides a more reliable and practical predictive tool for robotic rebar placement applications.

However, certain limitations should be acknowledged. For instance, further research is necessary to comprehensively identify and evaluate factors that lead to larger deflections, including rebar geometry and environmental influences such as corrosion and temperature variations. Moreover, this study is limited to the initial rebar intersection placement, leaving the complexities associated with subsequent intersections unaddressed. Exploring these challenges represents a valuable direction for future work.

References

[1] Pei Pei Heng, Hanizah Mohd Yusoff, and Rozita Hod. Individual evaluation of fatigue at work to enhance the safety performance in the construction industry: A systematic review. *PLoS one*, 19(2):e0287892, 2024.

[2] IronBOT | rebar lifting, carrying, and self-placing robot. URL `https://www.constructionrobots.com/ironbot`.

[3] Mahdi Momeni, Johan Relefors, Abubakar Khatry, Lars Pettersson, Alessandro V Papadopoulos, and Thomas Nolte. Automated fabrication of reinforcement cages using a robotized production cell. *Automation in Construction*, 133:103990, 2022.

[4] Jiansong Zhang, Chen Shen, and Ruilin Li. A robotic system method and rebar construction with off-the-shelf robots. In *Computing in Civil Engineering 2021*, pages 1204–1211. 2021.

[5] Kathrin Dörfler, Norman Hack, Timothy Sandy, Markus Giftthaler, Manuel Lussi, Alexander N Walzer, Jonas Buchli, Fabio Gramazio, and Matthias Kohler. Mobile robotic fabrication beyond factory conditions: Case study mesh mould wall of the dfab house. *Construction robotics*, 3:53–67, 2019.

[6] Oliver A Bauchau and James I Craig. Euler-bernoulli beam theory. In *Structural analysis*, pages 173–221. Springer, 2009.

[7] KE Bisshopp and DC13360 Drucker. Large deflection of cantilever beams. *Quarterly of applied mathematics*, 3(3):272–275, 1945.

[8] Ling Yuan, Bi Sun, Lu Ma, Weidong Wu, et al. Research on factors affecting deflection and an approach for evaluating web stress for steel octagon-web beam. *Heliyon*, 9(7), 2023.

[9] Intel realsense camera 400 series product family datasheet. Datasheet 337029-017, Intel Corporation, March 2024. URL `https://www.intelrealsense.com/wp-content/uploads/2024/10/Intel-RealSense-D400-Series-Datasheet-October-2024.pdf`. Revision 018.

[10] Richard Szeliski. *Computer vision: algorithms and applications*. Springer Nature, 2022.

[11] Chanda Bhabatosh et al. *Digital image processing and analysis*. PHI Learning Pvt. Ltd., 2011.

[12] American Concrete Institute. *Specification for Tolerances for Concrete Construction and Materials (ACI 117-10) and Commentary*. American Concrete Institute, Farmington Hills, MI, 2010.

[13] BSI BS 4449. Steel for the reinforcement of concrete—weldable reinforcing steel—bars, coil and decoiled product—specification. *Standard Policy and Strategy Committee*, pages 1–27, 2005.

[14] Betonstaal. Rebar isn't round at all, n.d. URL `https://www.betonstaal.nl/en/knowledge-centre/rebar-isnt-round-at-all`. Accessed: 2025-01-02.

Mobility-added multi tool end-effector for picking and nailing applications in robotic timber assembly

Mauricio Arredondo-Soto[1], Paolo Pancho-Ramirez [1], and Rafiq Ahmad[1*]

[1]Smart & Sustainable Manufacturing Systems Laboratory (SMART Lab), Department of Mechanical Engineering, University of Alberta, Edmonton, Canada

marredon@ualberta.ca, paoloand@ualberta.ca, rafiq.ahmad@ualberta.ca

Abstract -

Robotic timber assembly applications, such as nailing, often require the simultaneous execution of multiple tasks, necessitating the use of multiple devices and large working spaces. This poses a challenge for the development of transportable robotic manufacturing cells, which aim to bring the benefits of controlled environments closer to construction sites. Addressing this issue requires the design of manufacturing cells that optimize robotic resources and space by enabling multiple tasks to be performed simultaneously. In this work, the concept of Mobility-Added Multi Tool (MAMT) end-effector is introduced. Specifically, a MAMT end-effector designed for picking and nailing operations is proposed to eliminate the need for additional external devices during timber nailing processes. This end-effector is based on closed-loop linkages driven by linear actuators, providing ease of control, high force transmission, and structural stability. To demonstrate their practicality, application examples involving two-robot collaboration are presented and animated, illustrating the usability of the proposed end-effector in real-world scenarios. Future work includes the development of linkage kinematics and dynamics, as well as the exploration of alternative mechanisms to introduce additional mobility features, enabling the same robot to perform a broader range of tasks during the timber assembly process.

Keywords -

End-effector; Timber Nailing; Robotic Fabrication; Construction Automation; Linkages; Mechanisms

1 Introduction

Timber assembly plays a vital role in the construction industry. Timber is one of the most widely used construction materials globally and serves as the primary material in countries such as the United States and Canada [1, 2]. The construction industry is gradually transitioning towards greater automation, although at a slower pace compared to other industries [3, 4]. However, significant steps are being made in this direction. Among various construction processes, timber assembly is particularly well-suited for robotic fabrication due to the material's lightweight nature, flexibility, and ease of manipulation.

Currently, robotic timber assembly is predominantly carried out in off-site facilities, where robots are working collaboratively with other robots [5, 6] and automated devices such as stationary disc saws and CNC machines [7, 8, 9] to perform tasks like pick-and-place [10, 11], cutting [12], milling [13], drilling [14], screwing [15], and nailing [16]. These operations facilitate the prefabrication of small individual elements [17, 18] as well as larger products, such as complete building sections and facades [19, 20, 21].

While off-site timber assembly factories are advancing toward full automation, there is an urgent need to develop robotic manufacturing solutions closer to construction sites. Such solutions would retain the advantages of controlled factory-like environments while avoiding logistical challenges, such as transportation. Transportable and reconfigurable manufacturing cells have emerged as a promising solution [22]. To ensure portability and efficiency, these cells must optimize robotic and spatial resources, often relying on just one or two stationary robots equipped with the necessary capabilities to execute the entire manufacturing process, even for large timber products.

A significant challenge in timber assembly lies in the wide range of tasks required to complete a fabrication process. These tasks can lead to delays and increased human intervention for tool and material handling. To address this, additional automated devices have been incorporated into manufacturing cells for robot use [23]. Furthermore, robots have been equipped with the ability to perform multiple tasks by employing tool changers [24] or multi tool end-effectors [25, 26, 7]. While these innovations have improved timber assembly efficiency, they also present challenges. Adding more automated devices increases spatial and control system complexity, while tool changers and multi tool end-effectors often reduce robotic mobility. Crucially, these approaches fail to address the fundamental issue of performing multiple tasks simultaneously with a single robot. In many cases, tools remain fixed relative to the robot's end-effector, relying entirely on robot motion for operation.

Simultaneous task execution often requires multiple devices, such as robots or other automated tools, equal to the number of tasks involved. This is particularly problematic for applications like nailing, where at least two timber elements are joined using a nail, thus requiring the use of up to three devices during a single nail process. As a result, robotic nailing is challenging to implement with one or multiple robots. Currently, robotic nailing often involves a table—sometimes automated and equipped with additional Degrees of Freedom (DoF)—to hold timber elements while a robot performs the nailing operation [22, 27, 28, 29]. However, this approach limits manufacturing capabilities to the table's size and the robot's range of motion. Consequently, most robotic nailing applications are restricted to producing smaller products, such as panels and pallets, and transportable manufacturing cells are similarly limited to fabricating small-sized items.

In this work, the concept of Mobility-Added Multi Tool (MAMT) end-effector is introduced. This end-effector enhances robotic capabilities by providing additional mobility, enabling simultaneous execution of multiple tasks while optimizing resources, space, and time. The proposed MAMT end-effector allows a robot to hold a timber element and perform nailing simultaneously, with the ability to vary the nail's position and orientation. The MAMT end-effector is designed using closed-loop linkages driven by linear actuators, which enable precise adjustments to the nail gun's position and orientation. Additionally, these linkages provide high force transmission and structural integrity, which are characteristics of this type of linkages, meeting the demands of construction materials and processes. This paper describes the design and motion principles of the proposed MAMT end-effector. Furthermore, application examples involving collaborative tasks performed by two robots are presented to demonstrate their usability in real-world scenarios.

The remainder of this paper is structured as follows: Section 2 details the proposed MAMT end-effector and their mobility principles. Section 3 presents three application examples of timber nailing processes using the MAMT end-effector. Finally, concluding remarks are summarized in Section 4.

2 MAMT end-effector description

The general structure of the proposed MAMT end-effector is shown in Figure 1. This design is intended for installation on an industrial serial robot, specifically the KUKA® KR 250 R2700, as illustrated in Figure 1(a). The MAMT end-effector consists of four primary modules, depicted in Figure 1(b):

1. **Gripping module** (Module 1) for gripping timber studs.
2. **Linear translation module** (Module 2) for horizontal nail gun movement.
3. **Rotation module** (Module 3) for angular adjustments.
4. **Height variation module** (Module 4) for vertical positioning and nail gun triggering.

The main objective of this section is to present the MAMT end-effector concept and its operational principles without delving into detailed technical analyses. Synthesis, kinematic, and dynamic studies are beyond the scope of this work; however, key dimensions of the model are provided for reference.

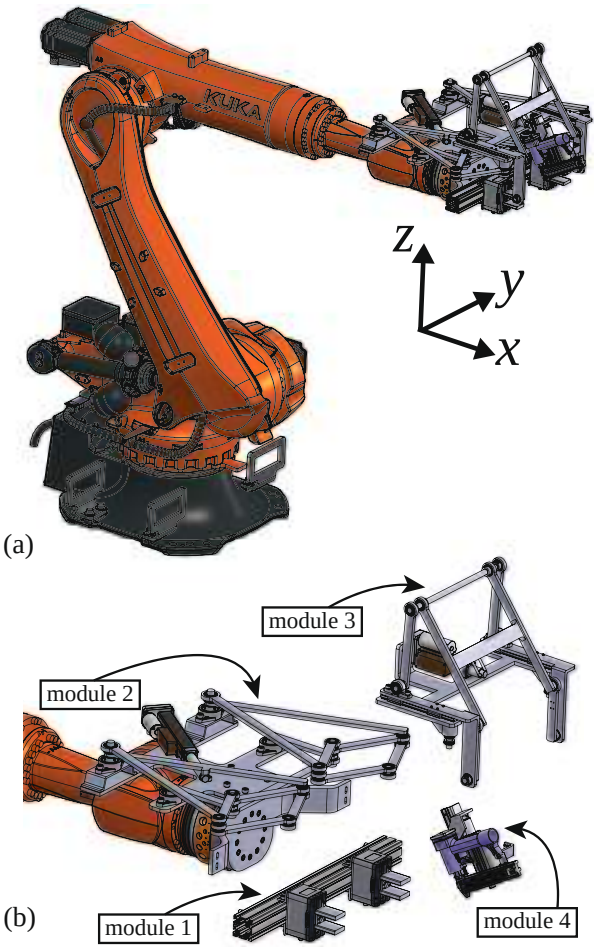


Figure 1. Proposed MAMT end-effector: (a) installed on a robot KUKA® kr 250 r2700, and (b) modules.

2.1 Module 1 - gripping

Module 1, shown in Figure 2, consists of two grippers symmetrically positioned along the end-effector's y-axis. This module enables the end-effector to pick and securely

hold timber studs during the nailing process. The grippers' tips are fixed at a distance w_1 from the center of the robot's final link. Distances h_1 and v_1 can be adjusted by repositioning the grippers. The dimensions used in the computational model are listed in Table 1.

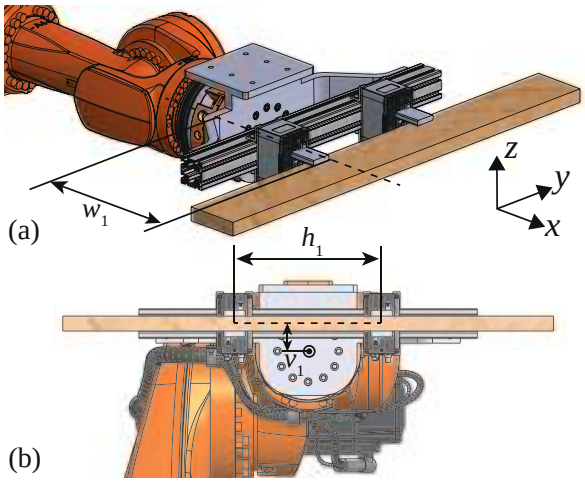


Figure 2. Module 1 - gripping: (a) isometric view, and (b) front view.

Table 1. Dimensions of Module 1 - gripping

Element	Dimension [mm]
h_1	300
v_1	56
w_1	447

2.2 Module 2 - linear translation

Module 2, shown in Figure 3, utilizes a double Peaucellier planar mechanism mounted in a plane parallel to the xy-plane of the robot's end-effector. The Peaucellier mechanism, a classic straight-line generator, comprises eight links and six revolute joints, resulting in one degree of freedom (DoF). Connecting two Peaucellier mechanisms, as shown in Figure 2(a), not only generates straight-line motion but also achieves straight-line body translation. This setup enables the nail gun to move linearly (parallel to the y-axis) over the gripped timber stud without undesired rotations, as illustrated in Figure 3(b).

The entire linkage maintains a single DoF and allowing to be driven by a single linear actuator. The Peaucellier mechanism's link dimensions are listed in Table 2. This double mechanism architecture provides a linear displacement d_1 of 700 mm with an output/input ratio of 3:1, meaning the nail gun moves three times the displacement of the linear actuator.

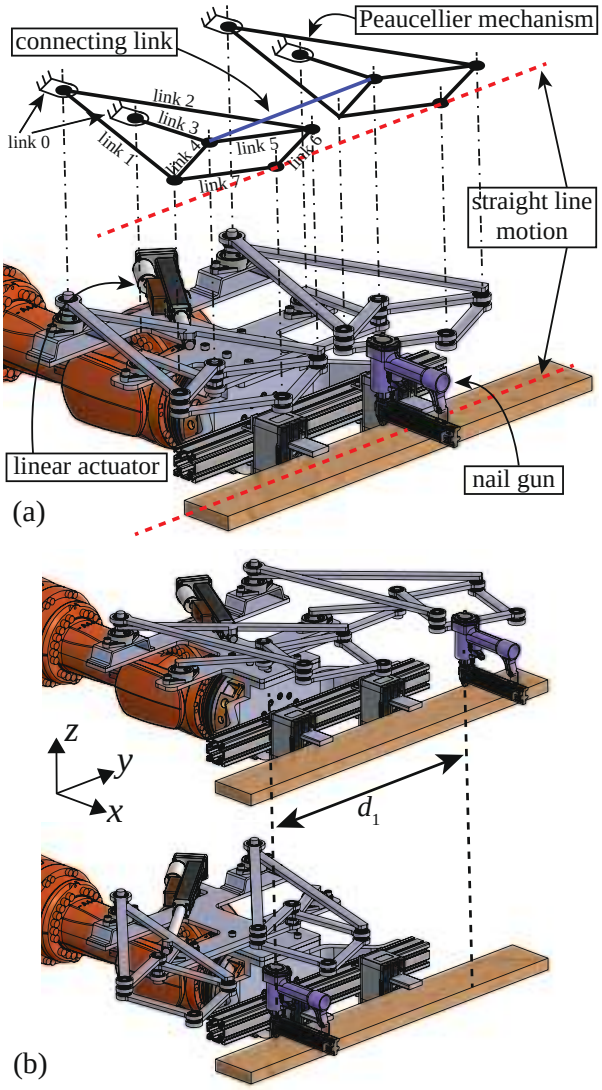


Figure 3. Module 2 - linear translation: (a) base linkage, and (b) displacement range. Modules 3 and 4 that physically connect module 2 with the nail gun are not included for better visualization of the base linkage.

Table 2. Dimensions of Module 2 - linear translation

Element	Dimension [mm]
link 0	200
link 1	530
link 2	530
link 3	200
link 4	210
link 5	210
link 6	210
link 7	210
connecting link	450

2.3 Module 3 - rotation

Module 3, shown in Figure 4, is based on two parallel four-bar planar mechanisms. Two design variants, to be used independently, are proposed:

1. **Variant 1** (Figure 4(a)): Mechanisms placed in planes parallel to the xz-plane of the robot's end-effector.
2. **Variant 2** (Figure 4(b)): Mechanisms placed in planes parallel to the yz-plane.

Each four-bar mechanism comprises four links connected by four revolute joints, resulting in one DoF, making it suitable for actuation by a single linear actuator. In these mechanisms, links 1 and 3 rotate about fixed axes. The nail gun is attached to link 3 in both variants, allowing rotation about axes parallel to the y and x, respectively.

The nail gun tips are positioned along these axes at a distance w_2 from the robot's last link. This distance can be adjusted using carriage and rail systems that enable translation along the x-axis between Modules 2 and 3. The link dimensions of the four-bar mechanism are listed in Table 3.

Module 3 provides a rotation range θ of 50° to 120° for variant 1 and 40° to 140° for variant 2. The rotation range of Variant 1 is shown in Figure 4(c).

Table 3. Dimensions of Module 3 - rotation

Element	Dimension [mm]
link 0	306
link 1	220
link 2	350
link 3	150
w_2	460

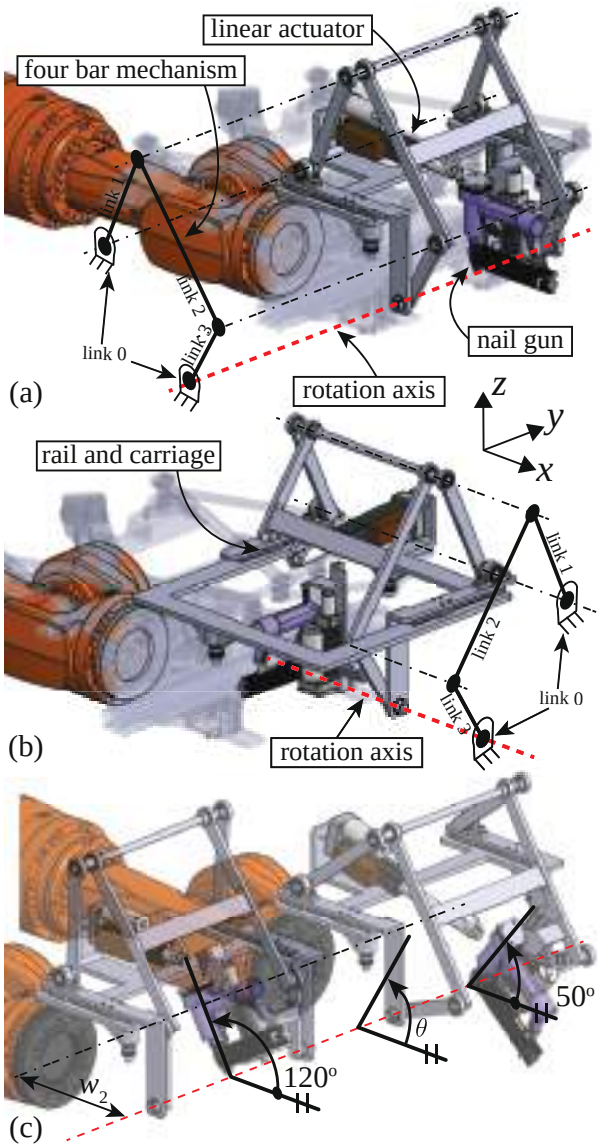


Figure 4. Module 3 - rotation: (a) base linkage of variant 1, (b) base linkage of variant 2, and (c) rotation range.

2.4 Module 4 - height variation

Module 4, shown in Figure 5, includes the nail gun and a linear actuator connected via a linear track. This module enables direct adjustment of the nail gun's distance d_2 relative to the gripped timber stud. In the initial position, as described in Section 2.3, the nail gun tip aligns with the rotation axis of Module 3.

Module 4 is rigidly attached to link 3 of Module 3, ensuring that any displacement produced by Module 4 moves the nail gun tip along an axis intersecting the rotation axis of Module 3. This functionality allows precise positioning of the nail gun tip close to the timber stud surface, regardless of the position and orientation defined by Modules 2 and 3. Additionally, Module 4 can trigger certain nail guns by moving the gun tip until it contacts the timber stud.

3 Application examples

This section provides visual demonstrations of the MAMT end-effector functionality in real timber nailing processes. Two robots in collaboration, each equipped with the proposed MAMT end-effector, were 3D modeled and animated to fabricate three types of timber-nailed structures:

1. **Straight on-plane structure**.
2. **Rotated on-plane structure**.
3. **Rotated off-plane structure**.

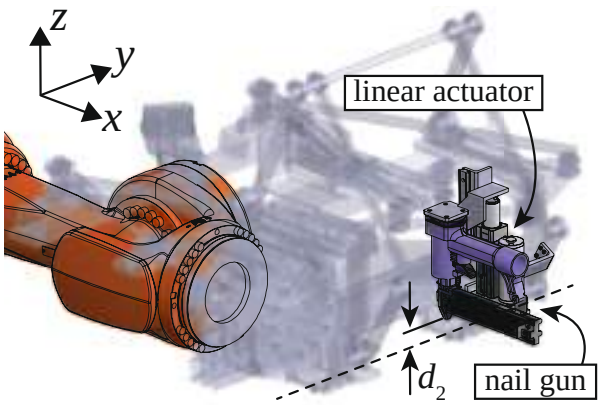


Figure 5. Module 4 - height variation.

The dimensions and configurations of these structures are illustrated in Figure 6.

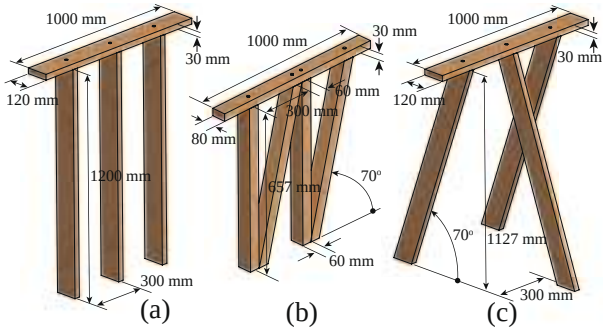


Figure 6. Timber structures: (a) straight on-plane, (b) rotated on-plane, and (c) rotated off-plane.

The fabrication process begins with Robot 1 picking the base timber stud and holding it horizontally at a height of 1797 mm, as shown in Figure 7. Simultaneously, Robot 2 picks the additional timber studs and aligns them with the base stud for subsequent nailing. This process of Robot 1 holding the base stud while Robot 2 places additional studs is repeated for the entire assembly of the three timber structures.

The initial steps of the fabrication processes for the three timber structures are depicted in Figure 8. The procedures for each structure are described as follows:

Straight On-Plane Structure (Figure 8(a)): i) Module 3 fixes the nail gun at 90°, Module 2 translates the nail gun 300 mm along the negative y-axis from the center, and Module 4 lowers the nail gun to join the two timber studs. ii) Module 4 raises the nail gun, Module 2 translates the nail gun 300 mm along the positive y-axis to align with the next stud placed by Robot 2, and the nailing process is repeated.

Rotated On-Plane Structure (Figure 8(b)): i) Module 2 translates the nail gun 300 mm along the negative y-axis, Module 3 (Variation 2) rotates the nail gun to 90°, and Module 4 lowers the nail gun to join the two timber studs. ii) Module 4 raises the nail gun, Module 2 translates the nail gun 300 mm along the positive y-axis to align with the next stud placed by Robot 2, Module 3 rotates the nail gun to 70°, and Module 4 lowers the nail gun to repeat the nailing process.

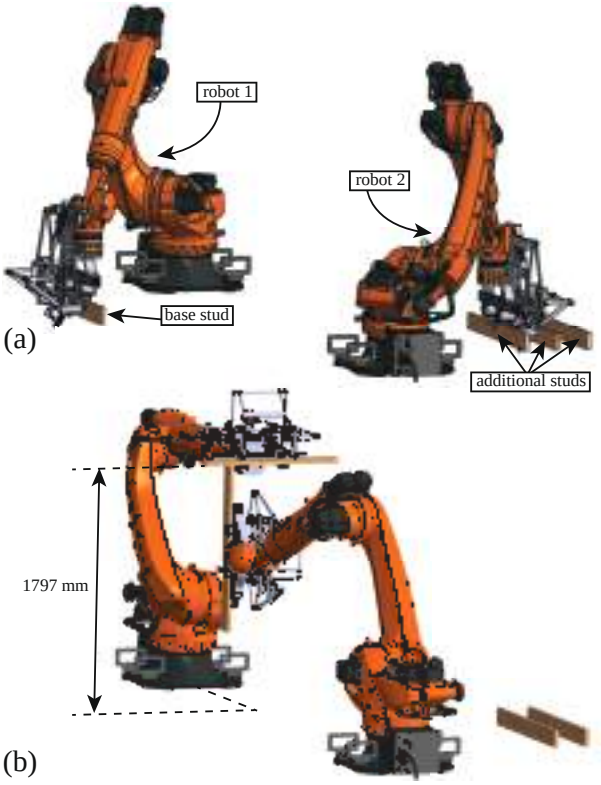


Figure 7. First stage of the timber nailing process with the robot team: (a) picking the timber studs, and (b) placing the timber studs.

Rotated Off-Plane Structure (Figure 8(c)): i) Module 2 translates the nail gun 300 mm along the negative y-axis, Module 3 (Variation 1) rotates the nail gun to 70°, and Module 4 lowers the nail gun to join the two timber studs. ii) Module 4 raises the nail gun, Module 2 translates the nail gun 300 mm along the positive y-axis to align with the next stud placed by Robot 2, Module 3 rotates the nail gun to 110°, and Module 4 lowers the nail gun to repeat the nailing process.

An animated video illustrating the assembly processes for all three timber structures is available at the following link [1].

[1] https://youtu.be/VaK4FC0Dgo4

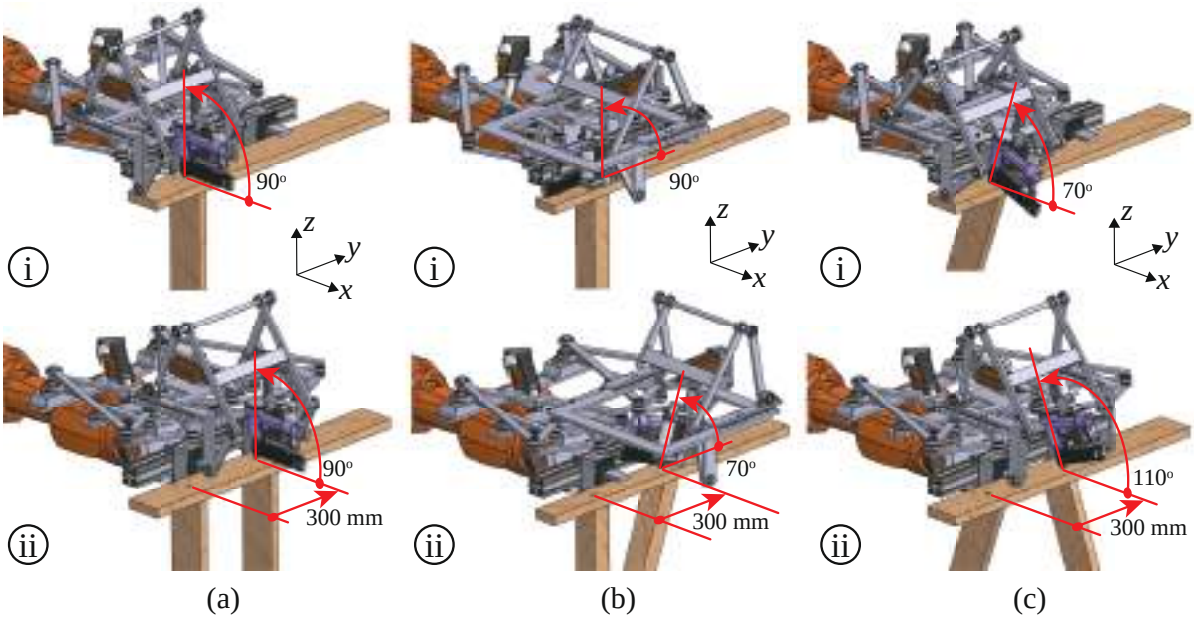


Figure 8. Fabrication processes for the three timber structures: (a) straight on-plane, (b) rotated on-plane, and (c) rotated off-plane. i) stands for first step and ii) for second step.

4 Discussion

Timber assembly is one of the most widely robotized construction processes due to the favorable characteristics of timber, which makes it highly suitable for robotic manipulation. However, specific operations such as nailing remain particularly challenging because of the wide range of tasks involved. This poses a significant obstacle to the development of compact and transportable manufacturing cells that can be deployed near construction sites.

This research introduces the concept of MAMT end-effectors for robotic timber assembly, addressing the need for more efficient and versatile solutions in nailing operations. By enabling robotic systems to perform multiple tasks simultaneously, the proposed MAMT end-effectors optimize space, time, and robotic resources.

The MAMT end-effectors leverage closed-loop linkage mechanisms driven by linear actuators, which provide high force transmission and structural stability essential for timber construction processes. Additionally, the mechanisms are easy to control, making them suitable for seamless integration into industrial robotic systems. The design includes four specialized modules: a gripping module for holding timber studs, a linear translation module for precise straight-line motion, a rotation module for angular adjustments, and a height variation module for accurate positioning and triggering of the nail gun.

The effectiveness of the MAMT end-effectors was illustrated through application examples, where collaborative two robots successfully fabricated three distinct timber-nailed structures: straight on-plane, rotated on-plane, and rotated off-plane. These examples showcased the system's ability to perform complex tasks with minimal reliance on external devices and within compact working spaces.

By eliminating the need for additional external devices during nailing, the MAMT end-effectors enable a more streamlined and efficient approach to robotic fabrication. Their modular and compact design enhances the adaptability and versatility, making them particularly valuable for applications where space and resources are limited. Furthermore, the linkage-based mechanisms ensures precise, robust, and reliable operation, even in demanding construction scenarios.

Some limitations of the proposed MAMT end-effector include the manual pre-setting of parameters h_1 and v_1, which requires prior knowledge of the timber stud dimensions. Additionally, the end-effector does not allow active translation of the nail gun along the x-axis. However, the nail position relative to the timber studs in this direction can be slightly adjusted by varying the portion of the timber stud grasped by the grippers.

While this work presents the feasibility and functionality of the MAMT end-effectors, further research will focus on conducting comprehensive technical analyses, including position synthesis, kinematic and dynamic modeling, and Finite Element Analysis simulations. Fabrication and experimental validation of physical prototypes are also planned as future work. Additionally, exploring alternative mechanisms to enable additional tasks such as cutting

and milling is also planned as part of ongoing research.

5 Conclusion

The concept of MAMT end-effectors represented in this paper goes beyond the additional mobility seen in many existing robotic end-effectors. It emphasizes equipping the end-effector with the necessary capabilities to perform tasks independently of the robot's motion. This independence significantly advances the automation of timber assembly, offering a compact, efficient, and versatile solution for robotic systems. By bridging the gap between off-site prefabrication and on-site manufacturing, the MAMT end-effectors have the potential to address critical challenges in the construction industry, paving the way for more sustainable, scalable, and adaptable processes.

Acknowledgments

We would like to acknowledge the financial support of NSERC and Alberta Innovates for funding this project through NSERC AOO 586056-23 Ahmad & AB In CASBE 232403727 Ahmad.

References

[1] B. Xiao, C. Chen, and X. Yin. Recent advancements of robotics in construction. *Automation in Construction*, 144:104591, 2022. doi:10.1016/j.autcon.2022.104591.

[2] E. Martinez Villanueva, J. A. Cardenas Castañeda, and R. Ahmad. Scientometric analysis for cross-laminated timber in the context of construction 4.0. *Automation*, 3(3):439–470, 2022. doi:10.3390/automation3030023.

[3] Japan Federation of Construction Contractors. *Construction Industry Handbook 2012‘ [Research report]*. Construction Industry Handbook 2012, Tokyo, 2012.

[4] J. J. Cisneros-Gonzalez, A. Rasool, and R. Ahmad. Digital technologies and robotics in mass-timber manufacturing: a systematic literature review on construction 4.0/5.0. *Construction Robotics*, 8(2): 1–26, 2024. doi:10.1007/s41693-024-00143-9.

[5] A. Thoma, D. Jenny, M. Helmreich, A. Gandia, F. Gramazio, and M. Kohler. Cooperative robotic fabrication of timber dowel assemblies. *Research culture in architecture.*, pages 77–88, 2020. doi:10.1515/9783035620238-008.

[6] J. J. González Cisneros, A. Rasool, Y. H. Chui, and R. Ahmad. Robotic cell design and simulation system for cross-laminated timber panel post-construction operations. In *Transforming Construction with Offsite Methods and Technologies (TCOT) Conference*, New Brunswick, Canada, 2024. URL https://conferences.lib.unb.ca/index.php/tcrc/article/view/2006.

[7] P. Eversmann, F. Gramazio, and M. Kohler. Robotic prefabrication of timber structures: towards automated large-scale spatial assembly. *Construction Robotics*, 1:49–60, 2017. doi:10.1007/s41693-017-0006-2.

[8] ABB Robot. Transforming construction with autovol. On-line: https://www.youtube.com/watch?v=9rkLPEWwFv4, Accessed: 08/22/2024.

[9] Homag. Robots in wood processing. On-line: https://www.homag.com/en/machines/automation/robotics, Accessed: 08/22/2024.

[10] A. Adel, D. Ruan, W. McGee, and S. Mozaffari. Feedback-driven adaptive multi-robot timber construction. *Automation in Construction*, 164:105444, 2024. doi:10.1016/j.autcon.2024.105444.

[11] E. P. G. Bruun, S. Adriaenssens, and S. Parascho. Structural rigidity theory applied to the scaffold-free (dis) assembly of space frames using cooperative robotics. *Automation in Construction*, 141:104405, 2022. doi:10.1016/j.autcon.2022.104405.

[12] H. Chai, C. So, and P. F. Yuan. Manufacturing double-curved glulam with robotic band saw cutting technique. *Automation in Construction*, 124:103571, 2021. doi:10.1016/j.autcon.2021.103571.

[13] P. Devadass, F. Dailami, Z. Mollica, and M. Self. Robotic fabrication of non-standard material. In *Annual Conference of the Association of Computer Aided Design in Architecture, CUMINCAD*, pages 206–213, Michigan, USA, 2016. doi:10.52842/conf.acadia.2016.x.g4f.

[14] A. Adel, A. Thoma, M. Helmreich, F. Gramazio, and M. Kohler. Design of robotically fabricated timber frame structures. In *Annual Conference of the Association of Computer Aided Design in Architecture, CUMINCAD*, pages 206–213, Mexico City, Mexico, 2018. doi:10.52842/conf.acadia.2018.394.

[15] R. Naboni, A. Kunic, A. Kramberger, and C. Schlette. Design, simulation and robotic assembly of reversible timber structures. *Construction Robotics*, 5:13–22, 2021. doi:10.1007/s41693-020-00052-7.

[16] H. J. Wagner, M. Alvarez, A. Groenewolt, and A. Menges. Towards digital automation flexibility in large-scale timber construction: integrative robotic prefabrication and co-design of the buga wood pavilion. *Construction Robotics*, 4(3):187–204, 2020. doi:10.1007/s41693-020-00038-5.

[17] M. Fleischmann and A. Menges. Icd/itke research pavilion: A case study of multi-disciplinary collaborative computational design. In C. Gengnagel, A. Kilian, N. Palz, and F. Scheurer, editors, *Computational Design Modelling*, Berlin, Heidelberg, 2011. Springer. doi:10.1007/978-3-642-23435-4_27.

[18] N. M. Larsen and A. K. Aagaard. Exploring natural wood. In *Annual Conference of the Association of Computer Aided Design in Architecture, CUMINCAD*, pages 500–509, Texas, USA, 2019. doi:10.52842/conf.acadia.2019.500.

[19] A. Thoma, A. Adel, M. Helmreich, T. Wehrle, F. Gramazio, and M. Kohler. Robotic fabrication of bespoke timber frame modules. In J. Willmann, P. Block, M. Hutter, K. Byrne, and T. Schork, editors, *J. Willmann, P. Block, M. Hutter, K. Byrne, T. Schork (Eds.), Robotic Fabrication in Architecture, Art and Design 2018 (ROBARCH 2018)*, pages 447–458, Cham. Springer. doi:10.1007/978-3-319-92294-2_2.

[20] CoDeFa. Collaborative design of prefabricated facade systems. On-line: `https://codefacades.ch/`.

[21] E. Martinez Villanueva, P. Martinez, and R. Ahmad. Target-path planning and manufacturability check for robotic clt machining operations from bim information. *Automation in Construction*, 158:105191, 2024. doi:10.1016/j.autcon.2023.105191.

[22] H. J. Wagner, M. Alvarez, O. Kyjanek, Z. Bhiri, M. Buck, and A. Menges. Flexible and transportable robotic timber construction platform–tim. *Automation in Construction*, 120:103400, 2020. doi:10.1016/j.autcon.2020.103400.

[23] J. Willmann, M. Knauss, T. Bonwetsch, A. A. Apolinarska, F. Gramazio, and M. Kohler. Robotic timber construction - expanding additive fabrication to new dimensions. *Automation in construction*, 61:16–23, 2016. doi:10.1016/j.autcon.2015.09.011.

[24] H. Hasan, A. Reddy, and A. Tsayjacobs. Robotic fabrication of nail laminated timber. In M. Al-Hussein, editor, *Proceedings of the 36th International Symposium on Automation and Robotics in Construction (ISARC)*, pages 1210–1216, Banff, Canada, 2019. International Association for Automation and Robotics in Construction (IAARC). doi:10.22260/ISARC2019/0162.

[25] L. Wang, T. Naito, Y. Leng, H. Fukuda, and T. Zhang. Research on construction performance evaluation of robot in wooden structure building method. *Buildings*, 12(9):1437, 2022. doi:10.3390/buildings12091437.

[26] A. Kramberger, A. Kunic, I. Iturrate, C. Sloth, R. Naboni, and C. Schlette. Robotic assembly of timber structures in a human-robot collaboration setup. *Frontiers in Robotics and AI*, 8:768038, 2022. doi:10.3389/frobt.2021.768038.

[27] Ekatech Automation. Ekatech automation - your partner in industrial automation and robotics. On-line: `https://www.ekatech.se/en/`, Accessed: 12/12/2024.

[28] NCR Advanced Robotics. Ncr advanced robotics - your partner in robotics. On-line: `https://www.nc-r.com.pl/en/strona-glowna-en`, Accessed: 12/14/2024.

[29] 7Robortics. Robot pallet nailing. On-line: `https://www.youtube.com/watch?v=lznSq-VgBjY&t=41s`, Accessed: 12/14/2024.

Learning-based 6 DOF Camera Pose Estimation Using BIM-generated Virtual Scene for Facility Management

Thai-Hoa Le[1], Ju-Chi Chang[1], Wei-Yi Hsu[1], Tzu-Yang Lin[2], Ting-wei Chang[2] and Jacob J. Lin[1]

[1]Department of Civil Engineering, National Taiwan University, Taiwan
[2]Lab for Service Robot Systems, Delta Research Center, Taiwan
D10521027@ntu.edu.tw, yuiochang@caece.net, R12521610@ntu.edu.tw,
TZUYANG.LIN@deltaww.com, TINGWEI.CHANG@deltaww.com, jacoblin@ntu.edu.tw

Abstract –
Image-based indoor localization is a promising approach to enhancing facility management efficiency. However, ensuring localization accuracy and improving data accessibility remain key challenges. Therefore, this research aims to automatedly localize images captured during facility inspections by matching the viewpoint of the camera with a corresponding viewpoint in a Building Information Modeling-based (BIM) simulated environment. In this paper, we present a framework that generates photorealistic synthetic images and trains a deep learning model for camera pose estimation. Synthetic datasets are generated in a simulation environment, allowing precise control over scene parameters, camera positions, and lighting conditions. This allows the creation of diverse and realistic training data tailored to specific facility environments. The deep learning model takes RGB images, semantic segmented maps, and corresponding camera poses as inputs to predict six-degree-of-freedom (6DOF) camera poses, including position and orientation. Experimental results demonstrate that the proposed approach can enable indoor image localization with an average translation error of 5.8 meters and a rotation error of 69.05 degrees.

Keywords –
Synthetic data; Image localization; Pose estimation; Facility management; BIM

1 Introduction

In facility management, accurate camera pose estimation is crucial for tasks such as building surveying, indoor navigation, and asset tracking. Within indoor built environments, reliable positioning data is the foundation for numerous location-based services. By determining the precise orientation and location of the camera, facility managers can automate the detection of structural changes, schedule efficient maintenance, and enhance occupant safety. Additionally, camera pose estimation enables advanced applications like augmented reality guidance and remote inspections, ultimately improving the overall efficiency and effectiveness of facility operations.

A traditional approach to getting a camera pose from a given 3D model of the scene is to match 2D image features with their corresponding 3D points using an n-point pose solver in a RANSAC loop [1]. Another common method is reconstructing a 3D scene via Structure-from-Motion (SfM), in which 3D points are triangulated from multiple local features and can be associated with the corresponding image descriptors [2, 3]. In recent years, the integration of SfM with Multi-View Stereo (MVS) has transformed 3D reconstruction by making data collection and processing more broadly accessible [4]. By capturing and computing multiple overlapping images, it is possible to derive 3D information without relying on knowing camera poses, camera calibration, or surveyed reference markers in the scene. However, in construction, capturing ordered photo collection typically follows a strict schedule and is implemented at predetermined intervals. Furthermore, reconstructing a reality model is time-consuming, and its quality may decrease if random photos are taken at different times. On the other hand, deep learning approaches have demonstrated strong capabilities in providing reliable and real-time image localization [5, 6]. Deep learning-based methods can automatically learn discriminative representations that are robust to large variations in scene appearance and viewpoint changes. Their ability to handle complex environments, varying lighting conditions, and scene dynamics leads to improved accuracy and efficient inference speeds,

making them suitable for real-time deployment [7].

Nowadays, Building Information Modeling (BIM) model is commonly available in the project, and it serves as a rich source for generating synthetic imagery of the scene. By rendering the model from multiple viewpoints, synthetic data can be produced to augment or replace labor intensive image capture. Such synthetic datasets are particularly beneficial for training deep learning–based (e.g., CNN-based) localization models, as they provide diverse examples of the scene without requiring extensive onsite photo documentation. Especially this approach can work with a low level-of-detail BIM model [8], thereby reducing the required time for the indoor camera pose regression process.

In this paper, we present a deep learning-based image localization approach that is suitable for dynamic and complex construction indoor environments. The framework includes two main phases: synthetic data generation and image localization.

2 Related Work

2.1 Traditional Approaches for Camera Pose Estimate

The goal of camera pose estimation is to estimate 6 degrees of freedom (DoF) of a given camera, including 3D translations and 3D orientations. Classical methods for camera pose estimation primarily rely on geometric principles and feature-based matching. Techniques such as Perspective-n-Point (PnP) solvers, including Direct Linear Transform (DLT) [9], EPnP [10], and advanced variants like UPnP and OPnP [11, 12], estimate camera pose from 2D-3D correspondences with improvements in efficiency and robustness. Feature-based approaches, like Structure-from-Motion (SfM), reconstruct 3D scenes by detecting and matching local features (e.g., SIFT, SURF), performing triangulation, and refining camera pose and 3D structures through bundle adjustment [3]. Multi-view geometry techniques, such as essential or fundamental matrix estimation, recover relative camera motion between image pairs. Additionally, model-based methods align 2D image features with known 3D models, while marker-based systems, using fiducial markers like ARTag or AprilTag provide reference points for simplified pose estimation [13, 14]. These methods then form the theoretical foundation for modern deep learning solutions in camera pose estimation.

2.2 Deep Learning-based Image Localization

Compared to traditional methods, which rely heavily on geometric principles, recent advancements in deep learning have transformed the field, offering robust and scalable solutions. A comprehensive overview of state-of-the-art methods for deep learning-based image localization, particularly in indoor environments, is presented in the work of [15], showing numerous benefits of using convolutional neural network (CNN) models for camera pose estimation. CNNs trained on large datasets have demonstrated exceptional capability in learning feature representations. Pose regression can be performed using RGB-D images, which incorporate depth information to encode the 3D location of each pixel within the scene relative to a known environment. Pair-wise registration is a common method in which corresponding 3D key points between consecutive frames are utilized to estimate the relative camera pose [16]. However, depth cameras are not as popular in construction as ordinary cameras, leading to limitations for using RGB-D photos. Deep learning-based approaches using RGB images only, such as PoseNet and its variants, have become prominent in the construction domain. PoseNet directly regresses the camera's 6-DOF (degrees of freedom) pose from an input image, bypassing the need for explicit feature extraction and geometric matching [17]. Variants of PoseNet further enhance accuracy and robustness by introducing loss functions tailored to pose regression, attention mechanisms, or incorporating spatial constraints [8, 18, 19]. Data augmentation is also used to increase fine-tuning samples [20, 21]. Deep learning-based localization methods not only improve the robustness and scalability of image-based localization systems but also adapt well to diverse application domains. By leveraging neural networks' ability to learn feature representations directly from data, these approaches overcome many limitations of traditional methods, such as sensitivity to occlusion, lighting variations, and large-scale environments.

2.3 Synthetic Data for Model Training

Recent advances in synthetic data generation have significantly impacted various computer vision applications, including image localization and camera pose estimation. Synthetic datasets, created through graphics engines or procedural modeling, allow researchers to overcome the challenges posed by limited real-world data, expensive annotation processes, and environmental variability. In construction, synthetic data generated from BIM models can be useful for training CNN models for different tasks, including image localization. Acharya et al. [8] introduced a pose estimation approach to determine the 6-DOF camera pose, which can serve as an initial location input for techniques like PDR, SLAM, and visual odometry. A deep Bayesian recurrent CNN is fine-tuned with synthetic image sequences generated from a BIM model to predict the poses of real image sequences [22]. Hong et al. [23] proposed an image retrieval method for

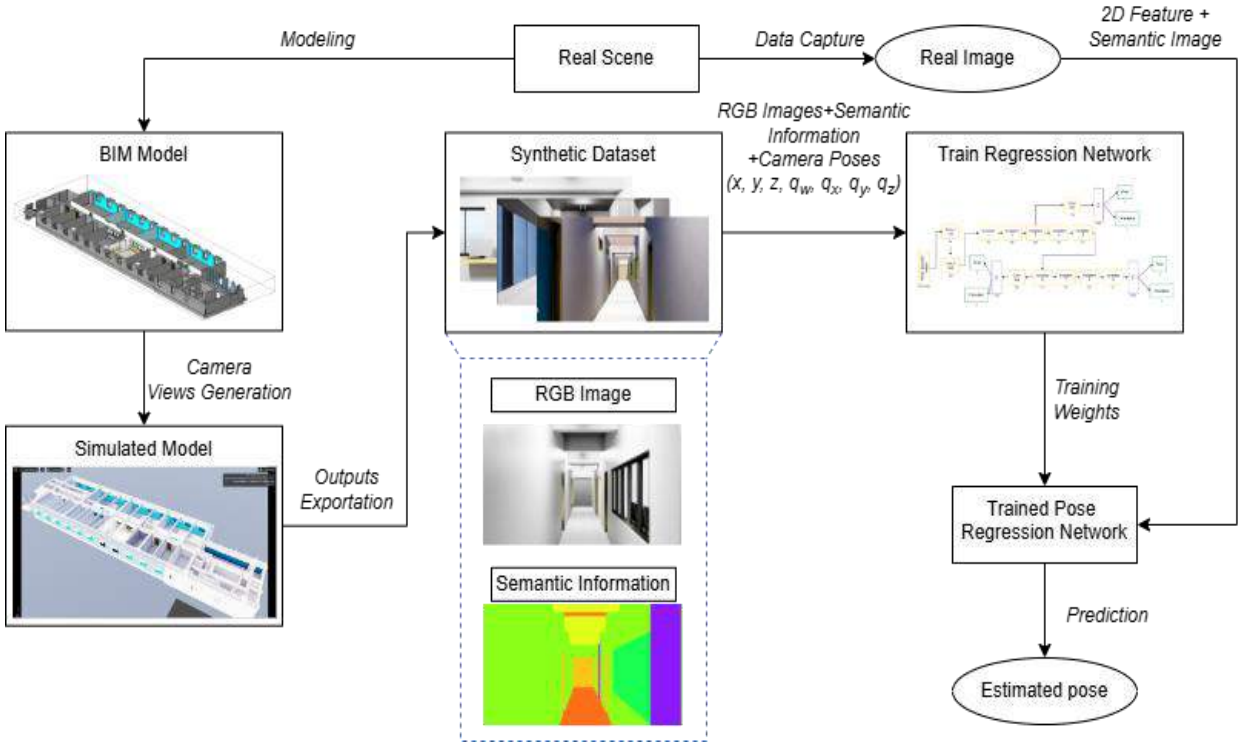


Figure 1. Overview of the framework for CNN-based image localization in this study

achieving coarse indoor localization utilizing BIM. It can be said that synthetic data has emerged as a flexible source for supplementing building image datasets and enhancing the performance and reliability of learning-based visual localization frameworks. The effectiveness of deep learning models trained on synthetic images may be influenced by the reality gap, which refers to discrepancies in color and texture between synthetic and real-world images [24].

In our framework, we combine different platforms to render photorealistic images with similar appearances to real-world images. The rendered outputs are used to train a PoseNet-based CNN model for image localization. The contributions of this study include: (1) propose an automated BIM-based localization for indoor inspection images, and (2) apply cross-domain feature integration to train a deep learning model to improve the indoor image localization.

3 Methodology

3.1 Framework of the CNN-based image localization

Figure 1 indicates our framework for CNN-based image localization, which combines synthetic data generation and deep learning to achieve accurate and robust image localization in dynamic construction scenarios. The BIM model is the basis of real-world environment simulations, including geometric, material, and spatial information, which are necessary to generate a synthetic dataset comprising images rendered from multiple viewpoints under diverse conditions. Furthermore, BIM software provides essential functionalities, such as its API, which allows automated creation and manipulation of camera poses within the model environment. These capabilities enable the efficient generation of diverse camera views, facilitating the development and testing of camera pose estimation techniques. The BIM model is then exported to an advanced simulation platform to generate the required synthetic dataset. This dataset is used to train a deep learning-based pose regression network, which learns to estimate the camera pose directly from input images. Real images captured from the construction scene are fed into the trained network, producing pose predictions.

3.2 Synthetic Data Generation

In this research, camera poses are generated from a 3D model in Revit by defining the positions and orientations of cameras to ensure the coverage of inner spaces in the building. A set of camera poses, as illustrated in Figure 2, is the result of our developed algorithm in Revit API. The process in Revit API begins by setting key parameters such as offset distance and camera position spacing. A transaction is started to

handle model modifications, and room boundaries are identified. For each room, the algorithm extracts boundary segments and generates viewpoints along the offset segments. It then determines valid points for camera placement and assigns viewpoints accordingly. Each viewpoint is processed to compute the camera's position and orientation.

In facility management, the inspectors often need to capture numerous images following a specific path. Thereby, camera poses should be determined on such a camera path that simulates the actual inspection path. After that, the model with camera poses is exported to Nvidia Isaac Sim to add more textures of objects and simulated lights and adjust the camera's parameters. The export process from RVT file in Revit to USD file in Isaac is directly supported by the NVIDIA plugin in Revit, ensuring optimal model preservation. Figure 3 shows the simulated environment in Isaac Sim software, in which camera parameters can be adjusted to match the real camera parameters. Other adjustments can also be executed, such as adding realistic materials and changing light intensity. Using this simulation is necessary because the current camera setting in Revit lacks customized functions to simulate the different cameras. The simulation environment in Isaac Sim is then used to render outputs, including RGB images, depth maps, and instance segmentation maps, which were saved with filenames corresponding to the camera names and timestamps for traceability. This process enabled the generation of high-fidelity visualizations of the architectural model, with realistic camera perspectives derived from the Revit-defined poses. Simultaneously, the position (x, y, z) and orientation (in quaternions) of the virtual cameras are exported into a CSV file as preparation for our CNN model training.

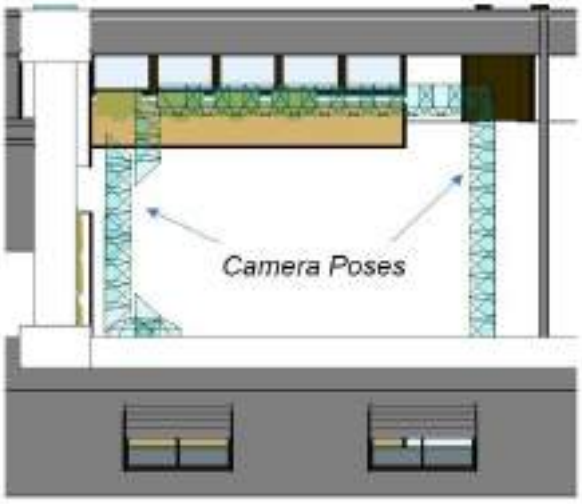


Figure 2. Camera poses defined in Revit



Figure 3. Simulated environment with generated camera poses in Isaac Sim

3.3 Image Localization

The localization of the camera pose is performed using a PoseNet-based CNN network, which has been proposed in our previous research [21]. This network is designed with 22 layers in its deep neural network architecture (Figure 4). In this study, the dataset used for training includes RGB and semantic images generated by Isaac Sim software, as well as their relevant 6-DOF poses. The poses are stored in a CSV file and linked to the relevant input images.

At first, the images are resized to 224×224 pixels and normalized using ImageNet-specific mean and standard (mean = [0.485, 0.456, 0.406] and standard = [0.229, 0.224, 0.225]) as required by the pre-trained GoogLeNet. The encoder, which includes convolutional layers and inception modules, extracts features from input images. This module uses pre-trained weight from the original GoogLeNet. An illustration of the inception module can be found in Figure 5. Each inception module captures a deeper level of abstraction from the input data. Originally in GoogLeNet architecture, the final layer of the network, representing the highest level of abstraction, along with

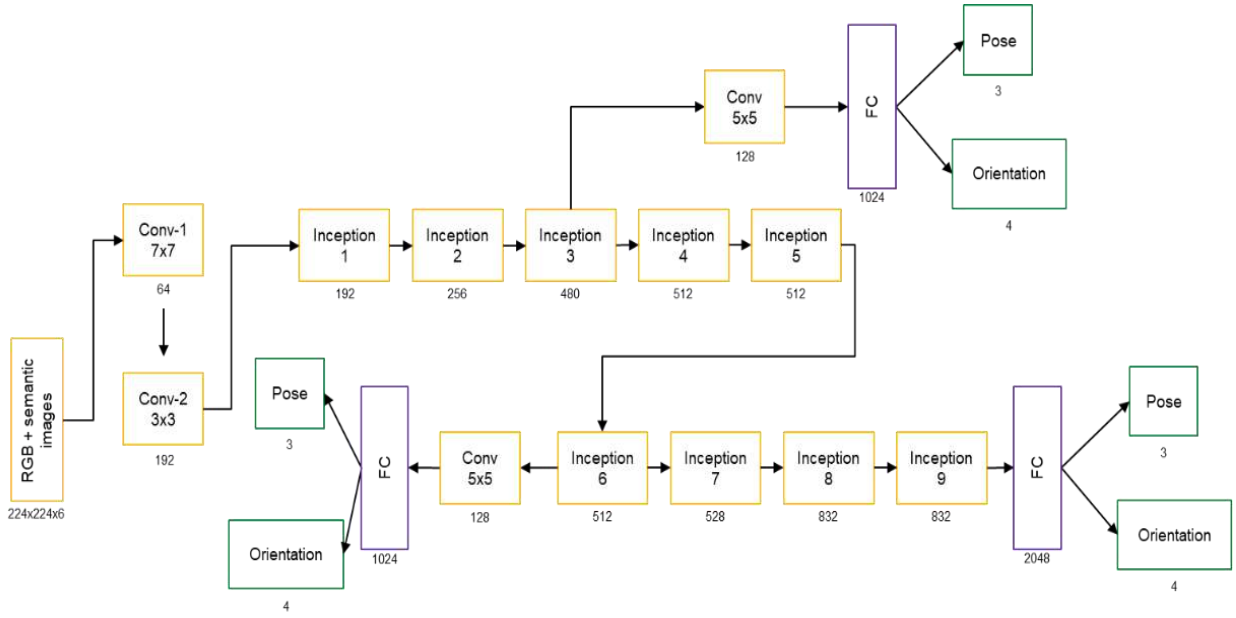


Figure 4. The architecture of CNN-based pose regression network

two intermediate abstraction levels, is passed to fully connected and Softmax layers to classify the objects.

PoseNet, however, replaces the Softmax classifier with two parallel fully connected layers with 3 and 4 units, respectively. These regressors output a 7-dimensional vector representing the camera's 6-DOF pose, which includes the 3D position $\hat{X}_c$ and orientation $\hat{q}_c$ in quaternion format. The network is trained using a loss function that minimizes the position error $\|\hat{X}_c - X_c\|$ and the orientation error $\|\hat{q}_c - q_c\|$ with a balancing hyperparameter to ensure optimal performance. Xc and qc are the ground truth position and orientation of the input camera poses, respectively.

$$Loss(I) = \|\hat{X}_c - X_c\|_2 + \beta\|\hat{q}_c - q_c\|_2 \qquad (1)$$

The hyperparameter β is used to balance the weights of position and orientation losses, keeping them with in a similar range. Since position errors and orientation errors are measured in different units and magnitudes, directly optimizing them together may lead to an imbalance, where one dominates the loss function. Thereby, β acts as a scaling factor that adjusts the relative contribution of orientation loss compared to position loss. If β is too high, orientation accuracy improves at the cost of position accuracy, and vice versa. Previous experiments suggest that β is typically assigned values between 250 and 2,000 [17]. The closest camera pose is determined by calculating the distance in position (Euclidean distance) and angular difference in orientation (quaternion angular distance) for each camera pose in the dataset

4 Initial Result and Discussion

In our study, an academic building model was used, containing different offices and a long hallway. The synthetic dataset generated by Isaac Sim includes 1922 RGB images and 1922 semantic images. Each camera position has five different orientations. These images' poses are controlled by creating camera paths offset from the room walls. These camera paths simulate the tracks when implementing inspections for facility management. The integration of Revit for camera pose generation, and Isaac Sim for advanced rendering provides an effective workflow for applications such as architectural visualization, training datasets for AI models, or construction progress monitoring. The resulting outputs demonstrate the feasibility of leveraging a combined pipeline for precise and visually rich simulation data.

When preparing the dataset, 80 percent of the images were used to train the pose regression model. Testing and validation used 10% of the data relatively. The testing results show an error of 5.81 meters in translation and 69.05 degrees in orientation. This section aims to evaluate the accuracy of our network for single-image localization using only synthetic images. The goal is to analyze the network's performance when trained exclusively on synthetic data and tested directly on real images without employing any techniques to bridge the domain gap.

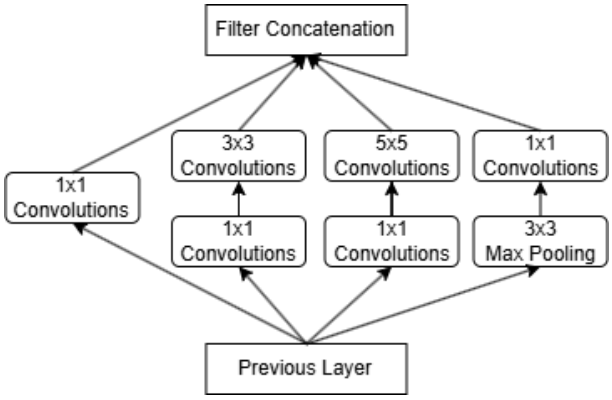


Figure 5. The inception module based on [17]

The test result can be considered acceptable in predicting the pose of an actual image, especially when the model was trained with a small dataset (less than 4000 images compared to other common models). Some causes can be listed for this result. At first, only images of the virtual environment and semantic information were used to train the mode. It is noticeable that there is a vast difference between the simulated environment and real-world space, including the furniture and objects in the scene. We acknowledge the complexity of the environment captured in the real image, including different finishing materials and furniture. Therefore, the images in our synthetic dataset subsequently lack features compared to the images captured in the real world. When using the pre-trained GoogLeNet backbone, the convolutional layers learn to recognize lines and geometric patterns as part of the training process. These features do not exist in the images in the synthetic dataset, thereby issuing challenges for the network. Furthermore, the network's ability to estimate the camera's rotation is less accurate, likely due to the complexity of capturing fine-grained orientation details. Without refinement techniques, the network struggles to bridge this gap, leading to suboptimal performance on real-world data.

For future work, additional features should be utilized to train the regression model. The simulation software Isaac Sim can generate depth information from images, which is essential for predicting spatial details. This input will support the prediction process. On the other hand, PnP and RANSAC algorithms are preferred for refining the poses predicted by the CNN in the earlier stage. The process will use a set of virtual images taken from the initial predicted poses, which serve as the starting point for the pose refinement process. The PnP algorithm leverages 2D-3D point correspondences to estimate the camera's position and orientation, minimizing reprojection error. In detail, it solves the transformation between a set of 2D image points and their corresponding 3D points in a known reference model such as a 3D point cloud. Then, PnP estimates the projection matrix, which aligns the image features with their 3D world coordinates. However, the presence of mismatched or noisy correspondences can degrade accuracy. This is where RANSAC becomes essential, as it ensures that only inliers contribute to the pose estimation. The algorithm repeatedly selects small random subsets of point matches to estimate the pose using PnP and evaluates the consistency of the remaining points. By using RANSAC, mismatched feature correspondences or erroneous keypoints are rejected, ensuring that only inliers contribute to the final pose estimation. As validated in our previous study [21], PnP and RANSAC enhance the robustness and precision of camera pose predictions, making them particularly effective for scenarios involving noise or incomplete data.

5 Conclusion

In conclusion, this study introduces an automated method for creating photorealistic synthetic images of BIM scenes depicting building interiors. In addition, a deep learning-based framework for image localization tailored to challenging and dynamic construction environments has been presented. The proposed framework includes two stages: the first stage involves synthetic dataset generation using BIM software and an advanced simulation platform; the second is a CNN-based pose regression. The pose regression model can use inputs, including RGB images and semantic images for prediction. The findings demonstrated that currently, using only a synthetic dataset generated in advanced simulation software can be reliable for training a CNN-based pose regression model. The network performs well for the initial coarse localization. However, in order to improve the performance of localizing a real-world image, implementing pose refinement will be the next step in our research. Future work will improve this initial prediction by matching the query image with virtual images generated from the BIM model, using geometric pose optimization methods such as PnP and RANSAC. The refined poses ensure higher accuracy, enabling the alignment of real images with the virtual model. This combination of synthetic data, deep learning, and geometric optimization creates a robust pipeline for localizing images in complex and evolving construction environments.

References

[1] M. A. Fischler and R. C. Bollers. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381-395, 1981

[2] M. Pollefeys, et al. Visual Modeling with a Hand-Held Camera. *International Journal of Computer Vision*, 59(3):207-232, 2004

[3] T. Sattler, B. Leibe and L. Kobbelt. Efficient & Effective Prioritized Matching for Large-Scale Image-Based Localization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(9):1744-1756, 2017

[4] M. W. Smith, J. L. Carrivick and D. J. Quincey. Structure from motion photogrammetry in physical geography. *Progress in Physical Geography: Earth and Environment*, 40(2):247-275, 2016

[5] A. Kendall and R. Cipolla. Geometric loss functions for camera pose regression with deep learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5974-5983, 2017

[6] J. M. H. Noothout, et al. Deep Learning-Based Regression and Classification for Automatic Landmark Localization in Medical Images. *IEEE Transactions on Medical Imaging*, 39(12):4011-4022, 2020

[7] Y. Shavit and R. Ferens. Introduction to camera pose estimation with deep learning. *arXiv preprint arXiv:1907.05272*, 2019

[8] D. Acharya, K. Khoshelham and S. Winter. BIM-PoseNet: Indoor camera localisation using a 3D indoor model and deep learning from synthetic images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 150:245-258, 2019

[9] Y. Liu, T. S. Huang and O. D. Faugeras. Determination of camera location from 2-D to 3-D line and point correspondences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):28-37, 1990

[10] V. Lepetit, F. Moreno-Noguer and P. Fua. EPnP: An Accurate O(n) Solution to the PnP Problem. *International Journal of Computer Vision*, 81(2):155-166, 2009

[11] L. Kneip, H. Li and Y. Seo. UPnP: An Optimal O(n) Solution to the Absolute Pose Problem with Universal Applicability. In pages 127-142,Cham, 2014

[12] Y. Zheng, Y. Kuang, S. Sugimoto, K. Astrom and M. Okutomi. Revisiting the pnp problem: A fast, general and optimal solution. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2344-2351, 2013

[13] M. Fiala. ARTag, a fiducial marker system using digital techniques. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, pages 590-596 vol. 2, 2005

[14] E. Olson. AprilTag: A robust and flexible visual fiducial system. In *2011 IEEE International Conference on Robotics and Automation*, pages 3400-3407, 2011

[15] X. Bai, M. Huang, N. R. Prasad and A. D. Mihovska. A Survey of Image-Based Indoor Localization using Deep Learning. In *2019 22nd International Symposium on Wireless Personal Multimedia Communications (WPMC)*, pages 1-6, 2019

[16] D. R. d. Santos, M. A. Basso, K. Khoshelham, E. d. Oliveira, N. L. Pavan and G. Vosselman. Mapping Indoor Spaces by Adaptive Coarse-to-Fine Registration of RGB-D Data. *IEEE Geoscience and Remote Sensing Letters*, 13(2):262-266, 2016

[17] A. Kendall, M. Grimes and R. Cipolla. Posenet: A convolutional network for real-time 6-dof camera relocalization. In *Proceedings of the IEEE international conference on computer vision*, pages 2938-2946, 2015

[18] D. Jia, Y. Su and C. Li. Deep convolutional neural network for 6-dof image localization. *arXiv preprint arXiv:1611.02776*, 2016

[19] R. Zhang, Z. Luo, S. Dhanjal, C. Schmotzer and S. Hasija. PoseNet++: A CNN framework for online pose regression and robot re-localization. *Tech. Rep*, 2018

[20] M. S. Müller, S. Urban and B. Jutzi. Squeezeposenet: Image based pose regression with small convolutional neural networks for real time uas navigation. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, IV-2/W3:49-57, 2017

[21] T. H. Wang, A. Pal, J. J. Lin and S.-H. Hsieh. Construction Photo Localization in 3D Reality Models for Vision-Based Automated Daily Project Monitoring. *Journal of Computing in Civil Engineering*, 37(6):04023029, 2023

[22] D. Acharya, S. Singha Roy, K. Khoshelham and S. Winter. A Recurrent Deep Network for Estimating the Pose of Real Indoor Images from Synthetic Image Sequences. *Sensors*, 20(19):5492, 2020

[23] Y. Hong, S. Park, H. Kim and H. Kim. Synthetic data generation using building information models. *Automation in Construction*, 130:103871, 2021

[24] H. Ying, R. Sacks and A. Degani. Synthetic image data generation using BIM and computer graphics for building scene understanding. *Automation in Construction*, 154:105016, 2023

Slate detection in orthophotos of building roof panels

Jiajun Li[1], Frédéric Bosché[1], Chris Xiaoxuan Lu[2], Lyn Wilson[3] and Boan Tao[1]

[1]School of Engineering, University of Edinburgh, UK
[2]Department of Computer Science, University College London, UK
[3]Historic Environment Scotland, UK

Jiajun.Li@ed.ac.uk, f.bosche@ed.ac.uk, xiaoxuan.lu@ucl.ac.uk, lyn.wilson@hes.scot, boan.tao@ed.ac.uk

Abstract -

Traditional roof inspection methods rely heavily on manual labour, which poses challenges in efficiency and safety. In this paper, we propose an overall approach and 4 specific methods for slate detection in orthophotos of building roof panels, using edge detection by Canny detector (Method 1), SAM (Segment Anything Model) (Method 2), SAM with bounding box prompt (Method 3), and edge detection on original SAM masks (Method 4). The methods are developed using a dataset composed of 4 sample roof segments and 3 complete orthophotos from real roof panels. Method 4 reaches the best results (overall results for all slates: mIoU = 0.67, Recall = 0.81) with Method 1 also performing well (mIoU = 0.57, Recall = 0.74). Such performance demonstrates the potential of the proposed approach and these methods to provide a robust foundation for slate detection and subsequently classification tasks.

Keywords -

Slate; Edge detection; Segmentation; Roof orthophoto

1 Introduction

Roof systems are a critical building fabric element serving to protect interior spaces and inhabitants against the elements (in particular rain). Slated roof systems are common type of roofs in both historic and modern buildings, requiring monitoring and timely maintenance to prevent damages, including structural damages. This is being made increasing challenging by climate change. As global weather patterns become increasingly erratic, slated roofs are subject to more frequent and severe fluctuations in temperature, moisture, and wind intensity. These environmental stressors can exacerbate the natural vulnerabilities of slates, a metamorphic rock that, while durable, has specific material properties—such as its anisotropic nature and inherent brittleness—that make it susceptible to damage under variable climate conditions [1, 2]. Besides, the nails used to secure slates can rust and fail. As a result, slated roofs often experience a range of deterioration issues, including broken, displaces, or detached (missing) slates.

The repair and maintenance of slated roof necessitate a targeted approach due to the localised nature of damage. Given that damage typically affects individual slates rather than large sections of the roof, restoration efforts often focus on a slate-by-slate basis. This time-saving and cost-effective approach is essential, not only to maintain the structural integrity and aesthetic quality of roofs but also to extend the lifespan of overall roofing systems. Therefore, a meticulous analysis of each individual slate is warranted to assess the type (if any) and extent of deterioration.

Current research on roof defect detection methods has advanced from manual inspection towards automation. For flat roofs, visual inspection by skilled experienced professional, with auxiliary means by sensors such as electric and ehermohygrometric methods, is the dominant method for non-destructive testing [3]. But, to capture data from buildings without the need for human access provision, Unmanned Aerial Vehicle (UAV)-based close-range sensing is increasingly used [4]. With the development of Artificial Intelligence (AI), deep learning models such as U-Net and HRNet are being developed to detect and categorise flat roof defects from aerial images, such as hollows, swellings, etc. These defects show continuous and large-area features [5]. In studies targeting at generic roof types, Mostafa et al. [6] apply Convolutional Neural Networks (CNNs) to classify roof damage types, achieving high accuracy in prioritising repairs.

However, during the evaluation and repair process of pitched roofs, workers must locate, assess, and restore individual slates at elevated heights, posing safety risks and inefficiencies. This further underscores the importance of precisely identifying the position and defect status of each slate to enable targeted physical intervention. As a result, pitched roofs, especially traditional pitched roofs, present unique challenges for automated and remote inspection: 1. Different from flat roofs, defects (mostly structural damages) of slated roofs are linked to single slate units. Therefore, defect detection should ideally start with slate mapping. 2. The complexity of slate arrangement and the variability in appearance, alongside variations in lighting and occlusions in roof images, further complicates automated analysis. 3. Defect categories of flat roofs differ

from those of pitched roofs, making existing defect detection model inapplicable. Addressing these challenges necessitates an advanced methodology capable of detecting individual slates and evaluating their condition in a robust and scalable manner.

Using UAV-based photogrammetric reconstruction, Digital Twin -based data processing, and Artificial Intelligence have been shown to be useful to automated slated roof monitoring Li et al. [7], from orthophoto generation for each roof panel [8] to roof sub-component segmentation [9]. These processes realise the gradual deconstruction of roof in a more measurable and analysable way. To further detect slate defects, one approach can be to first detect each individual slate in the roof orthophotos, and then classify each detected slate according to its condition, supporting high-precision diagnostics that can automate and streamline the repair process. This study proposes and compares a few methods of detecting individual slates, such as using traditional image processing in edge detection, the innovative Segment Anything Model (SAM).

2 Related work

2.1 Object detection using construction images

Image processing has emerged as a cornerstone of construction informatics, facilitating automated analyses for tasks like defect detection, progress monitoring, and material classification. Among them, object detection is the most common application. Simple yet effective techniques such as thresholding (e.g., Otsu's method) and filtering, are commonly applied in crack detection [10]. Canny edge-detector and watershed transformation are used in building component (column) detection [11]

With the advent of deep learning, modern models are built for object detection, like convolutional neural networks (CNNs), transformer-based vision architectures or more advanced zero-shpt SAM. For example, CNN refine input data for edge detection, making it easier for the network to learn spatial features. Such integrations enable applications like real-time construction site monitoring, where drones capture data that is processed for safety analysis [12], equipment detection [13]. While SAM-related research enables provide more accurate and comprehensive results [14]. Image processing in construction not only enhances operational efficiency but also addresses challenges like labour shortages and the need for faster project delivery, making it indispensable for modern construction workflows.

Based on object detection, segmentation helps further clarify the boundary of each object. To realise precise segmentation, edge detection-based methods and direct segmentation methods can be considered.

2.2 Edge detection -based methods

Traditional edge detection -based methods focus on explicitly identifying intensity discontinuities in images, often marking the boundaries of objects. These methods typically involve gradient-based techniques, such as:

1. *Canny Edge Detection* combines Gaussian smoothing, gradient calculation, non-maximum suppression, and hysteresis thresholding to produce clean and precise edges [15].
2. *Sobel and Prewitt Operators* compute gradients in horizontal and vertical directions to highlight regions with high intensity variation, making them ideal for detecting straight lines or simple contours [16, 17].
3. *Laplacian of Gaussian (LoG)* identifies edges by detecting zero-crossings in the second derivative of the image, providing robust results for blob-like structures [18].

Recent advancements in deep learning have significantly enhanced traditional edge detection methods, shifting from purely gradient-based approaches to data-driven, learning-based frameworks. Modern edge detection methods leverage convolutional neural networks (CNNs) and other architectures to achieve robust and adaptive performance in diverse scenarios. For instance, models like Holistically-Nested Edge Detection (HED) use deep CNNs to learn hierarchical edge features, allowing for precise edge maps with context-aware refinement [19]. Transformer-based approaches, such as DETR (DEtection TRansformer), implicitly model edge-like information by attending to boundary regions during object localisation and segmentation tasks [20].

2.3 Segment Anything Model (SAM)

Segment Anything Model (SAM) is a groundbreaking zero-shot learning model designed for universal image segmentation. It is trained using segmentation dataset with over 1 billion masks on 11M images. Unlike task-specific models, SAM employs a flexible prompting mechanism, enabling it to generalise across diverse image types and tasks without fine-tuning. It accepts various prompts such as points, bounding boxes, and even textual descriptions, making it adaptable to many domains. For example, SAM can segment complex natural scene images with minimal human intervention, showcasing its versatility [21].

Traditional edge detection methods offer efficient, lightweight solutions for extracting structural details, especially in scenarios where computational resources are limited, while deep learning approaches excel in capturing complex patterns and contextual information through data-driven feature learning. SAM adds a high-level semantic understanding to segmentation tasks, enabling precise object delineation across diverse and challenging vi-

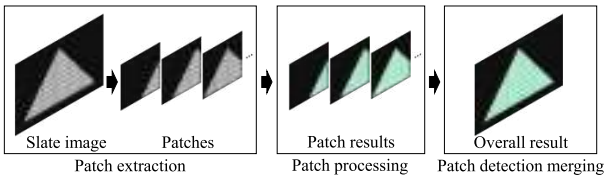


Figure 1. Approach overview

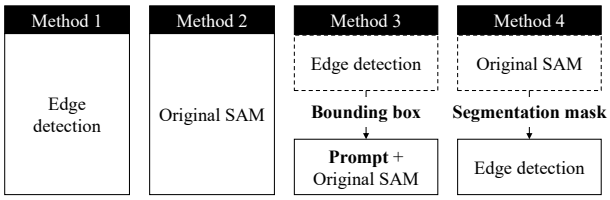


Figure 2. Method development of slate detection

sual contexts.

3 Approach overview

Roof panels, and consequently roof panel orthophotos come in various sizes. Since the performance slate detection algorithms is dependent on the size of the slates in the processed image, it is necessary to process each roof panel orthophoto in patches representing roofs at similar scale. As a result, the overall pipeline for slate detection in roof panel orthophotos includes the following three steps (fig. 1):

1. **Patch extraction**: the image is split into an overlapping set of patches;
2. **Patch processing**: slates are detected in each patch; and
3. **Patch detection merging**: all detections are merged in the original image.

Sections 4 to 6 detail those three steps.

4 Patch extraction

Small fixed-size patches (200×200 pixels) are extracted from the complete slate images using a sliding window approach. Square patches are the best option for inputting in SAM model to prevent distortion. The step size is 25% of the window size (i.e. 50 pixels), creating overlapping patches. This ensures that each slate will appear fully in at least one patch. The mask of the slated area in the orthophoto is also a known input from the orthophoto generation theory [8].

5 Slate detection

Four slate detection methods are investigated in this study (see fig. 2). They are developed according to the following logic: **Method 1** is developed based on traditional image processing using edge detection. Thanks to the well-developed AI model by meta, SAM, **Method 2** is developed by using SAM as is. However, as will be shown, **Method 2** shows lower performance than **Method 1**. As a result, Methods 3 and 4 are investigated. **Method 3** uses the bounding boxes obtained using **Method 1** as prompt to SAM. **Method 4** employs the same detection strategy of **Method 1** but applied to the SAM segmentation map. The 4 methods are described in more detail in the following sub-sections.

5.1 Method 1: Edge detection by Canny detector

The process of the application of **Method 1** to each patch is as follows:

1. *Noise removal*: Gaussian blur is applied to the patch (or the overall image) to reduce noise.
2. *Edge detection*: The Canny edge detector is applied to the patch, followed by oriented horizontal edge detection using Sobel.
3. *Slate row detection:* Probabilistic Hough Transform-based line detection is applied to the edge map to identify horizontal slate rows in the image. Only lines where the absolute difference between the y-coordinates of the two endpoints is less than 10 are retained. The most common distance between neighbouring lines — which should correspond to the roof panel *exposure*, also called *margin* — is calculated, and any line whose distances to neighbouring lines are smaller than the exposure is discarded. Similarly, lines are added between pairs of lines whose distance is twice (or three times) the exposure. The final set of lines is used to split the image into strips, with each strip expected to contain one slate row.
4. *Slate detection*: For each slate row, the average vertical grayscale value is computed across the strip, and valleys in this resulting 1D signal are assumed to correspond to the locations of vertical boundaries between slates in the strip, and the strip is segmented correspondingly into boxes defining initial slate boundaries, The slates boxes detected across all strips are gathered in the patch.
5. *Slate detection filtering*: Boxes that contain more than 90% black pixels (i.e. background pixels in the orthophoto that are not part of the roof panel) are discarded as noise.
6. *Slate detection adjustment*: Because slates may not be laid in a strict horizontal direction, each slate box is slightly resized by locally adjusting their top and bottom edges to better match the local minimum average grayscale along the edges. The resizing is restricted



Figure 3. Method 4 pipeline, left to right: grayscale image - SAM segmentation mask - line extraction based on Hough Transform - strip extraction - strip valley calculation - box visualisation - box shifting.

within a distance set to 30% of the slate exposure.

5.2 Method 2: Original SAM

In **Method 2**, the original SAM is used without any other input to generate the probability map of segments, followed by a segmentation mask by setting the threshold as the median value of the entire probability map. Bounding boxes are then generated by extracting the largest connected components after removing noise (small objects, pixel count < 25) and smoothing edges. Due to the transition area from slate to slate gap, segments by SAM tend to be smaller than the actual slates. Therefore, a two pixel extension is added to the four boundary sides to make the boxes closer to reality and touch each other. Theoretically, this approach may be able to better generalise than **Method 1**, since SAM offers strong generalisation capabilities and high-level semantic understanding.

5.3 Method 3: SAM with bounding box from edge detection as prompt

In **Method 3**, the original SAM is used with the bounding box information from **Method 1** used as prompt, with the aim of increasing the probability of a slate segment inside the box while retaining the theoretical capability of SAM to detect slates that may not have a rectangle appearance. The detection box refinement steps of **Method 1** are then applied the same way.

5.4 Method 4: Edge detection by original SAM mask

Because we noticed that the segmentation masks produced by SAM have clear slate boundaries, possibly clearer than in the original grayscale image, in **Method 4** SAM is applied to the grayscale image, outputting a probability map indicating pixel-level segmentation confidence. By setting the threshold as the median probability value, a segmentation mask is generated to which the Canny edge detector of **Method 1** is applied. The detection box refinement steps of **Method 1** are then applied the same way.

6 Patch slate detection results merging

All slate detections from all patches are gathered in the overall roof panel orthophoto. The resulting set of detections contains redundant detections from overlapping detections as well as boxes with abnormal sizes (noise). To merge and filter those results, the following steps are applied:

1. *Boundary boxes*: Boxes located on the boundary of overlapping patches are discarded.
2. *Small boxes*: Boxes whose height or width is $< 0.5 \times exposure$ are discarded, but only if $> 90\%$ of the area of the orthophoto covered by the box is already covered by another box.
3. *Redundant boxes*: Non-Maximum Suppression (NMS) is used to discard redundant (overlapping) boxes. When boxes are overlapping, NMS keeps the most significant one based on a predefined Intersection-over-Union (IoU) threshold (0.3). However, like above, a box is discarded only if $> 90\%$ of the area of the orthophoto that it covers is already covered by another box.

7 Experimental results

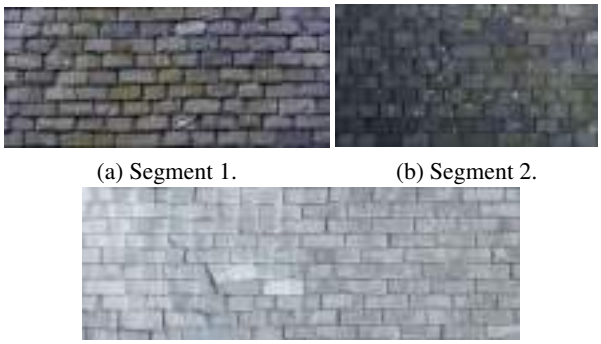
7.1 Dataset

The dataset used in the study is a set of orthophotos obtained from real slated roof panels using the method in [7]. The orthophotos are of roof panels with different shapes and deterioration conditions (see fig. 7) . Panel A and B are expected to be easier due to their smaller size, while Panel C is much larger and has the highest deterioration level. The three panels collectively contain 5,386 slates. Ground Truth (GT) labelling for each slate is done manually with axis-aligned bounding box (even though some slates are not rectangular). The results are presented for sample images first and then for the overall roof panels.

7.2 Performance evaluation metrics

Recall and Intersection over Union (IoU) are widely used metrics for evaluating object detection and segmentation tasks. Recall helps evaluate how well a model identifies all objects in an image, while IoU provides a robust measure of spatial overlap between predicted and ground truth (GT) regions[22]. Slate detection performance is thus evaluated here using both Recall and IoU. Note that each predicted box is matched to the GT box with highest IoU, and classified according to the IoU value as summarised below:

- *TP (True Positive)*, if $IoU >= 0.5$.
- *FP (False Positive)*, if $0.1 <= IoU < 0.5$. It is further classified into:



(a) Segment 1. (b) Segment 2.

(c) Segment 3.

Figure 4. Three segments extracted from the three roof panels.

Method	mIoU	Recall
1	0.86	0.97
2	0.62	0.65
3	0.76	0.89
4	0.86	0.97

Table 1. Performance (mIoU and Recall) of the four slate detection methods applied to the three sample orthophoto segments.

  - *FP - Over segmented*, if GT box size > Prediction box size.
  - *FP - Under segmented*, if GT box size < Prediction box size.
- *FN (False Negative)*, if $IoU < 0.1$.

Overall performance is evaluated using:

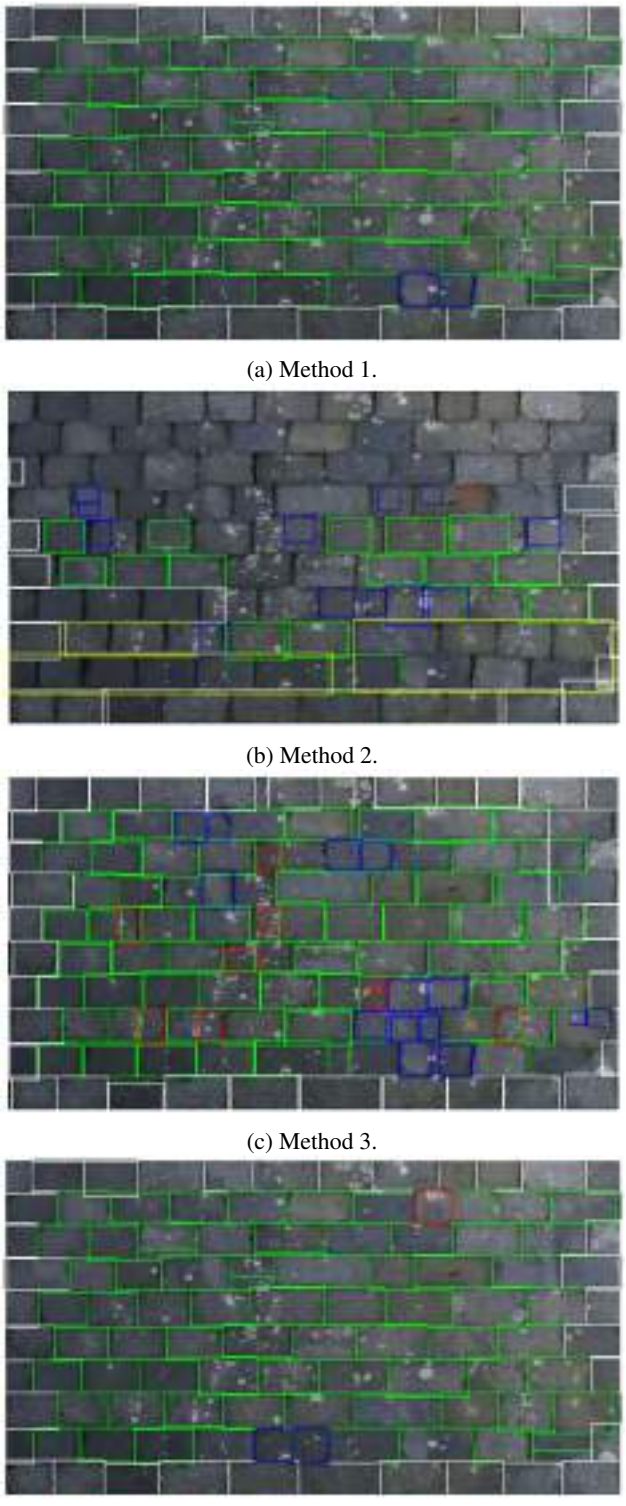
- *mIoU*, the mean IoU values;
- *Recall*, the proportion of slates correctly detected.

7.3 Detection results on sample roof segment

Results using the 4 different methods were first obtained for three roof segments extracted from the three panels, shown in fig. 4. The results for the three segments are summarised in table 1, and results can be visualised for one segment in fig. 5.

Note that the slates located at the boundaries of the segments are typically incomplete, and should thus be discarded from the performance analysis. The three segments collectively contain 202 non-boundary slates, and 114 discarded boundary slates, and the results in table 1 are obtained by considering only the non-boundary slates.

The first observation is that **Method 1** clearly outperforms **Method 2**. **Method 2** performs poorly because some its segmentation map sometimes includes small connections between the detecting areas of neighbouring slates leading to the detection of one large (that is subsequently filtered out) instead of two slates.



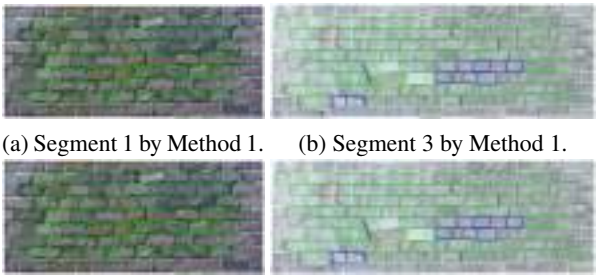
(a) Method 1.

(b) Method 2.

(c) Method 3.

(d) Method 4.

Figure 5. Slate detection results obtained by the four methods on Segment 2 (White: Boundary; Green: TP; Blue: FP - Over Segmented; Yellow: FP - Under segmented; Red: FN).



(a) Segment 1 by Method 1. (b) Segment 3 by Method 1.

(c) Segment 1 by Method 4. (d) Segment 3 by Method 4.

Figure 6. Visualisation of results obtained with Method 4 for Segments 1 and 3.

Method	Panel	Non-boundary		All slates	
		mIoU	Recall	mIoU	Recall
1	A	0.72	0.89	0.68	0.85
	B	0.57	0.78	0.54	0.73
	C	0.57	0.75	0.56	0.73
	Overall	0.58	0.77	0.57	0.74
4	A	0.78	0.93	0.76	0.91
	B	0.63	0.83	0.59	0.78
	C	0.69	0.83	0.67	0.81
	Overall	0.69	0.84	0.67	0.81

Table 2. Comparison between different segmentation methods in raw slate images.

The results of **Method 3** demonstrate an obvious improvement compared to **Method 2**. The bounding box prompt helps increase the probability of segments inside each prompt bounding box, making **Method 3** distinguish individual slates more robustly than **Method 2**.

Method 4 performs even better, to a similar level to **Method 1**. However, because of the mechanism behind the two methods (Canny detector and SAM segmentation mask), they differ in their sensitivity to edges, resulting in both methods seemingly failing in different locations.

fig. 6 show the results obtained with **Method 4** on the remaining two segments.

7.4 Detection results on complete slate images

The results obtained for the three orthophoto segments in section 7.3 clearly show the superiority of **Method 1** and **Method 4**. Therefore the performance on the complete orthophotos is now evaluated for only those two methods. The results are summarised in table 2 and those obtained with **Method 4** can be visually explored in fig. 7. Note that the analysis distinguishes the performance with and without *boundary* slates, that are the slates whose GT box intersects the slate mask, i.e. intersect the orthophoto background of black pixels. The reason for this distinction is that boundary slates typically present a greater range of shape (e.g. truncated) and appearance (e.g. presence of leadwork pixels), which may challenge the detectors.

Overall, **Method 4** performs best, with $mIoU = 0.67$ and $Recall = 0.81$. The better seems because the SAM segmentation map demonstrates clearer slate boundaries than the Canny edge detector. Such advantage is more obvious in the complete orthophotos that include a broader range of complex shape and conditions (e.g. boundary slates, holes, etc.).

However, **Method 4** still makes a number of errors. For example, there are a lot of noisy bounding boxes generated in areas containing biological growth (see the upper and right part of Panel C). The uneven brightness in the bottom part of Panel A also affects the vertical slate edge detection, leaving some non-boundary slates undetected. The unclear slate boundary in the top part of Panel B results in more slates detected in FP (Under segmentation) status.

The comparison of the results for all slates and non-boundary slates shows a drop of 2 percentage points in the performance for all slates. As noted above, a drop was expected due to the fact that boundary slates tend to have a greater range of shapes and appearance. However, the drop appears to be only minor.

8 Discussion

The results of **Method 1** and **Method 4** already demonstrate promising results, with mIoU=0.67 and recall=0.81 for **Method 4** and mIoU=0.57 and recall=0.74 for **Method 1**, which can form the basis for defect assessment of individual slates, e.g. through classification. However, the orthophoto quality noticeably influences the accuracy, especially in Panel B where the slates at the bottom of the panel are not as clear. This motivates further work to improve data collection and orthophoto generation. Besides, FPs are noticed to be common in areas where there is the presence of biological growth. Therefore, methods should be explored to better discriminate slate boundaries from biological growth in the image input to the Hough Transform-based line detection (from Canny edge detector or SAM segmentation map).

We note that the proposed slate segmentation approach is designed for typical traditional slated roofs (laid in in horizontal rows) and does not require any roof-specific parameters. The approach is thus applicable to most slated buildings. However, slated roofs may at times have different patterns, with slates cut and laid in other ways (e.g. fish scale patterns). For those roofs, **Method 1** and **Method 4** would not be applicable, but **Method 3** could prove suitable.

Finally, while SAM is shown to be able to highlight edge features, some alternative deep learning model for image segmentation (and slate detection) could be explored.



(a) Panel A.



(b) Panel B.



(c) Panel C.

Figure 7. Visualisation of complete slate image results obtained by Method 4.

9 Conclusion

This study proposed an overall approach and 4 specific methods for slate detection in orthophotos of building roof panels, comparing: edge detection by Canny detector (**Method 1**), SAM (**Method 2**), SAM with bounding box prompt from **Method 1** (**Method 3**), and edge detection on the original SAM mask (**Method 4**). All four methods were first evaluated on three cropped orthophoto segments, and **Method 1** and **Method 4** stood out with similar and high mIoU (0.86) and Recall (0.97). The two methods were then further evaluated on complete roof panel orthophotos, on which **Method 4** showed superior performance. The results highlight the encouraging levels of performance of **Method 1** and **Method 4**, but also their difficulty in areas containing significant biological growth and their sensitivity to roofs acquired in poorly lit conditions. Their distinct performance in different areas also suggest that future research could — alongside evaluating the methods on more extensive datasets — possibly explore integrating their results.

10 Acknowledgements

This paper was made possible thanks to research funding from Historic Environment Scotland (HES) and the Engineering and Physical Sciences Research Council (EPSRC) [grant reference EP/W524384/1]. The views and opinions expressed in this article are those of the authors and do not necessarily reflect the official policy or position of HES and EPSRC. The authors would also like to acknowledge the HES Digital Documentation and Innovation team, Stirling City Heritage Trust for providing us with the data used in the experiments reported in this paper.

References

[1] J. Wichert. *Mining of Slate and Production*, pages 159–201. Springer International Publishing, Cham, 2020. doi:10.1007/978-3-030-35667-5_6.

[2] V. Cárdenes, J. P. Cnudde, J. Wichert, D. Large, A. López-Mungira, and V. Cnudde. Roofing slate standards: A critical review. *Construction and Building Materials*, 115:93–104, 2016. doi:10.1016/j.conbuildmat.2016.04.042.

[3] J. Conceição, B. Poça, J. de Brito, I. Flores-Colen, and A. Castelo. Data analysis of inspection, diagnosis, and rehabilitation of flat roofs. *Journal of Performance of Constructed Facilities*, 33(1):04018100, 2019. doi:10.1061/(ASCE)CF.1943-5509.0001252.

[4] E. Adamopoulos and F. Rinaudo. Close-range sensing and data fusion for built heritage inspection

and monitoring—a review. *Remote Sensing*, 13(19), 2021. doi:10.3390/rs13193936.

[5] D. A. Yudin, V. Adeshkin, A. V. Dolzhenko, A. Polyakov, and A. E. Naumov. Roof defect segmentation on aerial images using neural networks. In *Advances in Neural Computation, Machine Learning, and Cognitive Research IV*, pages 175–183, 2021.

[6] K. Mostafa, T. Hegazy, R. D. Hunsperger, and S. Elias. Using image analysis to quantify defects and prioritize repairs in built-up roofs. *Facilities.*, 41 (7/8), 2023-04-26.

[7] J. Li, B. Tao, F. Bosché, C. X. Lu, and L. Wilson. Automated generation and semantic segmentation of roof orthophoto for digital twin -based monitoring of slated roofs. *Automation in Construction*, 167: 105725, 2024. doi:10.1016/j.autcon.2024.105725.

[8] J. Li, F. Bosché, C. X. Lu, and L. Wilson. Occlusion-free orthophoto generation for building roofs using uav photogrammetric reconstruction and digital twin data. In *Proceedings of the 40th ISARC*, pages 371–378, Chennai, India, July 2023. doi:10.22260/ISARC2023/0051.

[9] J. Li, B. Tao, F. Bosché, C. X. Lu, and L. Wilson. Extracting roof sub-components from orthophotos using deep-learning -based semantic segmentation. In *Proceedings of the 41st ISARC*, pages 675–682, Lille, France, June 2024. doi:10.22260/ISARC2024/0088.

[10] B. Kim and S. Cho. Image-based concrete crack assessment using mask and region-based convolutional neural network. *Structural Control and Health Monitoring*, 26(8):e2381, 2019. doi:10.1002/stc.2381.

[11] Y. Wu, H. Kim, C. Kim, and S. H. Han. Object recognition in construction-site images using 3d cad-based filtering. *Journal of Computing in Civil Engineering*, 24(1):56–64, 2010. doi:10.1061/(ASCE)0887-3801(2010)24:1(56).

[12] M. Z. Shanti, C. Cho, B. Garcia de Soto, Y. Byon, C. Y. Yeun, and T. Y. Kim. Real-time monitoring of work-at-height safety hazards in construction sites using drones and deep learning. *Journal of Safety Research*, 83:364–370, 2022. doi:10.1016/j.jsr.2022.09.011.

[13] H. Kim, H. Kim, Y. W. Hong, and H. Byun. Detecting construction equipment using a region-based fully convolutional network and transfer learning. *Journal of Computing in Civil Engineering*, 32(2):04017082, 2018. doi:10.1061/(ASCE)CP.1943-5487.0000731.

[14] M. Ahmadi, A. G. Lonbar, H. K. Naeini, A. T. Beris, M. Nouri, A. S. Javidi, and A. Sharifi. Application of segment anything model for civil infrastructure defect assessment, 2024.

[15] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698, 1986. doi:10.1109/TPAMI.1986.4767851.

[16] J. MS. Prewitt et al. Object enhancement and extraction. *Picture processing and Psychopictorics*, 10(1): 15–19, 1970.

[17] I. Sobel. Neighborhood coding of binary images for fast contour following and general binary array processing. *Computer Graphics and Image Processing*, 8(1):127–135, 1978. doi:10.1016/S0146-664X(78)80020-3.

[18] D. Marr, E. Hildreth, and S. Brenner. Theory of edge detection. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 207(1167): 187–217, 1980. doi:10.1098/rspb.1980.0020.

[19] S. Xie and Z. Tu. Holistically-nested edge detection. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, December 2015. doi:10.1109/ICCV.2015.164.

[20] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko. End-to-end object detection with transformers. In *Computer Vision – ECCV 2020*, pages 213–229, Cham, 2020. doi:10.1007/978-3-030-58452-8_13.

[21] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W. Lo, P. Dollár, and R. Girshick. Segment anything. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3992–4003, 2023. doi:10.1109/ICCV51070.2023.00371.

[22] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88:303–338, 2010. doi:10.1007/s11263-009-0275-4.

A Large Language Model as an Assistant for Software-Independent Structural Analytical Model Review and Editing

Justin S. Lee[1] and Ghang Lee[1,2*]

[1]Department of Architecture and Architectural Engineering, Yonsei University, Seoul, Republic of Korea
[2]Institute for Advanced Studies, Technical University of Munich, Munich, Germany
*Corresponding author
justinlee@yonsei.ac.kr, glee@yonsei.ac.kr

Abstract

While structural analysis software facilitates the modeling of complex structures, it requires significant effort to master these tools. With advancements in large language models (LLMs), this study explores the potential of leveraging an LLM as a software-independent assistant for reviewing and editing structural analytical models. The proposed framework was tested using GPT-4o and an analytical model from Midas Gen, focusing on tasks such as model information extraction and editing. The results demonstrated the framework's potential, achieving an accuracy of 91% when a system prompt with model data ontology information was used, compared to a 20% accuracy without the use of any prompt engineering.

Keywords –

Analytical Models; Structural Models; LLM; System Prompt; Code Interpreter

1 Introduction

Computer based structural analysis has become an indispensable part of the structural engineering profession. The emergence of structural analysis software has enabled engineering designs of much higher complexity and greater scale. However, even with the evolution of structural analysis software over the years, the general modeling approach has not evolved dramatically – they still rely heavily on heuristics and trial-and-errors which makes the structural design optimization a human-centric process [10]. In this task, the extent of the engineer's ability to efficiently control the analytical model constructed within the analysis software becomes an important factor. Especially for engineers who perform designs for multiple industries or regions, they may be required to obtain proficiency in multiple structural analysis software which is no small feat. Furthermore, even if adequate software proficiency is reached, engineers are usually limited to the native features of the software. This lack of flexibility in analytical model control leads to tasks with high interaction cost, especially when dealing with complex models.

Large language models (LLMs) have made great strides in recent years. With carefully designed prompts, LLMs demonstrate great performance on textual classification, information extraction, and summarization [11]. Considering the building blocks of any structural analytical model are simple textual and numeric arrays, an LLM based user interaction method for analytical model control can potentially provide much more freedom in communicating the engineer's intent to the structural analysis software. In addition to this already powerful textual processing abilities, multimodal LLMs allow for amalgamation of diverse set of data types [9].

This study aims to lay the foundation for developing a software-independent structural analytical model review and editing method using an LLM for future research endeavors. To perform a structural analysis, a structural model and its properties must be prepared, reviewed, and adjusted for the targeted analysis. The process of establishing a structural model frame can be broken down as shown in Figure 1. This study focuses particularly on the last two steps: review and editing steps. For the review step, the ability to select specific groups of elements and extract various information about them is evaluated. For the editing step, the capability to adjust element attributes, such as section types or boundary conditions, for a specified group of elements is assessed.

Selecting a specific group of elements within the modeling software, or filtering, is an essential feature in performing review and editing steps of any structural analytical model. Filtering is generally done with natively offered basic filter options. There are software packages that allow users to build custom filters such as SAP2000 and its advanced filter features using a

structural query language (SQL); however, while such approach widens the types of filters that can be used, it still does not offer enough flexibility to cover all possible scenarios and are often time consuming to build, test, and deploy when needs arise. For example, if a user needs to filter beams that are not directly connected to any columns to select all sub-beams or joists, such filtering would generally not be covered by native software filters. This element selection via filtering must be followed by the ability to apply edits to the element attributes, if necessary, completing the review and editing process.

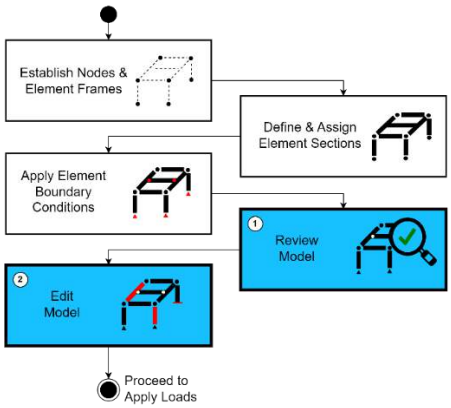


Figure 1. Review and editing steps of structural analytical modeling process

For testing, the geometric and element feature data of an analytical model from the most widely used structural analysis software in South Korea, Midas Gen, was uploaded to ChatGPT-4o along with a system prompt and model data ontology information to enhance its performance. Then, a series of tasks were given as user prompts where corresponding results were evaluated.

2 Related Studies

This section of the study consists of two parts: The first part will review literature on LLM applications in architectural design process where the primary focus will be LLM augmentation on design software. The second part will be an overview of artificial intelligence (AI) applications in structural engineering where model generation and design optimization will be discussed.

2.1 LLM Augmentation on Design Software

Natural language (NL) based user expressions allow for much more complex forms of queries to be used in software control. One application strategy is to extract user's intentions via LLM and map specific functions or features of the design software. For instance, an NL query to search for a specific element or additional information about the element from a building information modeling (BIM) entity, such as a Revit model, has been proposed [11]. In this literature, the NL query is first processed by the LLM to map its intent to the corresponding parameters and value dictionaries. Then, the appropriate search functions are called to interact with the application programming interface (API). However, this approach limits its focus on element searching and does not account for design generation or information management.

Jang et al. expands the prompt-based user interface concept to handle Revit element manipulation with their proposed framework Natural-language-based Architectural Detailing through Interaction with AI (NADIA) [4]. In this framework, the user request via the task prompt is first categorized based on its task type and relevant component type. Then, the necessary BIM authoring is executed to generate, modify, retrieve, or delete the relevant Revit object. This functionality more closely resembles the analytical model control which will be discussed in this study; however, the main difference will be that the geometric data and some of the element feature data will be entirely transferred to LLM for this study where analytical model data will be manipulated directly by LLM without any reliance on existing software functions or features.

2.2 AI Application in Structural Engineering

Design optimization is expectedly a popular topic of interest in many recent studies involving AI and structural engineering. As per most engineering design problems, structural design problems involve a large set of input variables and also many output variables to be optimized. Structural optimization problems can be categorized into cost, performance, environmental impact, and multi-objective optimization (MOO) problems [8]. For example, reinforced concrete section design is an MOO problem that is an excellent candidate for an optimization method using an artificial neural network (ANN). ANN can be used to explore various pareto curves to flexibly optimize what would traditionally be a computationally expensive problem with much less effort [2]. However, design process optimization, including user-software interaction, unfortunately falls outside of the purview of this popular view of engineering optimization.

Although it is not an optimization method, generative AI presents new opportunities for structural topology generation. AI models can be used to extract patterns from a large design dataset and reproduce resembling structural layouts accordingly [7]. Generative Adversarial Network (GAN) can be used to generate column and beam models based on 2D images [6], furthermore, physical rules can be implemented alongside GAN to enhance the quality of the generated structural topology [1]. Although these novel strategies

contribute to enhancing the qualities of the AI generated analytical models, the need for structural engineers to closely review and edit these models still remains.

3 Proposed Method

3.1 Analytical Model Data

For most commercial structural analysis software, such as Midas Gen or ETABS, analytical model data is typically stored in a series of 2D arrays or tables. Analytical model data composition for Midas Gen is summarized in Table 1. These tables reference each other – for example, the element table does not contain the coordinates of these elements explicitly, but rather they specify the two nodes each element is connected to. A part of the "Elements" table from Midas Gen is shown in Table 2.

Table 1. Midas Gen model data tables

Table Name	Contents
Elements	Element Number / Material / Section / Orientation / Node Connectivity
Nodes	Node Numbers / Coordinates
Materials	Material Number / Name / Type / Sub-type
Supports	Node Assignment / Translation & Rotation Release
Sections	Section Number / Name / Shape / Sizes

Table 2. "Elements" table from Midas Gen

Element ID	Material	Section	Beta Angle	Node 1	Node 2
1	1001	101	90	1	10
2	1001	103	0	1	11
3	1001	101	90	2	11
4	1001	103	0	3	11
5	1001	101	0	3	12

Compared to typical architectural models, structural analytical models are much simpler and hold vastly less information as they are based on directional stick-and-node graphs. This simpler and smaller model data can be uploaded to ChatGPT-4o for more direct control of the model data without relying on any built-in software functionalities for model manipulations. To ensure that ChatGPT-4o properly understands the data structure, an ontology representation file was created with Protégé as shown in Figure 2.



Figure 2. Midas Gen model data ontology

3.2 Data Processing

When tabular data, such as CSV files, is uploaded to ChatGPT-4o, it is stored in its memory as DataFrames which is a table-like data format provided by the Python library pandas. When the user prompts inquire about the information stored in the DataFrames, ChatGPT-4o then writes and executes its own Python codes to extract and edit the data (Figure 3). This functionality of ChatGPT-4o is called Code Interpreter. These codes are then immediately available to the user with the response. An example of the code generated by ChatGPT-4o is shown in Figure 4 in pseudo-code for clarity.

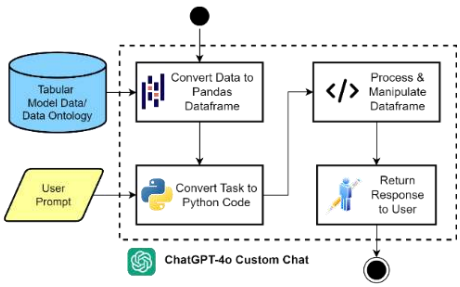


Figure 3. Tabular data processing of ChatGPT-4o



Figure 4. ChatGPT-4o generated code for the task 'How many columns are on the first floor?'

Because the analytical model review and edit assistant proposed in this study relies on ChatGPT-4o to correctly interpret the user tasks and successfully extract and manipulate the relevant information, the system heavily depends on ChatGPT-4o's ability to write effective Python codes.

3.3 System Prompt

The system prompt was developed through a trial-and-error approach, iteratively refining its content based on errors until ChatGPT-4o could visualize the analytical model successfully. After several iterations, the following three main components were selected to be included in the textual part of the system prompt: general role, geometric context, and element definitions as shown in Table 3. The general role explains the expected role of ChatGPT-4o followed by the descriptions of the CSV model data files and the data ontology file. The geometric context explains the contextual meaning of the coordinate system, for example, z = 0.0 means ground level. Finally, the element definitions describe the definitions of various structural components in terms of their nodal coordinates. The element definitions component of the system prompt is important because the user is expected to refer to the elements by the element types, such as beams or columns; however, ChatGPT-4o, in its native state, was not able to identify these element types solely based on the model data.

Followed by the textual portion of the system prompt, the data ontology file was attached to the system prompt of ChatGPT-4o; however, the CSV model data files were not uploaded with the system prompt for two reasons: first, the proposed framework needed to work with any model data, not just one set of CSV files. Second, ChatGPT-4o was unable to keep and recall all the information properly when textual system prompt, the ontology file, and the CSV files were all uploaded within the system prompt at once. In other words, ChatGPT-4o was unable to process multiple instructions and files all under one request as LLMs generally struggle with queries with multiple objectives and perform better when interrelated tasks are provided separately [3].

3.4 Framework Assembly

The analytical model review and edit framework is assembled using the custom chat feature of ChatGPT: The system prompt is first inserted in the instructions window, followed by uploading the data ontology file (.rdf format) prepared using Protégé. This custom chat is saved and initiated as an instance, expecting to receive the analytical model CSV files extracted from Midas Gen. For this study, the CSV files were directly uploaded to ChatGPT-4o's web interface. Once CSV files are processed, it is then ready to take on the user tasks. The overall flow of this framework is shown in Figure 5.

Table 3. The textual portion of the system prompt

Category	Contents
General Role	• You are a virtual assistant to help review and edit an analytical structural model. The model data is stored in several CSV files. These files reference each other. You can combine these files referencing an ontology (.rdf format) file which I will provide also. Let me give you the ontology file first.
Geometric Context	• Ground level is first floor (z = 0.0m). • The second floor is one story above the first floor. • Roof level is the highest level.
Element Definitions	• Columns are vertical elements (change in ONLY 'z' coordinates). • Beams are horizontal elements (no change in 'z' coordinates between nodes). • Braces are diagonal elements (change in ('x' or 'y') coordinates AND 'z' coordinates). • 'Beta_angle' means local orientation, not global.

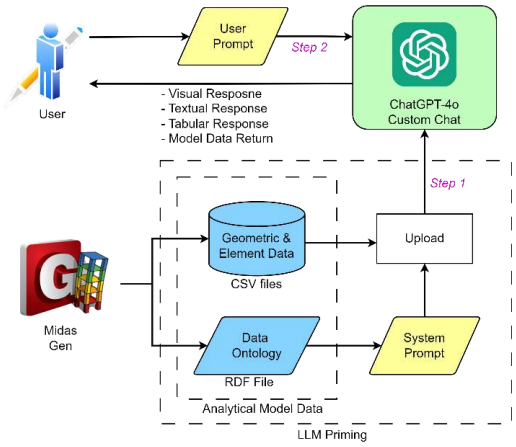


Figure 5. Structural analytical model review and edit framework

4 Validation

To evaluate the analytical model review and edit framework, this study focused on its ability to accurately execute user queries. A diverse set of tasks, including visualization, review, and editing were designed to

evaluate the key capabilities of the framework. The task set and the descriptions are summarized in Table 4.

4.1 Test Subject

A modified version of the American Society of Civil Engineers (ASCE) benchmark structure from the Earthquake Engineering Research Laboratory at the University of British Columbia (UBC) [5] was modeled in Midas Gen. Minor modifications were introduced to this model to add more geometric complexity and incorporate additional section types.

Table 4. Validation Task Set

Task Type	Task Sub-Type	Description
Visualization	Create 3D view	Render a 3D view of the model.
	Color by element type	Show beams, columns, braces and support nodes in assorted colors.
Extraction for Review	Single table reference	Inquire about information accessible from a single table.
	Multiple table reference	Inquire about information that requires accessing multiple tables that reference each other.
Extraction and Editing	Based on prompts	Edit model data based on user prompts.
	Based on images	Edit model data based on user prompts and image files.

4.2 Evaluation Task Categories

4.2.1 Visualization Tasks

The purpose of the visualization task category is to verify that the custom chat initialization has been correctly performed. Although these tasks do not simulate any potential user inquiries, for the validation of the study, it is a key step to ensure that the framework is able to assemble the model data correctly. The geometric assembly is validated by the 3D render task and the element classification is validated by the element color assignment task (Figure 6).

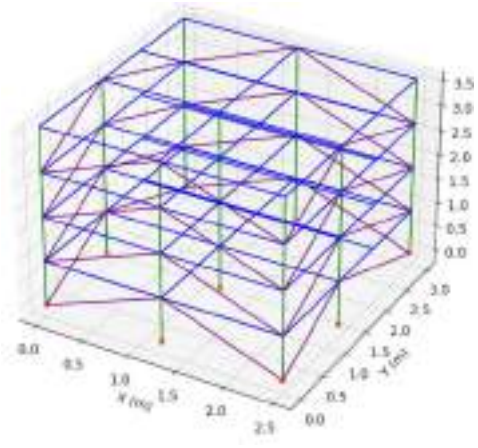


Figure 6. A successful visualization task result

4.2.2 Extraction for Review Tasks

The extraction for review task category is to simulate the user inquiries to extract model information for the purpose of reviewing various model attributes, such as element quantities, section types, material types or connection types. This task category includes basic filtering tasks to ensure that the LLM based approach can effectively replace the basic filters offered by most modeling software as well as more challenging tasks such as boundary condition-based filtering and filtering based on neighboring element types such as adjacency with columns.

For the single table reference tasks, the requested information is directly accessible from one of the DataFrames, while for the multiple table reference tasks, ChatGPT-4o must utilize the data ontology to reference multiple DataFrames to process and generate the appropriate response. An example of a multiple table reference task would be an element length calculation task where the element node numbers refer to the nodes table where the coordinates are stored. The coordinates are then used to calculate the Euclidean distance between the two nodes to evaluate the element length. An image file containing grid label assignment is also uploaded at this stage of testing.

4.2.3 Extraction and Editing Tasks

The extraction and editing task category consists of two important steps: firstly, the framework needs to understand the user request to extract element attributes and make changes to the model data stored in its memory, then secondly, it needs to return CSV files back to the user in a format that is consistent with the original model data. The second step is crucial to allow the CSV files to be loaded back to Midas Gen for an uninterrupted workflow.

The extraction and editing task based on images is to simulate a scenario where structural elements need to be

updated in accordance with drawings, either from architectural or structural design changes. For this task, an image (.png format) file containing a plan view with revised section type labels is uploaded along with the user prompt.

5 Results

A total of 11 tasks were tested 5 times each. In addition, to see the effects of the system prompt containing the textual instructions and the data ontology file, the same set of tests were repeated without any initialization (ChatGPT-4o in its vanilla state) and with the system prompt. In total, 110 task trials were conducted for this study. The test results are summarized in Figure 7 and detailed results along with the task prompts are shown in Table 5.

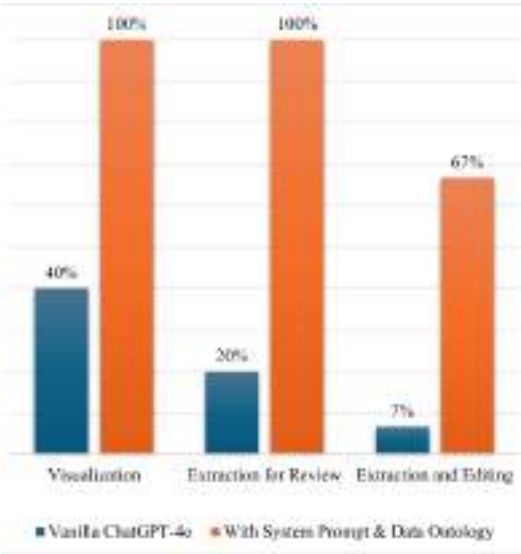


Figure 7 Test results summary

5.1 Results by Task Categories

5.1.1 Visualization Tasks

ChatGPT-4o was able to successfully perform the 3D visualization task 4 out of 5 times even without any assistance from the system prompt; however, it was not able to classify different element types, such as beams, columns, or braces. Although ChatGPT-4o has access to general definitions of what each of those element types are, it was not able to semantically connect the dots between element definitions and geometric orientations in the cartesian coordinate system.

When initialized with the system prompt that contains detailed element definitions in the context of the model geometry, both 3D visualization and color coding by element type tasks were performed perfectly.

5.1.2 Extraction for Review Tasks

The overall success rate of the extraction for review tasks without the system prompt was 20%. A notable result was when asked to check if all beams that are not directly connected to columns also pin connected, the correct response (‘yes’) was given 3 out of 5 times even without the system prompt. To ensure the successful responses were not ‘false positive’ answers, an additional inquiry was made to identify the relevant element numbers for confirmation. With the system prompt, all the tasks were performed perfectly for all trials.

5.1.3 Extraction and Editing Tasks

ChatGPT-4o was able to return the edited CSV files in the correct column arrangements (to allow for direct import back to Midas GEN) even without the system prompt; however, it was able to perform the requested edits correctly for only 1 out of the 15 trials. The most common errors were edits being applied to too many elements around the correct target elements. Prompt-based editing tasks were perfectly performed with the system prompt. Image-based editing tasks could not be performed successfully even with the system prompt as the target elements were always misidentified.

6 Discussion

Recent LLM driven BIM software control studies have generally relied on the software’s natively built-in functionalities for model control. This study demonstrates that leveraging ChatGPT-4o’s code interpreter feature to more directly control analytical model data is a feasible strategy, with one important caveat: the framework requires an elaborate system prompt, including detailed textual instructions along with data ontology information, to function effectively.

6.1 Limitations

While this study provides new insights into a prompt-based analytical model review and editing framework that is free of any predefined software functionalities, there are limitations that must be addressed.

The proposed framework was evaluated using the ASCE benchmark model, which is significantly less complex than most real-world engineering examples. This limitation restricts the generalizability of the results, as the structural simplicity of the ASCE model may not adequately represent the challenges encountered in practical engineering scenarios. In addition, the validation process was conducted with a relatively small and narrowly defined set of tasks, further constrained by a limited number of repetitions.

Table 5. Task user prompts and test results

Task Type	Task Sub-type	User Prompt	Success/Fail: Vanilla ChatGPT-4o	Success/Fail: With System Prompt & Data Ontology
Visualization	3D View of Model	a. Create a 3D visualization of analytical model.	4/5	5/5
		b. Show support nodes in red, and color beams, columns, and braces differently.	0/5	5/5
Extraction for Review	Single table reference	a. How many columns are there on the first floor?	1/5	5/5
		b. What are the element numbers of beams on the ‘roof level’ along grid X1?	0/5	5/5
		c. Give me the element numbers of beams that are not connected to columns.	0/5	5/5
	Multiple table reference	a. Give me the element lengths of all the beams on the third floor.	1/5	5/5
		b. Give me the element numbers and section names of beams that are on the third floor and beam end released	1/5	5/5
		c. Are all the beams that are not directly connected to columns also beam-end-released?	3/5	5/5
Extraction and Editing	Based on prompts	a. Replace all beam sections on the roof level to 'section 101'. Apply this change and return the element CSV in its original column format.	0/5	5/5
		b. Remove all beams with beam end released on the second level. Apply this change and return the element CSV in its original column format.	1/5	5/5
	Based on images	a. Apply beam section changes as per the uploaded PNG on the ‘roof level’. Select the section number according to the section name in the PNG file. Apply this change and return the element CSV in its original column format.	0/5	0/5
		Overall Success Rate	20%	91%

The framework's performance depends entirely on ChatGPT-4o's ability to generate effective Python code for manipulating model data. Also, heavy reliance on an LLM can lead to unreliable reproducibility. As demonstrated by the test results, variations were observed in the responses generated, even when identical input conditions were provided.

Failure to perform image-based tasks is a crucial drawback of this framework. Drawing based structural analytical model revision is a common and time-consuming task, and it would be an invaluable functionality to be implemented to this framework.

6.2 Future Work

Further testing with structural models that more closely resemble the size and complexity of real-world engineering examples with greater number of repetitions will be conducted. A more unified pipeline will also be developed for the proposed framework for proper testing and further development.

For the upcoming research efforts, the effects of the data ontology representation independently from the system prompt will be explored to better quantify its impact on the performance of the proposed framework.

Image-based task performance must be addressed by exploring new strategies such as few-shot learning or more refined system prompt designs.

7 Conclusion

This study explored the potential of LLMs as analytical model review and editing assistants for structural analysis. By using an LLM-driven approach, it demonstrates new ways to improve user interaction beyond traditional tools like rule-based filters. The results showed the framework’s potential, achieving an accuracy of 91% when a system prompt with model data ontology information was used, compared to a 20% accuracy without any prompt engineering.

Key limitations include evaluation with the relatively simple ASCE benchmark model with limited validation task set and repetitions, an overreliance on ChatGPT-4o's capability to produce effective Python code and difficulties in managing image-based model edit tasks. Future work will focus on more rigorous validation and developing a unified pipeline for seamless integration with industry-standard tools. The image-based task performance will also be addressed by exploring few-shot learning and more refined system prompts.

8 Acknowledgements

This work is supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant RS-2021-KA163269 and RS-2024-00407028) and the Hans Fischer Senior Fellowship program at the Technical University of Munich - Institute for Advanced Studies (TUM-IAS) in Germany.

References

[1] Fu B., Wang W., and Gao Y. Physical rule-guided generative adversarial network for automated structural layout design of steel frame-brace structures. Journal of Building Engineering, 86:108943, 2024.

[2] Hong W.-K. and Nguyen D.H. Pareto frontier for steel-reinforced concrete beam developed based on ANN-based Hong-Lagrange algorithm. Journal of Asian Architecture and Building Engineering, 22(6):3535–3551, 2023.

[3] Hu H., Yu S., Chen P., and Ponti E.M. Fine-tuning large language models with sequential instructions. On-line: https://doi.org/10.48550/arXiv.2403.07794, Accessed: 20/11/2024.

[4] Jang S., Lee G., Oh J., Lee J., and Koo B. Automated detailing of exterior walls using NADIA: Natural-language-based architectural detailing through interaction with AI. Advanced Engineering Informatics, 61:102532, 2024.

[5] Johnson E.A., Lam H.F., Katafygiotis L.S., and Beck J.L. Phase I IASC-ASCE structural health monitoring benchmark problem using simulated data. Journal of Engineering Mechanics, 130(1):3–15, 2004.

[6] Kösencig̃ K.Ö., Okuyucu E.B., and Balaban Ö. Structural plan schema generation through generative adversarial networks. Nexus Network Journal, 26(3):409–427, 2024.

[7] Málaga-Chuquitaype C. Machine learning in structural design: An opinionated review. Frontiers in Built Environment, 8:815717, 2022.

[8] Mei L. and Wang Q. Structural optimization in civil engineering: A literature review. Buildings, 11(2):66, 2021.

[9] Qi S., Cao Z., Rao J., Wang L., Xiao J., and Wang X. What is the limitation of multimodal LLMs? A deeper look into multimodal LLMs through prompt probing. Information Processing & Management, 60:103510, 2023.

[10] Zhao P., Liao W., Huang Y., and Lu X. Intelligent design of shear wall layout based on attention-Engineering Structures, 274:115170, 2023.

[11] Zheng J. and Fischer M. Dynamic prompt-based virtual assistant framework for BIM information search. Automation in Construction, 155:105067, 2023.

Roadmap for Exoskeleton Adoption in the Construction Industry: Leveraging Standard Tests and Minimum Criteria

Malcolm Dunson-Todd[1], Mazdak Nik-Bakht[1], Amin Hammad[1,2]

[1]Gina Cody School of Engineering and Computer Science, Concordia University, Canada
[2]Corresponding author: amin.hammad@concordia.ca

Abstract –

Occupational exoskeletons (OEs) offer significant potential to reduce the incidence and severity of musculoskeletal disorders (MSDs) in the construction industry. Despite this potential, OE adoption remains low due to unique challenges in this industry. By addressing the limitations of previous OE adoption frameworks and roadmaps, this paper proposes a framework to guide OE design, evaluation, and implementation specifically for the construction industry. The novel elements of the framework are: (1) incorporating sets of general and trade-specific standard OE test methods, and (2) incorporating trade-specific criteria for minimum OE design requirements, sufficient evidence of OE efficacy, and sufficient evidence of OE effectiveness. A roadmap is extrapolated from the proposed framework that recommends future research on trade-specific standard test creation and trade-specific stakeholder surveys to inform minimum criteria. Finally, the results of a pilot field study on a back-support OE for reinforcing ironworkers are linked to the framework as an example of its utility. The goal of this research is to define a process for generating a comprehensive body of scientific evidence that can support large-scale OE adoption in the construction trades.

Keywords –

Occupational Exoskeleton; Construction

1 Introduction

Trade workers in the construction industry suffer from work-related musculoskeletal disorders (MSDs). MSDs in the United States construction industry in 2016 were higher than the average of all other industries combined [1]. 46% of United States construction workers in 2019 self-reported MSD symptoms [2, 3]. Research in Québec, Canada in 2024 suggested that occupational exoskeletons (OEs) had the potential to have addressed 50% of all MSD cases that were observed in a large construction contractor's work force over the three-and-a-half-year study period [4]. The adoption of OEs by the construction industry, however, remains low.

Previous research has investigated the necessary elements for successful industry adoption for any type of occupation intervention. Van der Beek et al. [5] proposed a general roadmap for the adoption of occupational interventions which involves six steps: (1) Quantify MSD incidence and severity, (2) Identify the risk factors associated with MSDs, (3) Investigate the underlying mechanisms relating risk factors to MSDs, (4) Develop an intervention, (5) Evaluate the intervention, and (6) Implement the effective intervention. Step 5 returns to Steps 2, 3, or 4 if there is no evidence of an intervention's effectiveness when evaluated, thereby making the roadmap iterative. Step 6 returns to Step 1, ideally showing a reduction in MSD incidence and severity in a population of workers, making the roadmap circular. As an example-use of their roadmap, Van der Beek et al. [5] presented the hypothetical development of an intervention for bricklayers: change in workstation height.

As the interest in adopting OEs at a large scale has grown since the mid-2010s, frameworks have been proposed specifically for the design, evaluation, and implementation of OEs [6–12]. Pesenti et al. [6], De Bock et al. [7], and Golabchi et al. [8] categorized measures of an OE's effects as either muscular, biomechanical, metabolic, functional, or subjective, and recommended task-specific and posture-specific evaluations. De Bock et al. [7] proposed technology readiness levels (TRLs) to guide the choice of test environments, including laboratory efficacy studies, field effectiveness studies, and simulation studies. Balancing comprehensive evaluation with practicality, Hoffmann et al. [11] recommended standardizing evaluation methods and consolidating measures of OE effects. This has since been exemplified by standard test methods from the ASTM Committee F48 on Exoskeletons and Exosuits (ASTM F48).

To realize such frameworks, roadmaps towards the goal of large-scale OE adoption have also been proposed. Crea et al. [13] proposed a roadmap to facilitate OE adoption in any industry, e.g., manufacturing, logistics, or construction. The goal of the authors was to provide a path toward a holistic body of scientific evidence that is necessary for large-scale OE adoption [13]. A limitation,

however, of this roadmap, and of all the above frameworks, is the lack of specific mechanisms or processes by which to generate such a body of scientific evidence.

Clear processes for OE adoption may be especially important for the construction industry because of the unique challenges it faces. Zhu et al. [14] and Nnaji et al. [15] found that, compared to the manufacturing and logistics industries—where OE design and evaluation is more mature—the construction industry is unique because of the inherent variability and intensity of the tasks and conditions that construction workers face within each trade, and because of the variability between trades. These challenges complicate OE adoption in the construction industry [14, 15].

Previous work has attempted to addressed these unique challenges in construction by proposing realistic evaluation methods towards the development of standard tests, including the research from Kopp et al. [16] on the welding and timber framing trades. Similarly, Dunson-Todd et al. [17] proposed a realistic test method for evaluating back-support OEs for reinforcing ironworkers (rebar workers), towards a standard test that can be used to generate sufficient evidence of OE efficacy, safety and acceptance, as recommended by [11]. The question remains, however, of how these standard tests for the construction industry can fit within a larger framework for generating the holistic body of scientific evidence required for OE adoption, as proposed by [13].

Informed by the previous work discussed above, the current paper proposes such a framework and uses it to extrapolate a roadmap for future research required in the construction industry. The framework addresses the specific challenges inherent in the construction trades while providing a comprehensive path towards OE adoption. To provide insights into the utility of the proposed framework, the results of a pilot field study on a back-support OE for rebar workers are discussed. The pilot field study is a continuation of previous work toward a standard test for the rebar trade [17]. Examples of real-world construction OE challenges found from the pilot field study are linked to the framework to demonstrate its value. Section 2 presents the framework for OE adoption in the construction industry and the roadmap to realize it. Section 3 presents the results of the pilot field study.

2 Framework and Roadmap for OE Adoption in the Construction Trades

Informed by the previous research described in Section 1, Figure 1 shows the proposed framework for OE adoption in the construction industry. The framework is divided into Steps 1 to 4, and sub-steps, e.g., 2(a), 2(b), etc. Much of the framework has been adapted from previous research: Step 2, Design, is adapted from Van der Beek et al. [5], and Step 3, Evaluation, is divided into short-term and long-term laboratory and field evaluation, as proposed in previous frameworks and roadmaps, e.g., [13]. By convention, long-term can be defined as a study period greater than 3 months, as in [18]. Short- and long-term simulation have been included in Steps 3(a) and 3(b) of the framework. Simulation has been used to estimate an OE's immediate effects on the biomechanical system of the body in previous research, e.g., [19]. Although no methods exist for simulating the long-term effects of an OE, future longitudinal field testing may generate evidence that can be used to verify simulations of an OE's long-term effects.

Two elements of the framework are novel: (1) proposing a set of general and trade-specific realistic standard laboratory tests for each trade, and (2) proposing three minimum requirement decision points informed by a trade-specific industry stakeholder survey, further referred to in this paper as a trade-specific survey (TS). These novel elements are outlined with a bold border in Figure 1. The framework, and a roadmap to realize it, are described in detail in the following subsections.

2.1 OE Design

Van der Beek et al.'s [5] proposed framework for designing an intervention rested on an evidence-based understanding of the mechanisms behind MSDs in a population of workers. As shown in Figure 1, this has been adapted as follows: Step 1, looking to statistics to understand MSD incidence and severity in a construction trade population, (e.g., as in [4]); Step 2(a), conducting epidemiological observational studies to identify the MSD risk factors present in a given trade; Step 2(b), conducting laboratory and field studies to identify the etiological mechanisms linking risk factors to MSDs; and Step 2(c), a research and design process that targets these mechanisms to develop a new OE or modify an existing OE for the benefit of a specific construction trade.

2.2 Set of General and Trade-Specific Standard Laboratory Tests

As recommended by Gutierrez et al. [20], controlled laboratory studies that simulate construction tasks and environments will help identify the most appropriate situations to implement OEs. The Delphi survey from Mahmud et al. [21] identified seven potential solutions regarding policy and regulatory barriers to OE adoption, two of which were (1) scientific data to drive regulation and policy development, and (2) forming industry standard practices. These solutions can be achieved by collecting benchmarking data from a set of both general and trade-specific standard tests.

As an example, the welding trade would benefit from

a general standard test to evaluate OE capabilities in confined spaces, and ASTM F48 has already published such a test method: F3523 − 21 [22]. The trade-specific requirements of an OE for a welder, however, cannot be fully met by a set of general standard tests. A realistic trade-specific standard efficacy test for welders, as proposed by Kopp et al. [16], can be used in combination with general standard tests to generate sufficient evidence for OE efficacy. In this case, "sufficient" can be defined by the results of a TS for the welding trade, as will be described in Section 2.3. As a second example, a candidate for a general test method is the evaluation of an OE design's heat retention, which is especially applicable to all construction trades that work in outdoor and non-temperature-controlled environments, such as the rebar trade. The proposed trade-specific test method from Dunson-Todd et al. [17] could be used in combination with general standard tests, including a general test on heat retention, to build a set of test methods that comprehensively evaluates an OE design for the rebar trade.

The challenge of creating useful standard efficacy tests is in balancing the comprehensiveness of a set of measurements and conditions with the practicality of implementing them [11]. An example of balancing comprehensiveness of experiment conditions with implementation practicality comes from De Vries et al. [23]. The authors limited the scope of their evaluation for plasterers to three main ceiling and wall plastering activities. In this way, the authors limited test complexity while still capturing most of the ergonomic risk in the

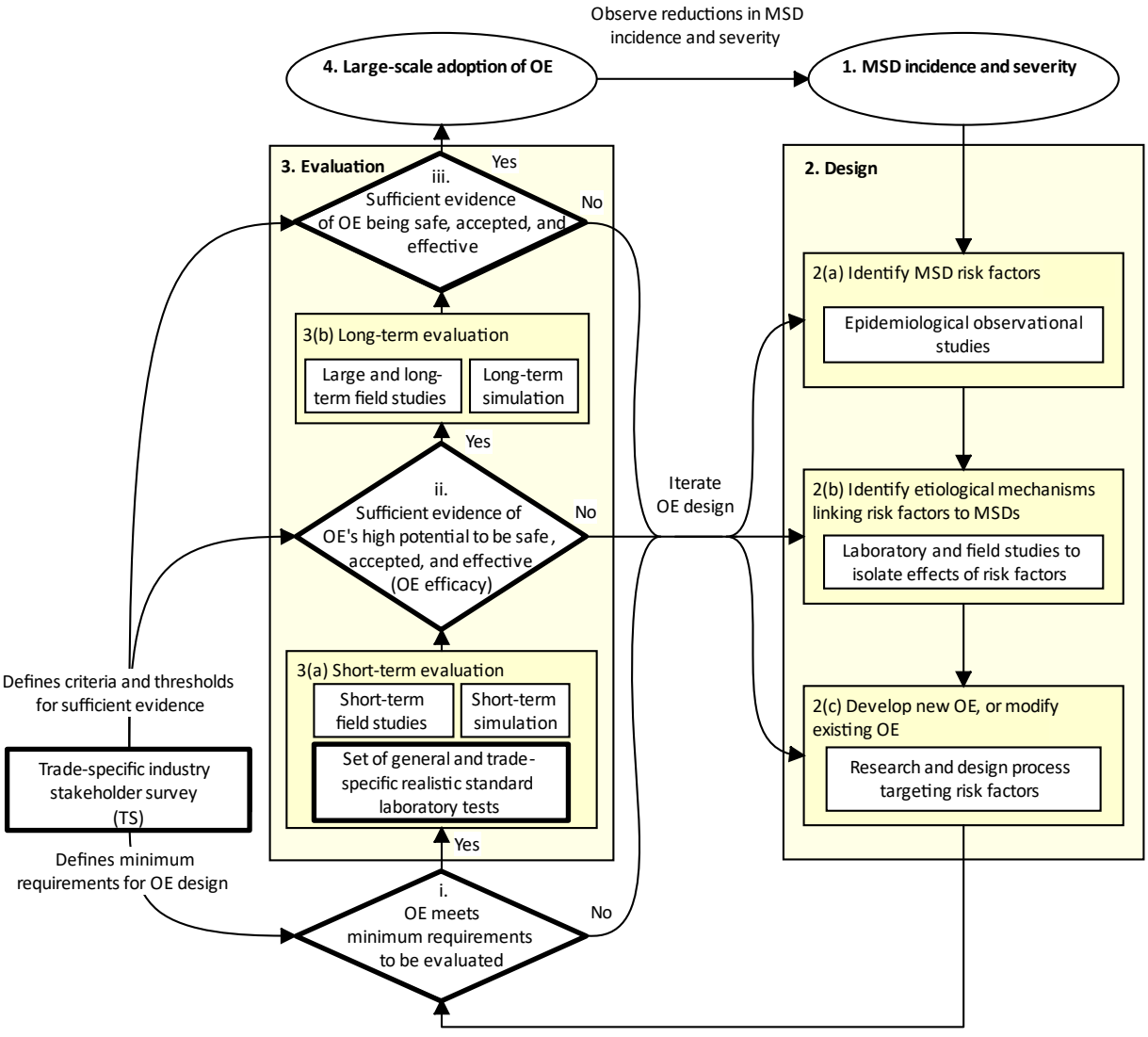


Figure 1. Framework for OE adoption in a specific construction trade

plastering trade. Moreover, the conditions captured the task variability inherent to the trade: results showed a significant reduction in perceived effort during overhead tasks but no significant reduction while applying gypsum to walls, likely due to the variable nature of the latter task. Future work can learn from this example of synthesizing the most important features of the environments and activities of a trade into a realistic efficacy test.

2.3 Trade-Specific Surveys to Define Minimum Criteria Decision Points

The framework shown in Figure 1 includes three decision points: (i) Does the OE meet minimum design requirements to be evaluated? (ii) Is there sufficient evidence of the OE's high potential to be safe, accepted, and effective (i.e., of OE efficacy)? (iii) Is there sufficient evidence of the OE being safe, accepted, and effective? Because of the variation of construction tasks and environments between each construction trade, the criteria for meeting each of these decision points will vary between trades. The criteria for passing each decision point in the framework may therefore be most readily determined through a TS that includes all relevant stakeholders for a given trade. Previous research (e.g., [20, 24, 25]) involved conducting industry surveys to collect construction stakeholders' perspectives. Various survey methods have been used, including Delphi, as in Nnaji et al. [15], in which a panel of experts come to a consensus agreement regarding questions that are important to the construction industry. Similarly, a Delphi TS can be used to inform criteria for passing each decision point in the proposed roadmap in Figure 1.

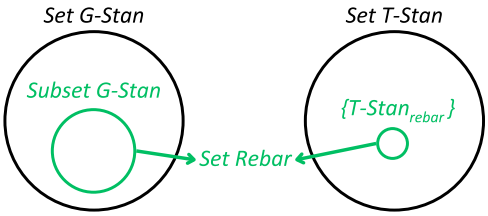
The roadmap from [13] recommended incorporating all stakeholder perspectives into the design and evaluation of an OE and included the following groups to encompass all OE stakeholders: (1) workers, (2) unions and worker associations, (3) corporate management and corporate decision makers, (4) health professionals, including but not limited to ergonomists, kinesiologists, and occupational medical doctors, (5) insurance companies, (6) policy makers, and (7) certifying bodies and standards development organizations (SDOs), e.g., ASTM F48. For the construction industry, two additional stakeholders are trade school instructors and students. These stakeholders can be added for inclusion in a Delphi TS, as they have a strong influence on the future of any major construction trade.

To further leverage the use of minimum criteria for design requirements, and for sufficient evidence of efficacy and effectiveness, a useful tool for OE adoption may be a minimum criteria labeling system. Labels that signify an OE's adherence to minimum design, efficacy, or effectiveness criteria would accelerate OE adoption in the construction industry, shifting the burden of finding suitable OEs away from individual contractors to the evaluation process outlined in Figure 1. This labeling system could be administered by an SDO to provide the industry with clear indicators of the suitability of any given OE for a specific trade.

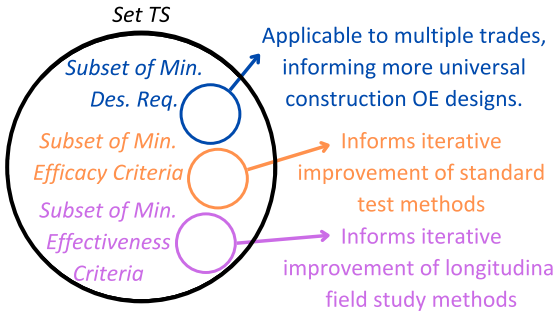
Labels may also be useful for market-neutral OE evaluation. The idea of market-neutral reporting used by Kopp et al. [16] promotes participation in the authors' proposed and implemented *Exoworkathlon*. The authors' reporting of aggregate results from multiple OEs, however, relied on the OEs being relatively similar passive shoulder-support designs with similar torque profiles [16]. This method of market-neutral reporting would therefore be impossible for a group of OEs that have significantly different designs and biomechanical effects. With defined criteria of sufficient OE efficacy and effectiveness—(ii) and (iii) in Figure 1, respectively—multiple disparate OE designs can achieve efficacy and effectiveness labels. Labels could thereby provide market neutrality while encouraging OE design diversity.

2.4 Roadmap Toward Standard Tests and Minimum Criteria

To realize the framework for construction OE adoption shown in Figure 1, a roadmap, the goals of which are shown in Figure 2, is proposed as follows.



(a) Goal of standard testing research effort.



(b) Goal of TS research effort.

Figure 2. Roadmap goals toward OE adoption

As outlined in the introduction, previous research has designed OEs for construction trades and has evaluated their efficacy in laboratory studies. In coming years, ongoing longitudinal field studies on construction OEs may be published. The non-novel steps of the framework

(without a bold outline in Figure 1) have therefore been touched upon by previous or ongoing research, with varying degrees of success, and for a limited number of construction trades. Future research may increase the probability of success for construction OE adoption by focusing on the novel components of the framework, which are outlined in bold in Figure 1 and described in Sections 2.2 and 2.3.

The set of general standard tests (*G-Stan*) and trade-specific tests (*T-Stan*) described in Section 2.2 can be represented visually, as shown in Figure 2(a). A set of standard tests for each construction trade can come from the union of a subset of tests from *G-Stan* with a subset of tests from *T-Stan.* As an example for the rebar trade, Figure 2(a) shows *Set Rebar* as the union of multiple general tests, *Subset G-Stan*, and one trade-specific test, $T\text{-}Stan_{rebar}$.

Set G-Stan, *Set T-Stan*, and *Set Rebar* can also be represented as equations, as shown in Equations (1), (2), and (3), respectively:

$$Set\ G\text{-}Stan = \{G\text{-}Stan_1, G\text{-}Stan_2, \dots, G\text{-}Stan_n\} \quad (1)$$

$$Set\ T\text{-}Stan = \{T\text{-}Stan_1, T\text{-}Stan_2, \dots, T\text{-}Stan_{nt}\} \quad (2)$$

$$Set\ Rebar = Subset\ G\text{-}Stan \cup \{T\text{-}Stan_{rebar}\} \quad (3)$$

In Equation (1), *n* is the total number of general standard tests. In Equation (2), *nt* is the total number of construction trades. Zhu et al. [14] classified 16 separate construction trades and their potential for OEs. An *nt* greater than 16 may be necessary, however, as some trade categories from publicly available occupational statistics are too broad. For example, "ironworker" encompasses rebar assembly and steel beam assembly—arguably two sub-trades involving significantly different tasks, environments, and ergonomic risk factors.

The set of trade-specific industry stakeholder surveys (*Set TS)* described in Section 2.3 can also be represented visually, as shown in in Figure 2(b). *Set TS* would provide a benchmark for minimum OE design requirements and useful measures of an OE's efficacy (i.e., evidence for Decision Point ii in Figure 1), and effectiveness (i.e., evidence for Decision Point iii in Figure 1). As represented in Figure 2(b), subsets of *Set TS* can be used as benchmarks to guide the design of future OEs that optimize capabilities and features to meet the requirements of multiple trades. Research conducting TSs would therefore support the development of OEs that are more universally applicable in the construction industry.

Set TS can also be represented as an equation, as shown in Equation 4:

$$Set\ TS = \{TS_1, TS_2, \dots, TS_{nt}\} \quad (4)$$

As in Equation (2), *nt* in Equation (4) is the total number of construction trades. To efficiently use the resources dedicated to researching OEs for the construction industry, the largest construction trade populations with the highest MSD incidence and severity should be prioritized in the pursuit of the goals shown in Figure 2.

3 Pilot Field Study on Rebar Installation

As discussed in Section 2.2, Dunson-Todd et al. [17] proposed a realistic trade-specific efficacy test method for evaluating back-support OEs for rebar workers. The proposed test method from [17] is currently under review by ASTM F48 to potentially contribute to a new standard. Following this research, a pilot field study was conducted to evaluate the same OE as in [17] in field conditions.

Three male construction workers participated in the pilot study, with an average (standard deviation) age of 45.7 (4.0) years, and an average (standard deviation) of 21.0 (2.6) years of experience in construction. The Research and Ethics Unit at Concordia University approved all ethics documents and surveys. The participants constructed a reinforced concrete slab foundation over the course of five days at an electrical substation construction site. Their tasks included assembling rebar, assembling formwork, and pouring and consolidating concrete. Figure 3 shows the construction site conditions and an example of a participant tying rebar while wearing the Biolift back-support OE. Although the sample size of participants was small, the pilot field study was aimed primarily at providing suggestive results on challenges regarding OE design and evaluation in construction. To that end, the following subsections discuss insights from the pilot field study regarding the utility of the design and evaluation framework proposed in Section 2.

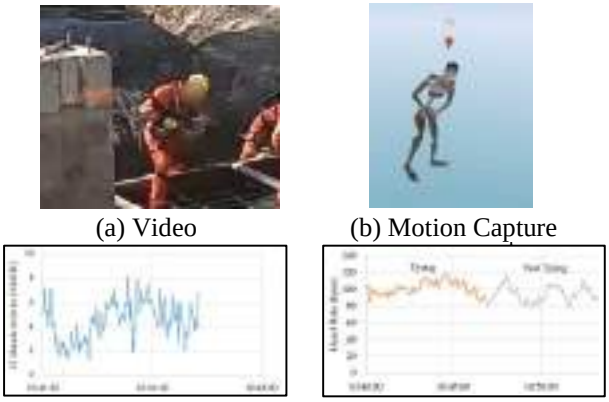


(a) Construction site (b) Tying task

Figure 3. Pilot field study conditions.

3.1 Challenges with Measuring OE Effects

As in Dunson-Todd et al. [17], electromyography (EMG) was used to record muscle activity of the lower back, electrocardiography (ECG) was used to record heart rate as a measure of physiology, and inertial measurement units (IMU) were used to record full-body motion capture (i.e., joint angles and segment positions). Figure 4 shows the three sensor data types synchronized to video reference of a participant tying rebar.



(a) Video (b) Motion Capture

(c) Lower-Back EMG (d) Heart Rate

Figure 4. Synchronized sensor data from the pilot field study.

Although the small number of participants and quantity of objective data was too low to be statistically significant, insights were gained regarding using these measures of OE effects in the field. The lack of control over when and for how long each task was performed created challenging conditions for collecting data in both the control (No OE) and intervention (With OE) conditions for each participant.

Data collection was also challenging due to environmental conditions, where hot days delayed construction, induced noise in muscle activity data, and made frequent readjustment of sensor placement necessary due to the participants sweating. Future field studies would benefit from more practical measures, and from measures that are more directly correlated with safety, acceptance, and effectiveness. The criteria created by a TS for the rebar trade could provide insight into what is sufficient evidence of efficacy and effectiveness, and therefore, what objective measures of an OE's effects are required.

3.2 Iterative Design Process

As examples of the iterative design process shown in Figure 1, the OE manufacturer has implemented design improvements informed by [17] and by internal studies, including: (1) providing a suitable replacement for a rebar worker's tool belt, (2) lowering the position of the lower-back pad and dividing it into two pieces on either side of the spine to improve comfort, and (3) providing more secure thigh strap fixtures to prevent strap misalignment.

The pilot field study exposed other potential OE design requirements that were not anticipated in the study from Dunson-Todd et al. [17]. Figure 5 shows participants' responses to a 5-point Likert survey adapted from [17]. Although opinions on support and on balance in the steep environment were positive, all participants found the heat retained by the OE uncomfortable. As discussed in Section 2.3, low heat retention—achieved through integrated cooling, low contact area, or other features—is a promising minimum design requirement for many trades.

Two unanticipated design requirements stemmed from the concept of OE acceptance: (1) A design that allows for both men and women to use the restroom without removing the OE, and (2) A design that allows for maintenance to be possible in field conditions without the need for specialized tools. Two additional design requirements stemmed from the safety measures necessary at an electrical substation: (1) fire resistant materials, and (2) covering all metal parts to eliminate the risk of electrical arcing. A further unanticipated design requirement was the OE's compatibility with a kneeling posture while tying, as one participant commented they were uncomfortable while kneeling with the OE activated. This difficulty, however, may stem from the pilot study's ad-hoc conditions. The design of the OE [26] evaluated in the pilot field study was informed by subjective feedback from experienced rebar workers on large slab construction sites, collected with surveys internal to the manufacturer. The pilot study field conditions were fundamentally different compared to large-slab construction sites, as the conditions involved a relatively small concrete slab with no need to carry rebars. Moreover, although the participants were experienced construction workers, they did not specialize in the rebar trade. As noted in [17], difference in experience in the rebar trade is correlated with significant differences in posture while tying.

The above observations highlight the challenge of designing OEs that accommodate a wide range of workers and use cases, and the challenge of ensuring field conditions that provide generalizable results. Identifying useful minimum design requirements, as described in Section 2.3, can help strike a balance between OEs that are versatile enough to support diverse trades while remaining cost-effective for widespread adoption.

4 Summary and Conclusions

Improving the design and metrology of construction OEs can accelerate the successful adoption of OEs in the construction industry. Informed by previous research, this paper provides a framework by which the OE research community may generate the holistic body of scientific evidence required for OE adoption while addressing the unique challenges faced by the construction industry. By aligning OE design and evaluation processes with trade-specific requirements, incorporating comprehensive stakeholder input, and by leveraging standard tests and minimum criteria, the roadmap offers a specific pathway for accelerating the adoption of safe, accepted and effective OEs.

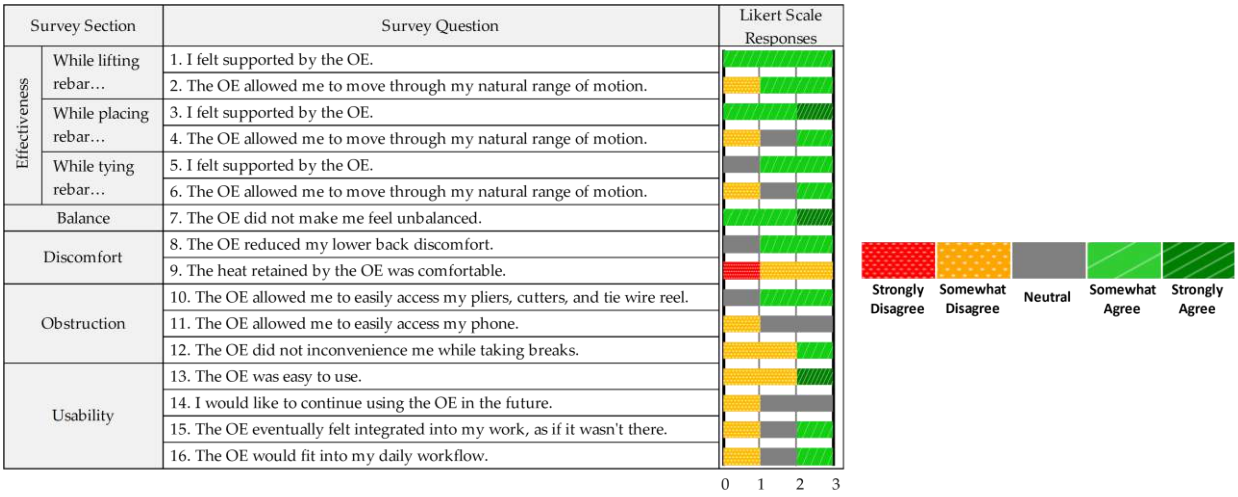


Figure 5. Subjective data from the pilot field study.

The framework outlines significant research initiatives, and this paper argues that investment in these initiatives is justified by the benefits they may bring to the construction industry: mitigating the unique challenges of OE design and evaluation which prevent large-scale OE adoption. To demonstrate the value of the framework with real-world examples of these design and evaluation challenges, the paper presented a pilot field study on a back-support OE for construction workers. The challenges identified in the study included choosing suitable measures for field evaluation, choosing representative field conditions in which to evaluate an OE, the difficulty of accounting for minimum design requirements in unpredictable environments, and other specific design challenges like heat retention. All these challenges can be mitigated by insights gained from a trade-specific industry stakeholder survey, and by this survey informing the other novel aspects of the proposed framework.

As described in the roadmap in Section 2.4, future work can focus on conducting trade-specific industry surveys to establish a consensus on the criteria for each decision point in the framework, eventually leading to benchmarks for minimum design requirements and useful measures of OE effects.

Acknowledgments

This research represents a joint effort with Hydro-Québec and Biolift Technologies, and the authors sincerely thank them for this collaboration. The authors would also like to thank the study participants for their enthusiasm and good humor throughout the research.

References

[1] U.S. Bureau of Labor Statistics, "Nonfatal cases involving days away from work: selected characteristics (2011 forward)." https://www.bls.gov/help/one_screen/cs.htm (accessed Aug. 24, 2022).

[2] "CPWR Chart Book (6th edition): Fatal and Nonfatal Injuries - Musculoskeletal Disorders in Construction and Other Industries," *CPWR*. https://www.cpwr.com/research/data-center/the-construction-chart-book/chart-book-6th-edition-fatal-and-nonfatal-injuries-musculoskeletal-disorders-in-construction-and-other-industries/ (accessed Mar. 24, 2024).

[3] J. T. Albers and C. F. Estill, *Simple Solutions: Ergonomics for Construction Workers*. U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, 2007.

[4] G. Gagnon, F. Sekkay, D. Imbeau, and M. Bourgault, "Analyzing Occupational Accidents and Exoskeleton Potential in the Construction Industry in Québec, Canada," *IISE Trans. Occup. Ergon. Hum. Factors*, pp. 1–19, Dec. 2024, doi: 10.1080/24725838.2024.2432450.

[5] A. J. van der Beek, J. T. Dennerlein, M. A. Huysmans, S. E. Mathiassen, A. Burdorf, W. van Mechelen, *et al.*, "A research framework for the development and implementation of interventions preventing work-related musculoskeletal disorders," *Scand. J. Work. Environ. Health*, vol. 43, no. 6, pp. 526–539, 2017, doi: 10.5271/sjweh.3671.

[6] M. Pesenti, A. Antonietti, M. Gandolla, and A. Pedrocchi, "Towards a Functional Performance Validation Standard for Industrial Low-Back

Exoskeletons: State of the Art Review," *Sensors*, vol. 21, no. 3, Art. no. 3, Jan. 2021, doi: 10.3390/s21030808.

[7] S. De Bock, J. Ghillebert, R. Govaerts, B. Tassignon, C. Rodriguez-Guerrero, S. Crea, *et al.*, "Benchmarking occupational exoskeletons: An evidence mapping systematic review," *Appl. Ergon.*, vol. 98, Jan. 2022, doi: 10.1016/j.apergo.2021.103582.

[8] A. Golabchi, N. Riahi, M. Fix, L. Miller, H. Rouhani, and M. Tavakoli, "A framework for evaluation and adoption of industrial exoskeletons," *Appl. Ergon.*, vol. 113, p. 104103, Nov. 2023, doi: 10.1016/j.apergo.2023.104103.

[9] P. M. Kuber, M. Abdollahi, M. M. Alemi, and E. Rashedi, "A Systematic Review on Evaluation Strategies for Field Assessment of Upper-Body Industrial Exoskeletons: Current Practices and Future Trends," *Ann. Biomed. Eng.*, vol. 50, no. 10, pp. 1203–1231, Oct. 2022, doi: 10.1007/s10439-022-03003-1.

[10] S. A. Elprama, B. Vanderborght, and A. Jacobs, "An industrial exoskeleton user acceptance framework based on a literature review of empirical studies," *Appl. Ergon.*, vol. 100, Apr. 2022, doi: 10.1016/j.apergo.2021.103615.

[11] N. Hoffmann, G. Prokop, and R. Weidner, "Methodologies for evaluating exoskeletons with industrial applications," *Ergonomics*, vol. 65, no. 2, pp. 276–295, Feb. 2022, doi: 10.1080/00140139.2021.1970823.

[12] M. de Looze, A. de Vries, F. Krause, and S. Baltrusch, "Three-Stage Evaluation for Defining the Potential of an Industrial Exoskeleton in a Specific Job," In *Proceedings of the 21st Congress of the International Ergonomics Association (IEA 2021)*, Cham, 2022, pp. 235–241, doi: 10.1007/978-3-030-74614-8_28.

[13] S. Crea, P. Beckerle, M. D. Looze, K. D. Pauw, L. Grazi, T. Kermavnar, *et al.*, "Occupational exoskeletons: A roadmap toward large-scale adoption. Methodology and challenges of bringing exoskeletons to workplaces," *Wearable Technol.*, vol. 2, 2021, doi: 10.1017/wtc.2021.11.

[14] Z. Zhu, A. Dutta, and F. Dai, "Exoskeletons for manual material handling – A review and implication for construction applications," *Autom. Constr.*, vol. 122, Feb. 2021, doi: 10.1016/j.autcon.2020.103493.

[15] C. Nnaji, I. Okpala, J. Gambatese, and Z. Jin, "Controlling safety and health challenges intrinsic in exoskeleton use in construction," *Saf. Sci.*, vol. 157, Jan. 2023, doi: 10.1016/j.ssci.2022.105943.

[16] V. Kopp, M. Holl, M. Schalk, U. Daub, E. Bances, B. García, *et al.*, "Exoworkathlon: A prospective study approach for the evaluation of industrial exoskeletons," *Wearable Technol.*, vol. 3, 2022, doi: 10.1017/wtc.2022.17.

[17] M. Dunson-Todd, M. Nik-Bakht, and A. Hammad, Forthcoming, "Towards a standard test to evaluate back-support exoskeleton efficacy for rebar workers: Test design and initial implementation," *J. Comp. Civ. Eng.*

[18] S. E. Kranenborg, C. Greve, M. F. Reneman, and C. C. Roossien, "Side-effects and adverse events of a shoulder- and back-support exoskeleton in workers: A systematic review," *Appl. Ergon.*, vol. 111, p. 104042, Sep. 2023, doi: 10.1016/j.apergo.2023.104042.

[19] S. Madinei and M. A. Nussbaum, "Estimating lumbar spine loading when using back-support exoskeletons in lifting tasks," *J. Biomech.*, vol. 147, Jan. 2023, doi: 10.1016/j.jbiomech.2023.111439.

[20] N. Gutierrez, A. Ojelade, S. Kim, A. Barr, A. Akanmu, M. A. Nussbaum, *et al.*, "Perceived benefits, barriers, perceptions, and readiness to use exoskeletons in the construction industry: Differences by demographic characteristics," *Appl. Ergon.*, vol. 116, 2024, doi: 10.1016/j.apergo.2023.104199.

[21] D. Mahmud, S. T. Bennett, Z. Zhu, P. G. Adamczyk, M. Wehner, D. Veeramani, *et al.*, "Identifying Facilitators, Barriers, and Potential Solutions of Adopting Exoskeletons and Exosuits in Construction Workplaces," *Sensors*, vol. 22, no. 24, Art. no. 24, Jan. 2022, doi: 10.3390/s22249987.

[22] ASTM Committee F48, "F3523-21 Standard Test Method for Exoskeleton Use: Confined Space: Horizontal Movement." ASTM International, 2021. doi: 10.1520/F3523-21.

[23] A. W. de Vries, F. Krause, and M. P. de Looze, "The effectivity of a passive arm support exoskeleton in reducing muscle activation and perceived exertion during plastering activities," *Ergonomics*, vol. 64, no. 6, pp. 712–721, Jun. 2021, doi: 10.1080/00140139.2020.1868581.

[24] N. Gonsalves, A. Akanmu, A. Shojaei, and P. Agee, "Factors influencing the adoption of passive exoskeletons in the construction industry: Industry perspectives," *Int. J. Ind. Ergon.*, vol. 100, 2024, doi: 10.1016/j.ergon.2024.103549.

[25] S. Kim, A. Ojelade, A. Moore, N. Gutierrez, C. Harris-Adamson, A. Barr, *et al.*, "Understanding contributing factors to exoskeleton use-intention in construction: A decision tree approach using results from an online survey," *Ergonomics*, vol. 0, no. ja, pp. 1–22, 2023, doi: 10.1080/00140139.2023.2289859.

[26] Biolift Technologies, "Biolift," 2024. https://biolift.co/en/home/

Feasibility of an EEG-based dynamic suboptimal cognitive monitoring for field neuroergonomics

Eren (M.) Shahrokhi[1], S. Nizam Ahmed[2], Gaang Lee[1] and SangHyun Lee[3]

[1]Hole School of construction Engineering and Management, Dept. of Civil and Environmental Engineering, University of Alberta, Edmonton, Canada
[2] Clinical Neurophysiology Laboratory, Division of Neurology, Department of Medicine, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, AB, Canada
[3] Tishman Construction Management Program, Dept. of Civil and Environmental Engineering, University of Michigan, 2350 Hayward St., Ann Arbor, MI 48109, USA
shahrok1@ualberta.ca, snahmed@ualberta.ca, gaang@ualberta.ca, shdpm@umich.edu

Abstract

Suboptimal cognitive states among construction workers significantly impact safety and productivity, with mental workload playing a key role in triggering these states. Determining if the mental workload fluctuation is leading to an error is challenging as the relationship between mental workload and suboptimal cognitive states is complex and non-linear, with traditional theories failing to map their fluctuations effectively. Recently, a two-dimensional space has been introduced to theoretically map mental workload fluctuations and suboptimal cognitive states using task engagement and arousal. However, there is currently no framework in place to continuously apply this theoretical knowledge in practical settings. To address this gap, this study investigates the feasibility of EEG-based frameworks for classifying four different cognitive states, namely comfort zone, mind wandering, effort withdrawal, and inattentional blindness, based on mental workload fluctuations. EEG signals were collected from 10 participants using a headset with dry electrodes, processed to extract relevant features, and classified using Support Vector Machine (SVM) and Artificial Neural Network (ANN) models. The ANN achieved superior performance in k-fold and leave one period out validation methods, though accuracy declined in leave one subject out validation. These findings underscore the potential of EEG-based differentiation of cognitive suboptimalities to enhance safety and productivity in construction by providing crucial information about when construction workers are most likely to make cognitive errors, which is essential for timely and appropriate interventions. Also, the low subject independent accuracy emphasizes the need to address individual differences in EEG signals for broader applicability.

Keywords

EEG, Wearables, Biosensors, Neuroergonomics, Construction safety

1 Introduction

Managing workers' mental workload is a critical aspect of construction management [1]. The dynamic nature of construction sites often subjects workers to varying levels of mental workload, ranging from excessively low to overly high. These suboptimal cognitive states increase the likelihood of unsafe or erroneous work behavior, making it essential to monitor and manage mental workload for both safety and productivity.

Traditionally, mental workload monitoring has been conducted through self-assessment methods, such as questionnaires and interviews [2,3]. However, these methods are highly subject to individuals' recall bias and sporadic due to their invasive nature, making them unsuitable for continuous and dynamic monitoring, which is critical for understanding the dynamics of mental workload oscillation during the ongoing task. To address these limitations, neuroergonomics, which is a field of research to study human brain function in the workplace [4], has been integrated with various wearable-based techniques capable of continuous monitoring and assessing the mental workload. These studies have applied different wearable biosensors, such as electrocardiogram (ECG) [5], electroencephalogram (EEG) [6], photoplethysmography (PPG) [7], and Electrodermal activity (EDA) [8], to measure individuals' mental workload levels during their ongoing tasks.

The previous efforts have enabled the monitoring of changing mental workload levels. For example, [9] applied EEG to quantify the construction workers mental workload and the findings suggest that Gamma band activity of specific EEG channels strongly correlates with tasks with high mental demands. However, challenges

remain in distinguishing problematically low or high levels—those that involve suboptimal cognitive states, such as mind wandering and blindness, and increased susceptibility to cognitive errors—from acceptable/appropriate levels during ongoing field tasks. The current literature does not establish thresholds for problematic mental workload, leaving no clear reference for identifying activities that may induce suboptimal cognitive states in workers. This limitation hinders the operationalization of biosensor-based mental workload monitoring to improve workers' cognitive states by identifying and addressing problematic moments and environmental factors likely to cause suboptimal cognitive states.

A recently introduced framework [10] has shown potential not only in addressing this limitation, but also in enabling the differentiation of different suboptimal states for interventions, by characterizing problematically low and high mental workload in relation to neurophysiological states. The framework, proposed by Dehais [10] and his team, reconceptualizes mental workload as a transactional interaction between an individual and a task's cognitive demands, influenced by intra-individual factors (e.g., emotional states, fatigue) and inter-individual variability (e.g., skill levels, personality). It categorizes neurophysiological states along with their associated suboptimal cognitive states and their symptoms that emerge when mental workload becomes problematically low or high. Specifically, an individual's neurophysiological state, depending on their mental workload, can be mapped onto a two-dimensional space defined by task engagement (the cognitive and emotional effort invested in achieving task goals) and arousal (the physiological readiness to respond to goals and external task demands, influenced by inherent traits and situational factors). This two-dimensional space is divided into four regions: three suboptimal cognitive states—Mind Wandering, Effort Withdrawal, and Mind Blindness—and one optimal cognitive state (Figure 1).

Mind Wandering occurs when mental workload is excessively low, leading to boredom and a lack of focus on the task. Effort Withdrawal and Inattentional Blindness, on the other hand, arise from excessively high mental workload levels, representing two distinct suboptimal states characterized by reduced task engagement and diminished awareness of unexpected stimuli, respectively. Differentiating these three mental workload-related suboptimal cognitive states not only enables the detection of problematic mental workload, but also provides insights for "state-specific" interventions, thereby significantly contributing to the field application of the wearable biosensor-based neuroergonomic approach. This framework takes a more detailed look at mental workload by considering both the intensity and quality of cognitive engagement. It offers a clearer way to understand how individual cognitive states influence performance on complex tasks.

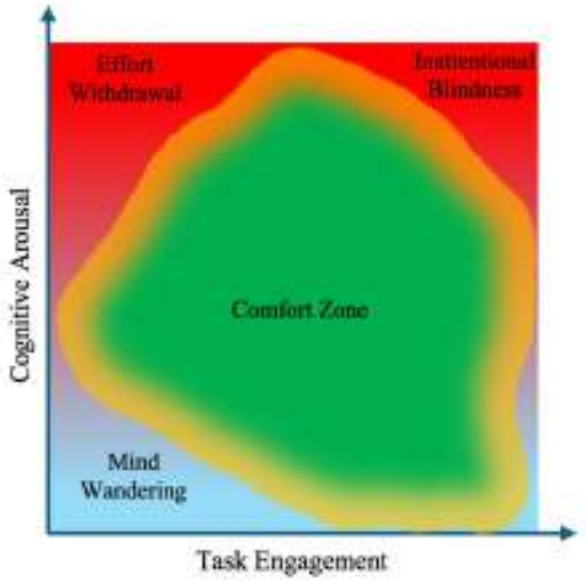


Figure 1. Dehais's framework characterizing mental workload-related suboptimal states [10]

To effectively implement this sophisticated framework in practical settings, an equally advanced method of analyzing the collected neurophysiological data is essential. In the field of neuroergonomics, traditional approaches to analyzing the signals collected during cognitive monitoring typically employ either post-hoc analysis or machine learning methods. While post-hoc analysis has been valuable for in-depth, after-the-fact evaluations, it does not facilitate the continuous and dynamic adjustments needed for ongoing cognitive state monitoring. Therefore, the integration of machine learning analysis and the new perspective on mental workload shows significant potential for effective and dynamic monitoring of field works. Despite this potential, the feasibility of this framework in accurately identifying distinct cognitive states remains untested. This gap signifies a crucial missed opportunity for enhancing operational efficiency, safety, and worker well-being through state-specific cognitive interventions. To address this knowledge gap, this study aims to investigate the feasibility of developing an EEG-based machine learning model that can differentiate between four cognitive states derived from the novel perspective on operationalized mental workload.

2 Methods

To achieve the research objective, a machine learning model was developed for classifying different suboptimal cognitive states (i.e., Mind Wandering, Effort

Withdrawal, and Mind Blindness), along with the optimal. To this end, an in-lab data collection was structured incorporating the characteristics of the suboptimal cognitive states stated in Dehais's framework and conducted. Following this, the data preprocessing techniques were employed to ensure the EEG signals were clean and reliable. The focus then shifted to the feature extraction process, which is crucial for preparing the data for effective machine learning analysis and classification model development.

2.1 Data collection

2.1.1 Participants

Data were collected from 10 participants aged 20 to 39 years, comprising 5 females and 5 males. All participants reported being in good physical and mental health, with no history of neurological disorders. To minimize confounding factors, participants were instructed to abstain from alcohol for 24 hours before the session. Each participant completed only the data collection session, with no additional activities scheduled before or after. The study protocols were approved by the University of Alberta's Research Ethics Board (Reference: Pro00135533), and participants provided informed consent via an electronic form before starting.

2.1.2 Tools

The study employed the DSI-24 dry-electrode EEG headset from Wearable Sensing for data acquisition. This device was chosen for its practical advantages, including reduced setup time, stable signal quality without the need for conductive gel, and comfort during extended use. In addition, the patented common-mode follower technology in this headset offers near immunity to electrical and motion artifact signals during data collection [11]. The headset features 19 channels based on the 10-20 system, a sampling rate of 300 Hz, and an active shielding system to filter environmental noise. Wireless data collection ensured flexibility and minimized interference during the session.

The data collection session was designed using the PsyToolkit platform [12,13], while the iMotion platform was utilized to integrate biosensors and stimuli for improved synchronization. This setup enabled the efficient execution of complex data collection designs.

2.1.3 Procedures

The data collection session involved three cognitive tests, each designed to target specific suboptimal cognitive states characterized by Dehais's framework [10]. First, an adapted Task-Switching test was used to induce inattentional blindness [14]. Participants alternated between rules for responding to stimuli, such as letters and numbers in a grid, under time constraints. The frequent rule changes introduced cognitive demands, simulating real-world multitasking challenges. Participants had only 4 seconds to respond to each stimulus; failing to do so triggered an error message. This setup required participants to concentrate intensely at the beginning of each block to apply the newly presented rule effectively and perform well. The intense focus and task engagement accompanied with high cognitive arousal due to time constraints, can lead to Inattentional blindness and deafness as shown in Figure 1. Each participant completed six Task-Switching sessions, each lasting 2.5 minutes, with randomized rules for each session. Breaks were allowed between sessions if needed.

The second test was the n-back task [15], designed to induce effort withdrawal. Participants were required to identify matches in a sequence of stimuli presented a specific number of steps (n) earlier, with task difficulty increasing incrementally from 1-back to 7-back. When the number of steps is higher than 5 the task looks impossible to individuals [16] and discourages them from fully engaging in the task since achieving the goal seems very unlikely. This procedure can induce Effort Withdrawal behavior among participants. Four n-back tasks were administered, each lasting 2.5 minutes, with no breaks between levels to maintain cognitive load and push participants toward a mental breaking point.

The third test was the Sustained Attention to Response Task (SART) [17], conducted to induce mind-wandering. Participants responded to any digit from 1 to 9 while withholding responses to an exception, the digit "3". Participants had 0.9 seconds to respond to the digits, and since the target digit was infrequent, they frequently had to respond to other stimuli. Over time, this repetitive response to non-target digits could become an automatic process requiring minimal attention, potentially allowing their freed-up focus to drift to other tasks. This cognitive shift increases the likelihood of Mind Wandering. Five SART sessions were presented toward the end of the session, leveraging reduced cognitive arousal after the more demanding preceding tasks. Each session lasted 4 minutes.

Upon arrival, participants were introduced to the biosensors and provided informed consent before setup. The tests were conducted in a quiet, isolated room to minimize distractions. The entire data collection session, including setup and breaks, lasted between 1.5 and 2 hours. This duration was carefully optimized to minimize fatigue while ensuring sufficient data quality and variability for the machine learning model.

2.2 Data preprocessing

EEG data were continuously recorded throughout the session without interruptions. To prepare the data for analysis, a 60 Hz notch filter was first applied (to account for the powerline frequency in Canada), followed by

bandpass filtering between 0.5 Hz and 50 Hz to reduce extrinsic artifacts. Despite these preprocessing steps, the data remained prone to noise from intrinsic artifacts such as eye blinks and muscle movements. To mitigate these, Independent Component Analysis (ICA) was employed to isolate and remove artifact-related components, enhancing the overall signal-to-noise ratio.

After preprocessing, the data were prepared for labelling. Since no self-reported measurements were used, the labelling strategy was based on the design and cognitive mechanisms of the tasks, supported by previous research. Each section's labelling was validated separately to ensure the tasks effectively induced the intended cognitive suboptimalities.

For Task Switching, it was hypothesized that participants initially struggle with new rules under time constraints but adapt as the session progresses, feeling more at ease. Therefore, first 30 seconds of each new trial was labelled as inattentional blindness. For effort withdrawal, the Task Engagement Index (TEI) was calculated during the 1-back, 3-back, 5-back, and 7-back tasks. The TEI increased from 1-back to 5-back but significantly dropped for the 7-back task, reflecting a state of effort withdrawal as participants found the task too demanding to complete effectively. Hence, the 7-back task data were used as effort withdrawal data.

For SART, errors were captured using a setup where an error message was displayed alongside a white box with a photodetector attached to the screen. This system pinpointed the exact moments when participants made errors. Following the methodology of previous studies [18], the 5 seconds preceding each error were labeled as mind-wandering states. After the error message appeared, participants were prompted to refocus and attempt better performance, and these moments were labeled as focus states.

2.3 Feature Extraction

After labeling the data, features were extracted from the EEG signals to prepare inputs for the machine learning model and analyze signal patterns linked to various suboptimal cognitive states. An effective window size for EEG data segmentation in cognitive state classification is typically determined through trial and error and generally ranges from 1 to 12 seconds [19]. The optimal window size in this study was found at using a 2-second window with a 25% overlap to ensure sufficient temporal resolution.

Feature extraction is a critical step in developing robust EEG-based machine learning models for classifying cognitive states. Selecting features that effectively capture the underlying neural dynamics is essential for accurate and meaningful classification. From the literature, features derived from both time and frequency domains, as well as non-linear analyses, have demonstrated their effectiveness in distinguishing between mental states and cognitive processes. Specifically, power spectral features, such as Alpha, Beta, Theta, and Delta Power, offer insights into the brain's frequency-specific activities related to relaxation, focus, drowsiness, and disengagement, respectively. Complementing these, non-linear and time-domain measures, such as Higuchi Fractal Dimension and Hjorth parameters (Activity, Mobility, and Complexity), provide valuable information on neural complexity, brain activity levels, cognitive flexibility, and adaptability. Together, these features encapsulate the multifaceted nature of brain activity, making them a suitable choice for the classification of various cognitive states [20].

A total of eight features were computed for each EEG channel: alpha, beta, theta, and delta power bands; Higuchi fractal dimension; and Hjorth parameters (activity, mobility, and complexity). With 19 channels in the EEG system, this resulted in 152 features (8 features × 19 channels) being extracted for each window. A summary of the selected features, along with the rationale for their selection and their significance in EEG signals, is presented in Table 1.

Table 1. Summary of feature selection

Feature	Description
Alpha Power	Represents relaxed but wakeful states and reduced sensory input, typically decreasing with increased cognitive load
Beta Power	Linked to focus, problem-solving, and sustained attention, often increasing with cognitive effort
Theta Power	Associated with drowsiness and working memory, helpful in identifying low vigilance or mental fatigue
Delta Power	Indicates low-frequency activity often tied to relaxation or cognitive disengagement in awake individuals
Higuchi Fractal Dimension (HFD)	Captures neural complexity, with higher values reflecting increased cognitive processing
Hjorth Activity	Reflects overall brain activity levels, useful for assessing mental workload or stress
Hjorth Mobility	Represents frequency variability, linked to cognitive flexibility
Hjorth	Indicates adaptability and mental engagement, highlighting the temporal

Complexity	intricacy of brain activity

These features provide a comprehensive representation of brain dynamics, enabling the model to differentiate between distinct cognitive suboptimalities effectively.

2.4 Multiclass classification of suboptimal cognitive states

After completing the feature extraction and labelling stages, a feature matrix and a target matrix were constructed to train machine learning models for classifying various suboptimal cognitive states: Mind Wandering, Effort Withdrawal, Inattentional Blindness, and the Comfort Zone.

In this study, Support Vector Machines (SVM) and Artificial Neural Networks (ANN) were utilized to classify EEG-based cognitive states due to their complementary capabilities in processing complex, high-dimensional data. SVM was employed for its effectiveness in identifying nonlinear decision boundaries and its suitability for smaller datasets, leveraging kernel functions to map EEG data into higher-dimensional spaces for enhanced separability. Fully connected ANN was selected for its capacity to model intricate patterns and automatically learn features from EEG signals, accommodating the data's non-stationary and multivariate characteristics. Both models were trained and evaluated on preprocessed EEG data to compare their classification performance across different cognitive states.

In machine learning research, validation methods play a crucial role in assessing the generalizability of models to unseen data [21,22]. Among the commonly used approaches is K-Fold Cross-Validation (KFCV), which splits the dataset into k equal-sized subsets or "folds." The model is trained on k-1 folds and tested on the remaining fold, cycling through until each subset has been used for testing. While this method is efficient and widely applicable, it assumes independence between data points. This assumption can lead to overestimation of model performance, particularly when applied to time-series or biosignal data, where neighboring data points are inherently correlated. The result of this validation method is used as a benchmark to compare the developed model with existing studies since this validation method is frequently used in the current literature [19,23].

The second method, Leave-One-Period-Out Cross Validation (LOPOCV), is designed to evaluate the generalizability of machine learning models across different environmental or contextual factors. In this approach, the dataset is divided so that data from one specific context—such as a particular environmental condition, task type, or experimental setup—is completely excluded during training and reserved for testing. This process is repeated iteratively, with each context taking a turn as the test set, ensuring that every context is evaluated independently.

Finally, Leave-One-Subject-and-Context-Out Cross-Validation (LOSCOCV) [24] was used to ensures both subject and context independence. By excluding data from one subject and one context during training and using them for testing, LOSCOCV provides a more stringent evaluation of generalizability. This method is particularly suited for biosignal research, where both individual and contextual variability significantly influence the data. Incorporating these three validation methods into the methodology ensures a comprehensive evaluation of model performance and generalizability under varying conditions.

To optimize performance, hyperparameter tuning was conducted for both the SVM and ANN models. The holdout method was employed for this process, with 8 subjects used for training and 2 set aside as the validation dataset. Models were trained and tested on this subset to identify the best hyperparameters, which were subsequently applied to the full dataset of 10 subjects. The selected hyperparameters ensured the models achieved robust performance across all validation methods.

3 Results

The classification models were trained using the SVM and ANN algorithms, with hyperparameters optimized to achieve the best performance. The implemented ANN has a fully connected architecture with 3 hidden layers and 2 drop out layers. The hyperparameters were determined using a grid search method. Final architecture and hyperparameter setting of SVM and ANN are presented presented in Table 2.

Table 2. Hyper parameter tunning results

Model	Hyperparameter and Architecture
SVM	Kernel: RBF
	C: 10 / Gamma: 0.1
ANN	Input layer: 152 neurons
	1st Hidden layer: 100 neurons
	1st Dropout layer: 15 percent
	2nd Hidden layer: 100 neurons
	2nd Dropout layer: 15 percent
	3rd Hidden layer: 75 neurons
	Output layer: 4 neurons

Model performance was evaluated using three validation methods—KFCV, LOPOCV, and LOSCOCV—and the results are presented in Table 3. The ANN outperformed the SVM across all validation methods, achieving the highest accuracy in both KFCV and LOPOCV validations. These results align with

previous studies that differentiate between levels of specific cognitive suboptimalities, providing benchmarks for comparison. Importantly, this study highlights the challenge of achieving high accuracy in LOSCOCV models, a critical area for future research and practical applications.

Table 3. Accuracy of different models for each validation method

Model	Validation method					
	KFCV		LOPOCV		LOSCOCV	
	Acc	F1	Acc	F1	Acc	F1
SVM	0.62	0.62	0.58	0.59	0.24	0.10
ANN	0.72	0.71	0.68	0.65	0.27	0.24
* ACC: classification accuracy						

4 Discussion

Traditional approaches to mental workload have faced challenges in distinguishing various suboptimal cognitive states, as the relationship between mental workload and cognitive errors induced by these states remained unclear. Recently, new perspectives have emerged, focusing on measuring task engagement and arousal to map suboptimal cognitive states related to mental workload, opening new opportunities to enhance field applicability. This study explores the feasibility of this approach by employing a neuroergonomic framework to map suboptimal cognitive states resulting from excessive mental workload. A machine learning model was developed to classify four distinct cognitive states using features extracted from EEG signals.

The study classifies cognitive comfort zone, mind wandering, inattentional blindness, and effort withdrawal with acceptable accuracy. Given that the task involves a 4-class classification, accuracies above 0.7 can be considered promising. Both classification accuracy and F1-scores were reported to mitigate the effects of data imbalance and validate the performance of the models. Among the models tested, the ANN trained with extracted features outperformed the SVM, achieving the highest accuracy. This result aligns with expectations, given the ANN's ability to handle the high number of features and capture complex relationships between input data and cognitive state labels. While the SVM demonstrated reasonable performance, its simpler architecture limits its ability to manage the intricacies of the dataset compared to the ANN.

Accuracy decreased as validation methods introduced greater independence across time periods or subjects, highlighting the challenges of generalization. However, the ANN demonstrated robust performance under LOPOCV, showcasing its potential for generalization within the same participant group. This validation method ensures that testing occurs on unseen time periods and data collection sessions, meaning the model can differentiate classes with acceptable precision even when the participant remains the same but the data collection setup changes. However, the accuracy for LOSCOCV was notably low across all models, underscoring the difficulty of achieving generalizability across different individuals. This limitation highlights the need for additional labeled data from new subjects to train models effectively, a process that is both time-consuming and inefficient.

This study demonstrates the feasibility of using machine learning combined with EEG-extracted features to classify cognitive states, including both the comfort zone and cognitive suboptimalities. This framework holds significant potential for continuous, dynamic monitoring of construction field workers with minimal invasiveness and without disrupting their tasks. Its neuroergonomic foundation enhances reliability by leveraging data directly from the central nervous system.

The benefits of such a framework are multifaceted. It enables timely interventions to mitigate risks associated with cognitive suboptimalities, thereby reducing cognitive errors and enhancing worker safety and performance. Furthermore, it offers indirect applications, such as integration with geographic information systems (GIS) or eye-tracking data for hotspot analyses. This integration can help identify areas or environmental factors where cognitive suboptimalities frequently occur, facilitating targeted interventions to improve workplace conditions.

This study sets a benchmark for multiclass classification of cognitive suboptimalities but has some limitations that need further exploration. First, the tested framework shows low generalizability across different subjects, which means there is a need to individualize the model for each user or increase the dataset diversity by data augmentation. Second, the proposed technique requires extensive labelled data collection eliciting cognitive suboptimalities, highlighting the need for a combination of improved protocols and techniques that reduce the reliance on learning from scratch to quickly build more reliable datasets. This prolonged data collection process means the developed model may require many hours of data acquisition and hyperparameter tuning whenever a new user is introduced, significantly increasing implementation time and effort. Third, as the models were developed and tested with data collected in a controlled environment, the additional finetuning and/or validation might be required in more realistic work scenarios to examine the field applicability.

Given the lack of a clear solution in the current literature for developing a generalizable model across individuals and the absence of a comprehensive dataset

with sufficient representation from diverse demographic groups, future research should explore strategies to simplifying the data collection and hyperparameter tuning process for new users. Reducing the time required for these tasks would enhance the framework's cost-effectiveness and scalability, opening new possibilities for industrial applications.

Addressing this technical challenge is crucial because the practical use of current EEG-based models is hindered by their limited subject independence. Future studies need to focus on enhancing subject independence to facilitate wider use and real-world implementation of these models.

5 Conclusion

This study demonstrates the feasibility of using mobile EEG-based machine learning models to detect and classify suboptimal cognitive states, highly relevant to construction workers' cognitive errors in the field. By leveraging neuroergonomic principles and advanced validation methods, the proposed framework successfully differentiates between four distinct cognitive states: mind wandering, inattentional blindness, effort withdrawal, and the comfort zone. Among the models tested, the ANN consistently outperformed the SVM in classification tasks, demonstrating superior capability in handling high-dimensional EEG data and capturing complex relationships. The findings underscore the potential of EEG-based frameworks for real-time and continuous monitoring of cognitive states in demanding environments like construction, with minimal invasiveness and without disrupting workers' tasks.

Despite its success, the study also identifies significant challenges, particularly in achieving generalizability across individuals. Accuracy decreased when validation methods introduced greater independence across subjects or contexts, highlighting the need for extensive retraining with new user-specific data. This limitation points to a critical area for future research: enhancing model generalizability to enable broader applicability without compromising accuracy. Additionally, the time-intensive process of collecting labeled data and tuning model hyperparameters further underscores the necessity of more efficient protocols and comprehensive datasets representing diverse demographics.

The implications of this work are both immediate and far-reaching. In the short term, the framework offers a robust tool for identifying cognitive suboptimalities in real-time, enabling timely interventions to mitigate risks, reduce cognitive errors, and improve worker safety and performance. In the long term, its integration with other technologies, such as GIS or eye-tracking systems, could provide actionable insights into environmental factors that contribute to cognitive errors. This integration could facilitate targeted interventions to optimize workplace conditions and enhance overall productivity.

Future research should focus on addressing these limitations by improving data collection, developing generalizable models that work across individuals and contexts, and validating the framework in larger and more diverse populations. Expanding its practical applications to other industries and refining its usability for real-world conditions could further enhance its impact. This study establishes a critical foundation for EEG-based neuroergonomic frameworks, advancing our ability to monitor and manage mental workload dynamically and effectively in the field settings.

Acknowledgement

This work was supported by the University of Alberta Research Exploration Funds, the Construction Innovation Centre, and the New Frontiers in Research Fund (NFRF), grant number NFRFE-2023-00174. The authors also wish to thank their industry partners for their support in data collection, as well as the anonymous participants who contributed their time and insights.

References

[1] J. Chen, X. Song, Z. Lin, Revealing the "Invisible Gorilla" in construction: Estimating construction safety through mental workload assessment, Autom Constr 63 (2016) 173–183. https://doi.org/10.1016/J.AUTCON.2015.12.018.

[2] W. Umer, H. Li, Y. Yantao, M.F. Antwi-Afari, S. Anwer, X. Luo, Physical exertion modeling for construction tasks using combined cardiorespiratory and thermoregulatory measures, Autom Constr 112 (2020) 103079. https://doi.org/https://doi.org/10.1016/j.autcon.2020.103079.

[3] H.J. Michielsen, J. De Vries, G.L. Van Heck, Psychometric qualities of a brief self-rated fatigue measure: The Fatigue Assessment Scale, J Psychosom Res 54 (2003) 345–352. https://doi.org/10.1016/S0022-3999(02)00392-6.

[4] R. Parasuraman, M. Rizzo, Neuroergonomics: The brain at work, Oxford University Press, 2006.

[5] H. Qu, X. Gao, L. Pang, Classification of mental workload based on multiple features of ECG signals, Inform Med Unlocked 24 (2021) 100575. https://doi.org/https://doi.org/10.1016/j.imu.2021.100575.

[6] W.K.Y. So, S.W.H. Wong, J.N. Mak, R.H.M. Chan, An evaluation of mental workload with

frontal EEG, PLoS One 12 (2017) e0174949. https://doi.org/https://doi.org/10.1371/journal.pone.0174949.

[7] W.-K. Beh, Y.-H. Wu, A.-Y. Wu, Robust PPG-based mental workload assessment system using wearable devices, IEEE J Biomed Health Inform 27 (2021) 2323–2333. https://doi.org/10.1109/JBHI.2021.3138639.

[8] C. Lee, M.J. Shin, D. Eniyandunmo, A. Anwar, E. Kim, K. Kim, J.K. Yoo, C. Lee, Predicting Driver's mental workload using physiological signals: A functional data analysis approach, Appl Ergon 118 (2024). https://doi.org/10.1016/j.apergo.2024.104274.

[9] J. Chen, J.E. Taylor, S. Comu, Assessing Task Mental Workload in Construction Projects: A Novel Electroencephalography Approach, J Constr Eng Manag 143 (2017). https://doi.org/10.1061/(asce)co.1943-7862.0001345.

[10] F. Dehais, A. Lafont, R. Roy, S. Fairclough, A Neuroergonomics Approach to Mental Workload, Engagement and Human Performance, Front Neurosci 14 (2020). https://doi.org/10.3389/fnins.2020.00268.

[11] Y. Mahdid, U. Lee, S. Blain-Moraes, Assessing the quality of wearable EEG systems using functional connectivity, IEEE Access 8 (2020) 193214–193225. https://doi.org/10.1109/ACCESS.2020.3033472.

[12] G. Stoet, PsyToolkit: A software package for programming psychological experiments using Linux, Behav Res Methods 42 (2010) 1096–1104. https://doi.org/10.3758/BRM.42.4.1096.

[13] G. Stoet, PsyToolkit: A Novel Web-Based Method for Running Online Questionnaires and Reaction-Time Experiments, Teaching of Psychology 44 (2017) 24–31. https://doi.org/10.1177/0098628316677643.

[14] S. Monsell, Task switching, Trends Cogn Sci 7 (2003) 134–140. https://doi.org/10.1016/S1364-6613(03)00028-7.

[15] M.J. Kane, A.R.A. Conway, T.K. Miura, G.J.H. Colflesh, Working memory, attention control, and the N-back task: a question of construct validity., J Exp Psychol Learn Mem Cogn 33 (2007) 615. https://doi.org/https://psycnet.apa.org/doi/10.1037/0278-7393.33.3.615.

[16] S.H. Fairclough, L.J. Moores, K.C. Ewing, J. Roberts, Measuring task engagement as an input to physiological computing, in: 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, 2009: pp. 1–9. https://doi.org/10.1109/ACII.2009.5349483.

[17] I.H. Robertson, T. Manly, J. Andrade, B.T. Baddeley, J. Yiend, Oops!': performance correlates of everyday attentional failures in traumatic brain injured and normal subjects, Neuropsychologia 35 (1997) 747–758. https://doi.org/https://doi.org/10.1016/S0028-3932(97)00015-8.

[18] C. Braboszcz, A. Delorme, Lost in thoughts: Neural markers of low alertness during mind wandering, Neuroimage 54 (2011) 3040–3047. https://doi.org/10.1016/J.NEUROIMAGE.2010.10.008.

[19] H. Jebelli, S. Hwang, S.H. Lee, EEG-based workers' stress recognition at construction sites, Autom Constr 93 (2018) 315–324. https://doi.org/10.1016/J.AUTCON.2018.05.027.

[20] J. Ruiz de Miras, A.J. Ibáñez-Molina, M.F. Soriano, S. Iglesias-Parro, Schizophrenia classification using machine learning on resting state EEG signal, Biomed Signal Process Control 79 (2023). https://doi.org/10.1016/j.bspc.2022.104233.

[21] M. Shahrokhishahraki, M. Malekpour, S. Mirvalad, G. Faraone, Machine learning predictions for optimal cement content in sustainable concrete constructions, Journal of Building Engineering 82 (2024). https://doi.org/10.1016/j.jobe.2023.108160.

[22] V. Asghari, M. Hossein Kazemi, M. Shahrokhishahraki, P. Tang, A. Alvanchi, S.-C. Hsu, Process-oriented guidelines for systematic improvement of supervised learning research in construction engineering, Advanced Engineering Informatics 58 (2023) 102215. https://doi.org/10.1016/j.aei.2023.102215.

[23] I. Mehmood, H. Li, Y. Qarout, W. Umer, S. Anwer, H. Wu, M. Hussain, M. Fordjour Antwi-Afari, Deep learning-based construction equipment operators' mental fatigue classification using wearable EEG sensor data, Advanced Engineering Informatics 56 (2023). https://doi.org/10.1016/j.aei.2023.101978.

[24] G. Lee, S. Lee, Importance of testing with independent subjects and contexts for machine-learning models to monitor construction workers' psychophysiological responses, J Constr Eng Manag 148 (2022) 04022082. https://doi.org/10.1061/(ASCE)CO.1943-7862.0002341.

Tracking algorithm for hazardous area in construction sites using PTZ cameras

Jaehwan Seong[1], Hyung-soo Kim[1] and Hyung-Jo Jung[1]

[1]Department of Civil and Environmental Engineering, KAIST, Korea
kevin8633@kaist.ac.kr, rlagudtn0934@kaist.ac.kr, hjung@kaist.ac.kr

Abstract

This paper proposes a PTZ (Pan-Tilt-Zoom) camera-based solution to enhance construction site safety, particularly for small to medium scale projects and those in the early or late stages of development. Traditional sensor-based methods often face challenges related to cost, installation time, and adaptability, limiting their suitability for dynamic construction environments. In contrast, the proposed system leverages PTZ cameras and deep learning techniques specifically Visual Place Recognition (VPR) and Monocular Depth Estimation (MDE) to dynamically redefine hazardous areas in real time as the camera's view changes. The system integrates multiple tracking strategies, including feature-based tracking, scene re-identification, depth-driven adjustments, moving object detection, and image stabilization, to adapt to diverse site conditions. Experimental validation in both virtual and real-world settings demonstrate high accuracy, achieving an average IoU of 0.8627 in virtual sites and 0.7463 in real sites, despite the increased variability in actual conditions. The virtual environment, developed using a game engine, enabled precise simulations of site dynamics such as lighting, weather changes, and object movements. Real site tests further validated the system's practical applicability, highlighting its potential for broader deployment in construction safety management. By enhancing hazard identification and monitoring capabilities, this approach contributes to more effective and comprehensive safety management in construction projects.

Keywords

Safety Management; Computer vision; Real-time hazardous area tracking; Pan-Tilt-Zoom (PTZ) cameras

1 Introduction

The construction industry has long been recognized as a sector with a high risk of occupational accidents due to several intrinsic factors: the use of large-scale equipment, complex processes, collaboration among various stakeholders, and prolonged project durations often conducted outdoors. These characteristics have continually drawn global attention to construction site safety, prompting sustained efforts toward improvement. According to the U.S. Occupational Safety and Health Administration (OSHA), there were 1,075 fatalities in the U.S. construction sector in 2023, accounting for over 20% of all workplace deaths, with falls representing the most frequent cause [1]. In China, data from the Ministry of Housing and Urban-Rural Development (MOHURD) show that between 2012 and 2019, approximately 4,100 accidents claimed 5,011 lives in the construction sector, with falls constituting 52.83% of these incidents. These statistics suggest that accidents frequently occur around static hazard areas (e.g., foundation excavations, edges, openings), underscoring a persistent risk in such environments [2]. Similarly, in Korea, there were 26,829 injuries reported in 2023, accounting for over 20% of all occupational accidents in the construction industry. Small-scale sites often lack systematic safety management measures, leading to a high incidence of falls and collapses [3].

A noteworthy aspect is that many of these fatal incidents occur in relatively predictable "static hazardous areas." That is, certain spatial features such as openings or edges of structures pose an inherent and constant risk throughout the construction process. When this latent hazard is compounded by managerial or technological limitations, the likelihood of accidents increases [4, 5]. As a result, continuously and efficiently monitoring static hazardous areas and detecting unauthorized entries into these areas in real time has emerged as a key priority for global construction safety management.

Against this backdrop, recent research in construction safety management has actively explored the use of computer vision (CV)-based technologies. Traditional manual surveillance in construction sites is labor-intensive, time-consuming, and susceptible to human judgment errors and observer fatigue. By contrast, computer vision technologies employ automated algorithms to analyze targets such as workers, equipment,

and structural conditions and have demonstrated utility in various applications, including worker safety monitoring, detecting facility defects, ergonomic posture assessments, and process monitoring [6-10]. Recently, approaches aimed at the real-time recognition of and intrusion detection into static hazardous areas have also emerged, such as the work by Mei et al., which proposed a more clearly defined hazardous area identification model based on computer vision [11]. Their system uses object detection to locate workers in each video frame and then determines the level of risk by checking whether the detected position falls within a polygon that represents a predefined hazardous area on the camera image.

However, most existing detection techniques for static hazardous areas rely on images from fixed cameras, requiring manual designation of these areas. The manually designated area is defined based on the image coordinates of the camera, and when the camera moves, the dangerous area is distorted and the worker's entry into the wrong area is detected. This approach presents significant constraints when applied in real-world scenarios. Nowadays construction sites are vast, and observation points change frequently as the project progresses, making it challenging to comprehensively monitor all areas through a static viewpoint. While the use of Pan-Tilt-Zoom (PTZ) cameras with adjustable fields of view is on the rise, each camera movement necessitates redefinition of the static hazardous area, a process that is both inefficient and lacking in adaptability as shown in Figure 1. In other words, current methods suffer from limited dynamic adaptability to changes in camera orientation, undermining both the reliability and practicality of real-time hazard monitoring.

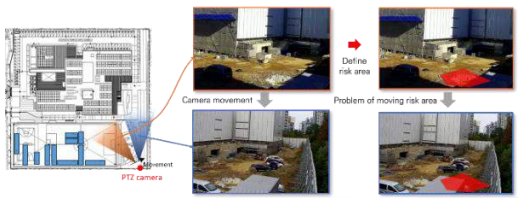


Figure 1. Problem of moving to a hazardous area

To address these limitations, this study proposes a safety management framework that estimates the pose of a PTZ camera using image data alone and automatically reconfigures predefined static hazardous areas according to camera movements. The key contribution lies in extracting spatial information solely from image input, without relying on additional sensors (e.g., IMUs), and using this information to dynamically map static hazardous areas in response to camera adjustments. As a result, the system can continuously track previously defined static hazard areas and determine whether intrusions occur, regardless of the camera's position and viewing angle. This approach is anticipated to improve both the operational flexibility of camera usage and the accuracy of monitoring in practical environments, thereby advancing real-time safety management systems that focus on static hazards. By applying the system proposed by Mei et al. to the tracked risk area, workers' access to the hazardous area can be detected. Moreover, by expanding the applicability of CV-based construction safety monitoring, this research may ultimately provide a practical method for managing various forms of risks across diverse project scales and processes in the future.

2 Proposed Framework

Construction sites are operated over extended periods, during which camera imagery is subject to significant variability due to equipment and worker movement, changes in workflows, illumination fluctuations, and weather conditions. To reliably track predefined hazardous areas under such variability, it is necessary to utilize robust algorithms against various factors such as camera motion, lighting changes, object rearrangements, and weather fluctuations. In this study, we propose a framework that, using only the footage captured from a PTZ (Pan-Tilt-Zoom) camera, estimates the camera pose and automatically reconfigures a predefined static hazardous area according to changes in the camera's viewpoint (see Figure 2).

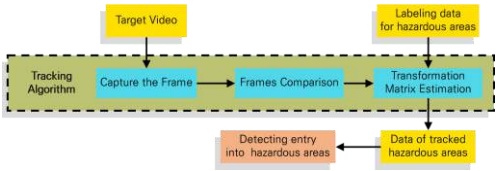


Figure 2. Framework of tracking algorithm for hazardous area

The proposed framework consists of three key stages:

1. **Camera poses estimation:** Using feature points extracting and matching technique, a transformation matrix between two viewpoints is computed. This method ensures real-time capability while maintaining relatively stable features matching under various environmental changes.
2. **Site verification based on Visual Place Recognition (VPR):** A VPR network is employed to determine whether the current image is from the same site where the hazardous area was previously defined. This allows for temporarily halting hazardous area tracking whenever the camera angle or position changes causes a completely different location to be captured, preventing safety incidents due to wrong identification.
3. **Robustness enhancement using depth information (MDE):** Depth information estimated

from a single image by a Monocular Depth Estimation (MDE) network is integrated into the tracking process. This approach enables stable background feature extraction despite lighting changes or temporary occlusions, thereby offering more robust adaptation to environmental changes when tracking the actual hazardous area.

2.1 Overview of the Framework

Figure 2 shows the overall structure of the proposed framework. First, features are extracted from images at two different points in time, then matched to estimate a transformation matrix. Traditional methods such as SIFT, SURF, and ORB are relatively easy to implement but are vulnerable to lighting changes and often struggle with speed-accuracy trade-offs. In contrast, the lightweight deep learning-based approach (XFeats) proposed by Potje et al. enables a better balance between real-time processing and accuracy [12]. By estimating camera pose changes between frames using this method, the static hazardous area is reprojected onto the image according to camera movements, enabling continuous tracking.

However, relying solely on feature matching can lead to erroneous matches caused by environmental changes (fog, rain, occlusion, changes in wall layout), ultimately resulting in incorrect hazardous area projections as shown in Figure 3. To mitigate this, the VPR network is employed to determine whether the captured footage is from the same site. If it is not, hazardous area tracking is suspended to prevent safety issues stemming from erroneous tracking. In addition, MDE is utilized to strengthen the structural understanding of the background beyond mere image-based features. By integrating depth information obtained from a single image, the framework achieves greater robustness against environmental changes during feature matching and VPR.

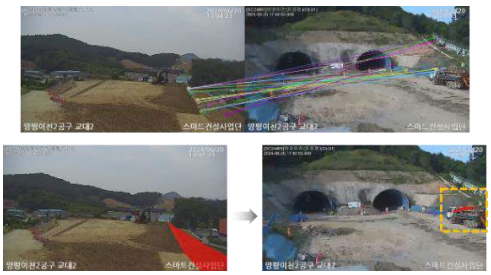


Figure 3. Failure of tracking due to PTZ camera operation (red area)

2.2 Estimation of Relative Camera Pose

Camera poses estimation is based on feature extraction and matching. Traditional methods (e.g., SIFT, SURF, ORB) are human-interpretable and easy to implement, but are sensitive to lighting and scale changes, and present a trade-off between speed and accuracy. Following Melekhov et al. CNN-based feature extraction and matching techniques have been explored [13]. Potje et al. introduced a lightweight approach, XFeats, combining real-time performance with accuracy [12]. In this study, XFeats is adopted to ensure stable tracking under diverse environmental changes in large-scale construction sites. Using the estimated transformation matrix, the camera pose is continuously updated, and the static hazardous area is reconfigured in real-time.

2.3 Visual Place Recognition (VPR)

VPR is a core component for identifying whether the captured images are from the same site where the hazardous area was previously defined. Traditional feature-based methods are highly vulnerable to changes in lighting (day/night) and weather (clear/rain/snow), resulting in misinterpretations that consider the same site as a completely different space. To address this issue, this study references the latest CNN- and Transformer-based VPR approaches [14, 15]. By employing the MixVPR method, high-dimensional features are mixed to generate a global descriptor. The similarity between global descriptors allows robust determination of whether the current site matches the previously defined hazardous area location.

Using the VPR results, the proposed framework maintains hazardous area tracking if it is the same site or suspends tracking or updates the reference images if not, preventing safety management issues caused by erroneous matches. The incorporation of MDE-derived depth information assists in robustly capturing background structures, further enhancing resilience to weather changes or temporary occlusions and improving overall recognition performance.

2.4 Monocular Depth Estimation (MDE)

Monocular Depth Estimation (MDE) is a technique that estimates the distance to each pixel in an image using only a single RGB camera input. It can be leveraged to acquire three-dimensional environmental information from construction sites without expensive equipment or complex installations, effectively serving as a substitute for RGB-D cameras. By employing an encoder-decoder neural network architecture, spatial features are extracted, and depth is predicted from the input image. Incorporating LSTM can capture temporal features, enabling more stable depth estimation. In addition, training on ground-truth depth maps that incorporate physical constraints ensures that the estimated depth is robust to environmental changes such as lighting variations, weather conditions, and occlusions.

In this study, a Monocular Depth Estimation (MDE)

network with a DINOv2 encoder and a DPT decoder is employed. We use a model pretrained on large-scale unlabeled data as demonstrated by Yang et al. [16]. This provides effective depth estimation, which, when integrated into feature matching and VPR, enhances the stability and accuracy of hazardous area tracking in construction sites.

3 Experiments

This chapter explains the preparations and approaches used in the experiments for the proposed PTZ camera-based static hazardous area tracking algorithm. It also defines the evaluation indicators that will be used to interpret the experimental results presented in Section 4. To overcome the challenges of collecting and labelling video data from actual construction sites while minimizing safety management burdens, a virtual construction site environment was developed using the Unity game engine. This approach enabled the simulation of realistic variables, including camera operations (pan, tilt, zoom), weather, and lighting changes, while ensuring accurate labelling information and controlled hazardous area configurations. The final performance was verified by applying the approach, which showed the highest performance in the virtual construction site, to the actual field.

3.1 Implementation Details

Section 3.1 describes the implementation details of the algorithms used in the experiment, the configuration of the virtual environment, and the process of labeling hazardous areas.

3.1.1 Virtual Construction Site Implementation

A virtual construction site was built using Unity as shown as Figure 4, incorporating various architectural elements (structures, equipment, terrain features, etc.), changes in lighting over time (day/night transitions), weather variations (rain, fog), and object movements. A PTZ (Pan-Tilt-Zoom) camera was set up to mimic actual site operations by randomly altering camera poses. Objects, such as equipment and workers, moved unpredictably, reflecting the dynamic nature of active construction sites.

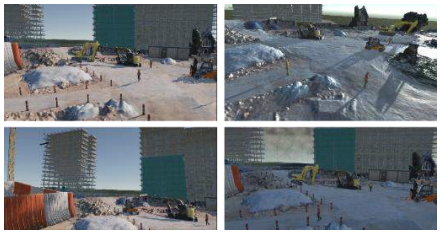


Figure 4. Virtual construction site

Camera operation times along a virtual timeline were categorized into four segments: less than 10 seconds (brief), 10 to 30 seconds (short), 30 seconds to 1 minute (mid), and over 1 minute (long). This categorization aimed to evaluate tracking performance over extended periods.

3.1.2 Hazardous Area Labeling and Ground Truth Generation

Within the virtual environment, certain areas were predefined as “static hazardous areas” and used as ground truth. As the camera moved, the projective representation of these hazardous areas changed. The proposed algorithm continuously reprojects these areas in real-time based on camera movement. The lighting environment was shifted over a 100 s day/night cycle, and intermittent rainfall was introduced to replicate weather-induced visibility changes. The ground truth information was utilized in subsequent evaluations (see Section 3.3 for evaluation metrics).

3.1.3 Detail Parameter of Applied Model

Feature extraction and matching for localization and tracking are performed using XFeats, a CNN-based feature extraction and matching model designed for real-time operation in resource-constrained environments. XFeats employs a lightweight CNN backbone that maintains a low number of channels, optimizing computational efficiency while preserving high-resolution feature representations. The model was pretrained on a combination of the Megadepth dataset and synthetically warped images from COCO to enhance its robustness. To further improve structural feature representation, MixVPR, a CNN-based descriptor specifically designed to handle severe illumination and weather variations, is utilized. MixVPR was pretrained on the Pitts30k-train dataset, and its encoder is employed to extract descriptors that capture the structural characteristics of each image. The cosine similarity between these descriptors is then used to assess structural similarity across images, enabling the determination of on/off states for risk area tracking. Additionally, for monocular depth estimation (MDE), a pretrained model was adopted following the approach proposed by Yang et al., as discussed in Section 2.4.

3.2 Experimental Setup and Approaches

This section outlines various approaches used to implement the proposed tracking mechanism. These include frame-to-frame comparison techniques, the use of a reference image, VPR (Visual Place Recognition) methods, and strategies for flexible hazardous area activation/deactivation by incorporating MDE-based depth information. Five approaches were proposed, and the structure of each is illustrated in Figure 5.

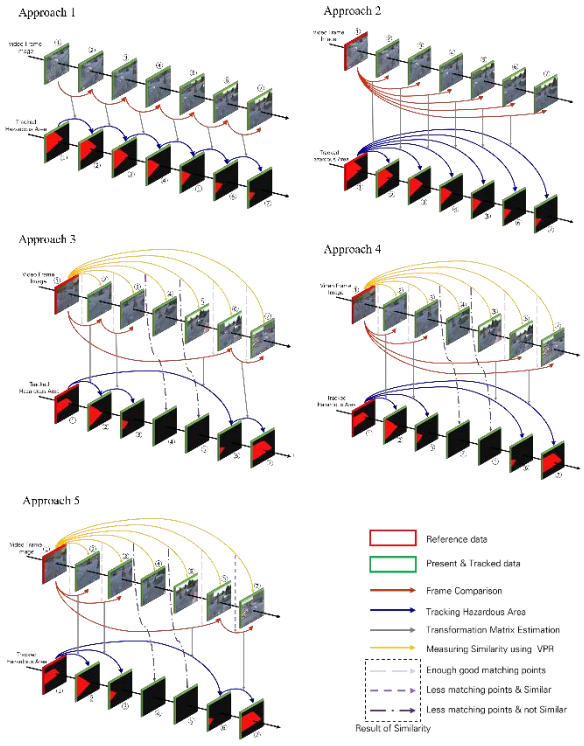


Figure 5. Architecture of each approach

3.2.1 Simple Transformation Approaches (Approaches 1 & 2)

Approach 1 (Continuous Frame Comparison) involves estimating the transformation matrix between consecutive frames for hazardous area tracking. This approach, commonly used in SLAM, effectively handles small incremental changes but suffers from cumulative errors over time, presenting challenges for long-term tracking.

Approach 2 (Reference Image-Based) defines the first frame as a reference and directly estimates the transformation matrix between the current frame and the reference frame for reprojection. This approach avoids long-term error accumulation but may face difficulties when the scene differs significantly from the reference image, making direct matching challenging.

3.2.2 VPR for Scene Change Adaptation (Approaches 3 & 4)

In construction sites, the camera may capture different scenes where the previously defined hazardous area no longer applies. To address this, 2 types of VPR methods were applied. Type 1 determines similarity based on feature points, while Type 2 integrates image descriptors with depth information from MDE. By evaluating whether the current frame and the reference frame depict the same location, the hazardous area can be deactivated when shooting a different site, mitigating tracking errors caused by scene changes. Approach 3 is a method of introducing the VPR plan based on Approach 1, and Approach 4 is a method of introducing the VPR plan based on Approach 2

3.2.3 Hybrid Approach (Approach 5)

Approach 5 overcomes the limitations of Approaches 1–4 by dynamically selecting between reference image-based tracking and VPR-guided results. If enough matching points are extracted, reference image-based tracking (Approach 2) is used; otherwise, previous-frame-based tracking (Approach 1) is employed. At the same time, VPR results guide the deactivation of hazardous areas when the scene changes. The integration of VPR and MDE enables stable activation and deactivation of hazardous areas, even under varying environmental conditions.

3.3 Evaluation Indicators

In this section, we define the quantitative evaluation metrics for interpreting the experimental results presented in Section 4.

3.3.1 Intersection over Union (IoU)

The accuracy and stability of hazardous area tracking are evaluated using various indicators. The performance of hazardous area tracking is quantitatively assessed using the Intersection over Union (IoU) metric and several supplementary indicators. First, the IoU (see Equation (1)) is the most representative indicator of similarity between the actual hazardous area (R_t) and the tracked hazardous area (R_g). When comparing the tracking results and the actual hazardous area in each frame.

$$IoU = \frac{R_t \cap R_g}{R_t \cup R_g} \quad (1)$$

An IoU value closer to 1 indicates a higher overlap between the two areas, signifying a closer match between the tracked and actual hazardous areas. However, to comprehensively evaluate performance across the entire video, a single frame value of IoU is insufficient. Therefore, the following additional indicators are employed to assess overall tracking performance and reliability. This metric evaluates the similarity between the ground truth hazardous area and the area reprojected by the tracking algorithm. It represents the ratio of the intersection to the union of the two areas, and a higher IoU value indicates greater tracking accuracy. Through the average IoU across all frames ($\overline{IoU}$), the ratio of frames where $IoU \geq 0.80$ (IoU_{80}), and the ratio of frames where $IoU \geq 0.95$ (IoU_{95}) the tracking performance can be evaluated from multiple perspectives.

3.3.2 Accuracy of Activation and Deactivation

When applying the VPR algorithm, changes in the scene can require the hazardous area to be deactivated; AoD measures the accuracy of deactivation in such cases. Conversely, in situations where the hazardous area should remain active, AoA indicates how accurately it remains activated. These metrics allow for an assessment of scene change adaptability and overall reliability in real-world site management.

3.3.3 Long-term Operational Stability

The degradation of the $\overline{IoU}$ based on the length of the video (brief, short, mid, and long) provides a view of the long-term tracking stability. We imply that the approach that does not significantly degrade the $\overline{IoU}$ for long videos compared to brief videos is the approach to maintain consistent performance in long-time operations. Therefore, the approach that does not significantly degrade for long videos compared to brief videos is more advantageous for construction site applications that require real-world long-term operations. These complementary metrics are used to compare and evaluate the performance of the proposed tracking Approaches (Approaches 1-5). Specific performance figures and comparison results are discussed in detail in Section 4.

4 Results

4.1 Results at Virtual Construction Site

4.1.1 Comparison of Basic Tracking Methods

Based on the $\overline{IoU}$, Approach 2, which tracks using a reference frame, showed a slightly better performance than Approach 1, which compares consecutive frames. In short and mid-length video scenarios, Approach 2 outperformed Approach 1. However, in long videos, both approaches exhibited similar performance, while in brief videos, Approach 1 performed slightly better. Overall, Approach 2 holds an advantage in terms of $\overline{IoU}$, but each approach comes with its own strengths and weaknesses when examining additional indicators. The performance of Approach 1 and Approach 2 is shown in Table 1 and shown as Figure 6.

4.1.2 Effect of Introducing VPR

Approaches 3 and 4 improved the $\overline{IoU}$, IoU_{80} and IoU_{95} compared to their respective counterparts, Approaches 1 and 2. Notably, Approach 4 demonstrated superior overall performance compared to Approach 3. However, in both approaches, AoA was observed to be significantly degraded compared to Approach1.

When comparing VPR types (Types 1 and 2), Type 1 accurately identified deactivation timing but often excessively deactivated hazardous areas, reducing its AoD value. In contrast, Type 2 achieved a higher AoA value, allowing for more reliable detection of actual worker access. However, since neither directly accounts for transformation inaccuracies, there are some limitations in determining the precise deactivation timing. The performance of each approach is detailed in Tables 2 and 3 and shown as Figure 6.

Table 1. Performance by tracking Approach 1 and 2

Tracking Method		Approach 1	Approach 2
$\overline{IoU}$		0.7214	0.7460
$\overline{IoU}_{type}$	brief	0.8513	0.8165
	short	0.7859	0.9127
	mid	0.7372	0.7908
	long	0.6898	0.6833
IoU_{80}		54.54%	76.11%
IoU_{95}		39.43%	54.92%
AoD		90.13%	58.80%
AoA		97.87%	89.73%

Table 2. Performance by tracking Approach 3

Tracking Method		Type 1	Type 2
$\overline{IoU}$		0.5988	0.6908
$\overline{IoU}_{type}$	brief	0.8193	0.7848
	short	0.6876	0.7823
	mid	0.5895	0.7110
	long	0.5662	0.6541
IoU_{80}		52.92%	53.26%
IoU_{95}		42.48%	38.59%
AoD		96.48%	87.84%
AoA		56.77%	86.82%

Table 3. Performance by tracking Approach 4

Tracking Method		Type 1	Type 2
$\overline{IoU}$		0.8326	0.7606
$\overline{IoU}_{type}$	brief	0.8820	0.8355
	short	0.9143	0.9168
	mid	0.8714	0.7934
	long	0.7930	0.7055
IoU_{80}		84.68%	77.58%
IoU_{95}		65.66%	56.44%
AoD		99.89%	62.98%
AoA		77.62%	89.46%

4.1.3 Performance of the Hybrid Approach

Approach 5 is a hybrid approach that dynamically decides whether to use the reference frame based on the number of matching points and relies on VPR results for activation and deactivation. When combined with descriptor-based VPR, this method achieved the highest $\overline{IoU}$ and most stable performance metrics overall. Approach 5 recorded top scores in most of the evaluated indicators. This highlights that integrating depth

information and adopting a flexible reference-frame selection strategy significantly improves long-term stability and accuracy in hazardous area tracking, especially under the complex, dynamic conditions of construction sites. Detailed performance results for each method can be found in Table 4 and shown as Figure 6.

Table 4. Performance by tracking Approach 5

$\overline{IoU}$		0.8627
$\overline{IoU}_{type}$	brief	0.9144
	short	0.9322
	mid	0.8789
	long	0.8360
IoU_{80}		87.12%
IoU_{95}		66.26%
AoD		97.20%
AoA		87.11%

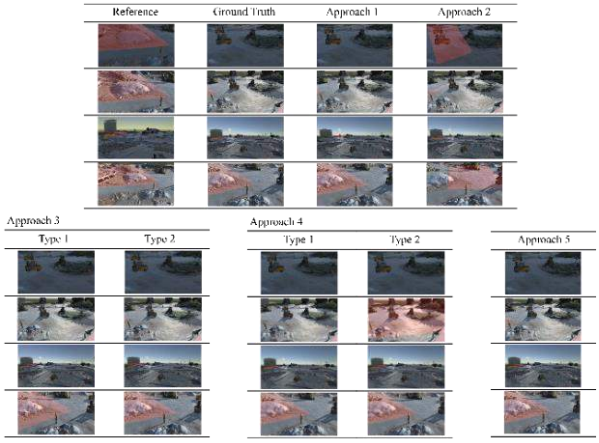


Figure 6. Tracking result at the virtual site

4.2 Results at Real Construction Site

The proposed algorithm (Approach 5 which is the highest performance method in virtual sites, illustrated in Figure 7) was validated using a PTZ camera at an actual construction site. For 20 videos collected from the PTZ camera at 13 sites, the initially set hazardous area and the hazardous area moved after the PTZ camera operation were manually labeled. In 20 videos, the hazard areas were manually defined based on the first frame, and the ground truth for the final tracking results was also manually established using the last frame. To evaluate the tracking performance, we calculated the IoU between the ground truth and the hazard areas tracked by the proposed method.

Performance was measured by comparing the hazardous area tracked by the algorithm with the ground truth of the hazardous area after PTZ camera operation. The tracking algorithm achieved average performance score of 0.7463 across 20 videos, with detailed results illustrated in Figure 8.

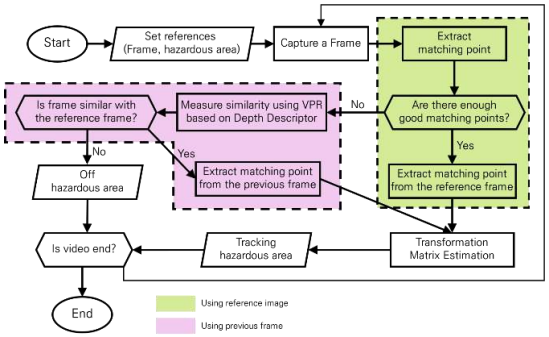


Figure 7. Flow chart of algorithm to track the hazardous area for PTZ camera

The complexity of real sites that was not considered in the virtual sites resulted in a lower result than the performance in the virtual sites. In general, segmentation models evaluated using the IoU metric are considered highly accurate when mean IoU exceeds 0.7. Based on this standard, our proposed algorithm appears sufficiently robust for hazard zone tracking in construction sites. However, from a safety management perspective, further improvements in tracking accuracy will be necessary for more precise hazard detection.

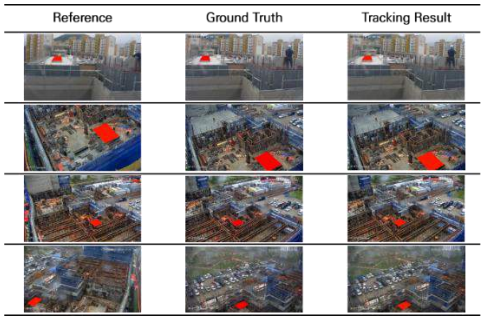


Figure 8. Tracking result at the real site

5 Conclusion

In this study, a hazardous area tracking algorithm was developed for PTZ cameras operated at construction sites. The tracking algorithm's performance was validated in both simulated and real-world construction site environments. The proposed algorithm was implemented by extracting well-matched point pairs from the reference image, deciding on reference image usage for transformation estimation, and activating hazardous areas based on descriptor cosine similarity. The proposed algorithm was verified at a virtual construction site and actual construction.

The proposed system achieved a tracking performance score of 0.8627 and an IoU_{80} of 87.12% in

virtual construction site simulations. In addition, by applying the estimation algorithm to actual construction sites, the average IoU was 0.7463, confirming that it could be properly applied to the actual site. As a result, a smart safety system that can detect construction workers' access to hazardous areas can be applied to PTZ cameras, thereby enhancing the safety of construction sites. The system is also expected to enable the projection of virtual object information for construction site safety management without requiring separate control points or sensors, thus contributing to more efficient site management.

This study focused solely on tracking methods for 2D objects. However, hazardous areas at actual construction sites often appear in the form of 3D, and it is necessary to implement the tracking function for 3D objects in subsequent studies. Additionally, while the hazardous area tracking algorithm was implemented, future research should evaluate its accuracy in detecting worker access to hazardous areas at real construction sites, especially when integrating previously proposed access detection algorithms.

Acknowledgements

This work is supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant RS-2020-KA156208)

References

[1] OSHA (Occupational Safety and Health Administration). Commonly used statistics. On-line: https://www.bls.gov/news.release/cfoi.nr0.htm, Accessed: May 03, 2025.

[2] MOHURD (Ministry of Housing and Urban-Rural Development). MOHURD's notification on the safety accidents in housing and municipal works in 2019. [In Chinese.] Stand. Eng. Constr. 2020 (7): 51–53, 2020.

[3] Kim, D. Status of industrial accidents in 2021. [In Korean.] Ulsan: Korea Occupational Safety & Health Agency. Status of industrial accidents in 2023, 2024

[4] Fang, Q., Li, H., Luo, X., Ding, L., Luo, H., & Li, C. Computer vision aided inspection on falling prevention measures for steeplejacks in an aerial environment. Automation in Construction, 93, 148-164, 2018

[5] Fang, W., Zhong, B., Zhao, N., Love, P. E., Luo, H., Xue, J., & Xu, S. A deep learning-based approach for mitigating falls from height with computer vision: Convolutional neural network. Advanced Engineering Informatics, 39, 170-177, 2019

[6] Hu, Q., Bai, Y., He, L., Cai, Q., Tang, S., Ma, G., ... & Liang, B. Intelligent framework for worker-machine safety assessment. Journal of Construction Engineering and Management, 146(5), 04020045, 2020

[7] Kim, H. [Construction Safety A to Z] Smart Construction Safety Using AI. [In Korean.] Daehan Economic Daily, 2023

[8] Koch, C., Georgieva, K., Kasireddy, V., Akinci, B., & Fieguth, P. A review on computer vision-based defect detection and condition assessment of concrete and asphalt civil infrastructure. Advanced engineering informatics, 29(2), 196-210, 2015

[9] Ray, S. J., & Teizer, J. Real-time construction worker posture analysis for ergonomics training. Advanced Engineering Informatics, 26(2), 439-455, 2012

[10] Han, K. K., & Golparvar-Fard, M. Appearance-based material classification for monitoring of operation-level construction progress using 4D BIM and site photologs. Automation in construction, 53, 44-57, 2015

[11] Mei, X., Zhou, X., Xu, F., & Zhang, Z. Human intrusion detection in static hazardous areas at construction sites: Deep learning–based method. Journal of Construction Engineering and Management, 149(1), 2023

[12] Potje, G., Cadar, F., Araujo, A., Martins, R., & Nascimento, E. R. XFeat: Accelerated Features for Lightweight Image Matching. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024

[13] Melekhov, I., Ylioinas, J., Kannala, J., & Rahtu, E. Relative camera pose estimation using convolutional neural networks. In Advanced Concepts for Intelligent Vision Systems: 18th International Conference, ACIVS 2017, Antwerp, Belgium, September 18-21, 2017, Proceedings 18 (pp. 675-687). Springer International Publishing, 2017

[14] Zhu, S., Yang, L., Chen, C., Shah, M., Shen, X., & Wang, H. R2former: Unified retrieval and reranking transformer for place recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023

[15] Ali-Bey, A., Chaib-Draa, B., & Giguere, P. Mixvpr: Feature mixing for visual place recognition. In Proceedings of the IEEE/CVF winter conference on applications of computer vision, 2023

[16] Yang, L., Kang, B., Huang, Z., Xu, X., Feng, J., & Zhao, H. Depth anything: Unleashing the power of large-scale unlabeled data. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024

Dynamic Hand-Signal Recognition and Prediction Framework for Crane Operations Using Deep Learning

Zhihao Wei[1,2], Ghulam Muhammad Ali[2], and Xinming Li[1]

[1]School of Mechanical Engineering, University of Alberta, Canada
[2]SensiImage Technologies Ltd. Canada
wzhihao@ualberta.ca, gmail@sensiimage.com, xinming.li@ualberta.ca

Abstract –

The study proposed a real-time sensor-based hand gesture signal classification framework to address communication challenges between signalmen and crane operators on construction sites. The system collects data from a 3-axis accelerator, 3-axis gyroscope, and three flex sensors installed on gloves to recognize and predict dynamic hand signals. A dedicated data structure is developed to capture temporal characteristics of gestures by combining five consecutive time step sensor data, making sure that the deep learning model can differentiate between complex, sequential gestures. A hybrid CNN-LSTM deep learning model is employed for gesture recognition, and an LSTM-based model is used for predicting sensor data for subsequent signals, helping crane operators improve operation accuracy. The proposed framework successfully classifies 18 predefined crane operation hand signals with 98% overall accuracy, and the prediction process achieves a R^2 0.872. For the real situation test, the framework achieves a correct classification rate of 95.72%. These results demonstrate the effectiveness and robustness of deep learning models with dynamic data structure in addressing communication challenges and improving work safety in crane operations.

Keywords –

Crane operations; Deep learning; Sensor-based systems; Safety in construction.

1 Introduction

Cranes have become indispensable tools on construction sites, especially with the growing adoption of prefabricated components and modular construction techniques. In crane operations, effective communication is vital to ensuring safety and efficiency. Traditionally, hand signals and two-way radios are used for communication, with crane operations being entirely dependent on the signalman's instruction.

However, the complexity of construction sites makes signal communication challenging. The main barriers include low visibility caused by obstructions of large size prefabricated modular components and weather conditions, as well as low clarity of radio communication caused by excessive construction site noise, low signal stability, and diverse workers accents. These difficulties increase the likelihood of miscommunication and the failure to execute instructions correctly. For crane operations, even the slightest error can result in fatal consequences. In fact, miscommunication between crane operator and signalman is a major factor in the high rate of fatalities in the construction industry [1]. According to the US Bureau of Labor Statistics 2017, the average number of fatalities in crane operations is 42 per year in the past 7 years [2]. Furthermore, a study [3] found that 34% of crane-related fatalities are caused by signalman and operator errors, with miscommunication accounting for 11% of these cases.

The alarming statistics highlights the need for a smart system to capture real-time signalman hand-signal to enhance existing communication methods and improve worker safety. In the presented study, we address this gap by proposing a real-time sensor-based hand signal classification framework for crane operations. A detailed deep learning algorithm to improve the classification accuracy of dynamic gesture recognition is developed and validated, while data with comprehensive temporal information, collected from HandCoMM glove set, is used to train the model. Additionally, a prediction process is also implemented for providing crane operators with more reaction time, enabling them to make the correct decisions and operations.

The proposed system can recognize signals, as illustrated in Figure 1. It is worth noting that an additional gesture, "*standby/activate*," is used to initialize the system proposed in this study. This gesture is performed with the forearm vertical and the index and middle fingers pointing upward. However, it is not shown in Figure 1.



Figure 1. Hand signal for hoist and crane operations

2 Literature Review

2.1 Internet of Things (IoT) in crane operation

Internet of Things technology has been developed, tested, and used in crane operation in construction sites. For instance, IoT technologies such as RFID [4] and GPS [5] positioning are used to integrate crane resources, which can reduce latency and improve efficiency in construction sites. Automation in crane operations is also a hotpot of IoT applications, X. Li et al. [6] used laser scanner and 4G communication to realize automatic crane operation for a prefabricated part. With AI-decision implementation, an automatic crane operation model has been achieved in a digital simulation model [7].

Sensor-based wearable devices represent another promising development direction. By integrating sensors into workers' personal protective equipment (PPE) rather than mounting them on crane arms or hooks, IoT systems can enable faster responses and provide more accurate hazard warnings [8], signal transmission [9], and other functionalities.

2.2 Deep Learning (DL) in crane operation

In crane operations, computer vision is one of the most widely applied deep learning technologies. Plenty of DL models, like YOLO (You Only Look Once), CNN (Convolutional Neural Network), are applied to achieve automation and intelligentization. CNNs are particularly effective in computer vision tasks, as they excel at capturing spatial features and local dependencies within image data. Some gesture recognition tasks rely on CNN-based architectures [15], while others use CNNs to extract combined feature representations from sensor data [14]. In addition, Long Short-term Memory (LSTM) model is crucial technology for processing time-series data. Many LSTM-based models are employed to achieve specific functions in construction sites. Wang et al. [10] used an optimized LSTM to achieve real-time and precise locating of the crane hook. LSTM is also applied to monitor and predict crane movement path [11], ensuring efficient crane operation planning and real-time adjustment.

2.3 Hand gesture recognition

With the widespread development of IoT technology and deep learning, more advanced systems for hand gesture recognition have further integrated to enhance the safety and efficiency of crane operations. Among these studies, vision-based sensors are the one of the most widely applied, with many efforts dedicated to combining visual signal with deep learning networks to achieve efficient hand signal recognition. Wang et al. [12] conducted a comparative study on different DL architectures, including ResNeXt-101 and Res3D + ConvLSTM + MobileNet, to evaluate their performance in image-based hand gesture recognition tasks in construction site. In addition to visual sensors, IMU sensors have been widely used to capture dynamic hand motion information. Mansoor et al. [1] integrated IMU sensors with traditional machine-learning algorithms and LSTM-based deep learning models for crane hand signal classification. Their study successfully demonstrated the feasibility of using wearable glove-based sensor layouts for reliable gesture recognition. Many other types of sensors have been selected to adopt to gesture recognition on under different conditions. EMG [13] helps to reduce restrictions on finger movement, flex sensors [9] improve the recognition of complex finger movements. LiDAR [14] minimizes the influence of lighting condition. Another effective approach is multimodal hand gesture recognition, by combining eye-tracking and IMU sensor [15], and adding context-aware sensing [16], researchers achieved more efficient the hand signal communication.

Building on these prior studies, this paper optimizes the use of accelerometers, gyroscopes, and flex sensors, while introducing a custom time-series data structure to better capture the temporal dynamics of hand gestures. Furthermore, we develop a CNN-LSTM hybrid model for precise gesture classification and incorporate an LSTM-based prediction module to anticipate future hand signals. This predictive capability extends the crane

operator's reaction time, enabling more accurate and timely execution of commands, ultimately enhancing operational safety and efficiency.

3 Research Methodology

The methodology for this study has been structured into three main parts: system design, data structure and preprocess, and deep learning model and implementation.

The overall workflow in the study is illustrated in Figure 2:

Step 1: Data is collected by sensors installed in the glove and transmitted to the central device.

Step 2: The data undergoes auto labeling, preprocessing and data structure management.

Step 3: The hybrid CNN-LSTM recognition model and LSTM prediction model are trained and tuned using processed data.

Step 4: The trained and fine-tuned models are deployed on the central device to enable real-time gesture signal recognition and prediction.

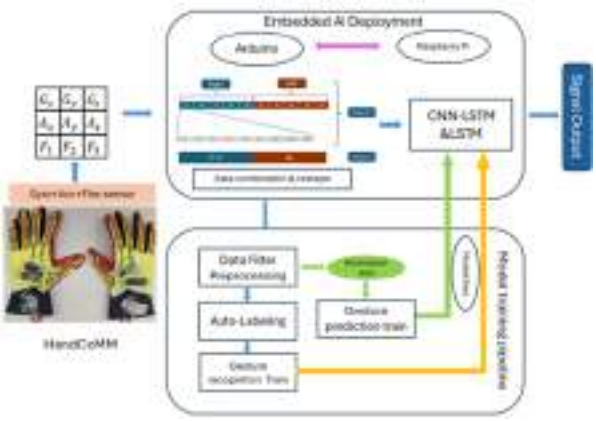


Figure 2. Dynamic crane signal recognition and prediction workflow.

3.1 System Design

The system consists of two main hardware components: the HandCoMM glove set, and a central device mounted on a belt. The glove set is responsible for collecting gesture signal data, while the central device handles data preprocessing for deep learning model and serves as the host for the deployed deep learning model using the preloaded weighted model file.

As shown in Figure 3, three flex sensors are strategically positioned on the thumb, index finger, and middle finger to precisely detect finger bending. As the ring and pinky fingers tend to move together with the middle finger, their bending pattens can be inferred from the flex sensor on the middle finger. This simplification not only reduces sensor complexity while maintaining high accuracy in gesture recognition but also lowers the communication burden and minimizes transmission delay. An Arduino development board, integrated with an accelerometer and gyroscope, is securely mounted on the back of the glove near the wrist. This configuration enables the system to monitor a wide range of hand movements, including hand pitch, wrist rotation, arm rotation, and the overall direction, speed, and acceleration of the hand. The battery and related accessories are securely fixed at the base of the glove.

The central device comprises an Arduino development board, a Raspberry Pi, and related supporting accessories etc. battery and power management unit. The gloves communicate with the central device via the Bluetooth protocol, while the Raspberry Pi and Arduino exchange data through a serial connection. These components are integrated and mounted on a belt to ensure the signalman's comfort during wear.

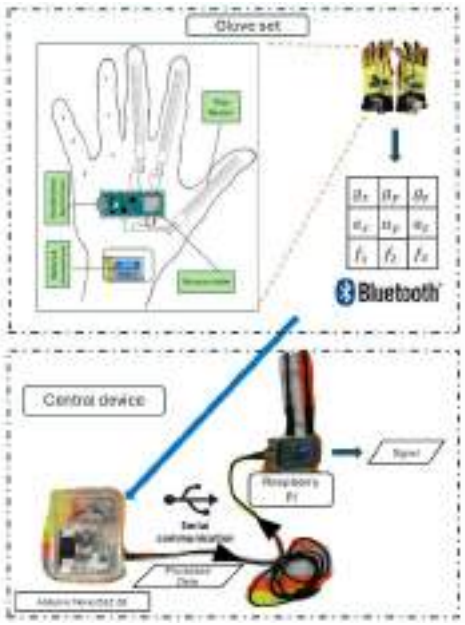


Figure 3. Layout of HandCoMM glove [17], [18] and system setup.

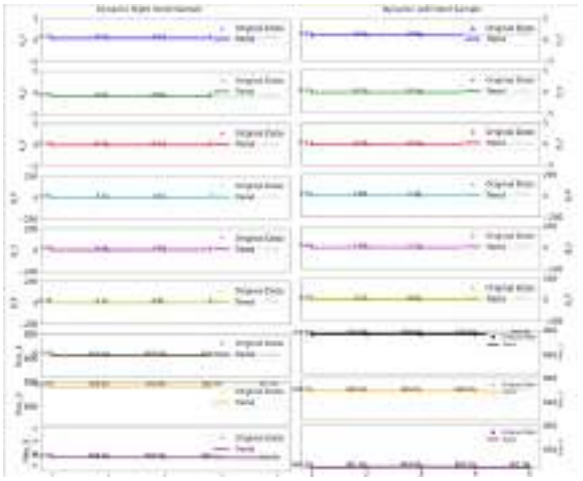
In this study, all devices are worn by the signalman. The gloves collect data from the accelerometer, gyroscope, and flex sensors, which are then transmitted via Bluetooth 5.0 Low Energy (BLE) to an Arduino development board for integration and preprocessing. This data transmission method ensures real-time performance and stability over short distances. The maximum Bluetooth packet size per transmission is 247 bytes, while the study's data size is approximately 200 bytes per packet, resulting in a transmission delay of just 1-2ms, depending on the communication mode. With a continuous communication interval of 7.5ms, this delay

is negligible compared to the 4-6Hz frequency of hand movements (150ms-250ms per motion cycle), ensuring smooth and uninterrupted communication. Although Bluetooth communication has a typical range of 10 meters, the actual distance between the gloves and the central device during operation is less than 1 meter, further enhancing the stability and reliability of the connection. The prepared data is subsequently sent to the central device (Raspberry Pi) via serial communication for deep learning inference, where gesture signals are generated and transmitted.

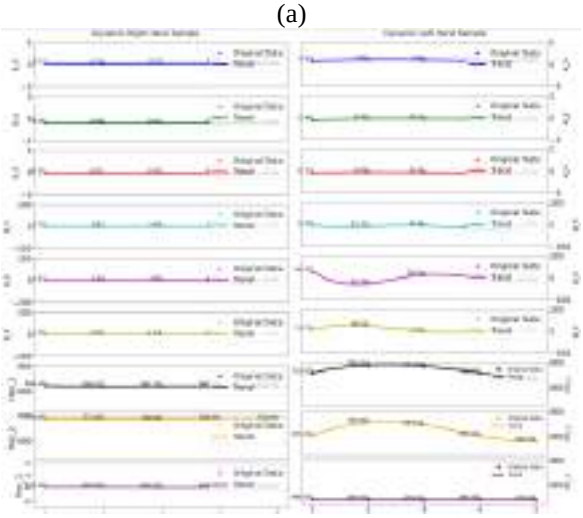
3.2 Data Structure and Preprocess

The dataset is designed to capture the temporal dynamics of hand gestures, essential for distinguishing complex dynamic gestures. Unlike static gestures, dynamic gesture signals require temporal context to represent transitions and movements. For example, the "*Boom up*" signal involves making a fist with the thumb raised, while "*Boom up hoist down (Option 1)"* requires sequentially bending four fingers while maintaining the thumb's position. Since these gestures share certain hand positions, single time-point data cannot differentiate them effectively. To address this, each sample integrates data from five consecutive time points, encoding the gestures' temporal characteristics. In Figure 4, the structural differences in sensor data between the two types of hand gesture signals are illustrated.

Each glove records nine types of data per time point: gyroscope readings (G_x, G_y, G_z), accelerometer readings (A_x, A_y, A_z), and flex sensor readings from the index finger, middle finger, and thumb (flex_1, flex_2, flex_3). Combining five consecutive time points results in a 45-dimensional feature vector per glove, and merging data from both gloves produces a 90-dimensional dataset per sample. With a sampling frequency of 30 Hz, the system captures samples at 6 Hz after combining five time points, ensuring efficient data processing and remaining gesture nuances. The selection of five-time steps in this study is determined by multiple factors. Research on human motion sensor sampling frequencies has shown that a sampling rate between 20Hz and 40Hz [19] provides a more accurate representation of human movements. Based on this, a baseline sampling frequency of 30Hz was adopted in this study. Additionally, studies on hand movement frequencies indicate a range of 1Hz to 4.5Hz [20], [21]. Considering the time required for data transmission, the effective sampling frequency was set to 6Hz, leading to the choice of five consecutive time steps. Furthermore, selecting a longer time window would increase communication time, as there are limitations on the data packet size per transmission. Exceeding this limit would require data segmentation, significantly increasing communication latency. As shown in Figure 4(b), five-time steps sufficiently capture the transition from finger bending to extension, including the two noticeable gyroscope fluctuations caused by the full bending and straightening of the fingers, as well as the rise and return of flex sensor values for the index and middle fingers.



(a)



(b)

Figure 4. Dynamic feature of sensor data comparison. (a) Gesture "*Boom up*". (b) Gesture "*Boom up Hoist down (Option 1)*"

For prediction tasks, due to the significant differences in absolute values among different sensor data, additional scaling is required to enhance the accuracy of the model's predictions. The gyroscope and accelerometer data possess directional properties, meaning their values can be both positive and negative; therefore, they are normalized within the range of (-1, 1). In contrast, the flex sensor measures resistance values, which are always positive, so it is normalized within (0, 1). The normalization process was conducted using the following formula:

$$x' = a + \frac{(x - x_{min})}{(x_{max} - x_{min})} \cdot (b - a)$$

Where x represents the original sensor value, x_{min} and x_{max} are the minimum and maximum values of the data, a and b are the lower and upper bounds of the target range, x' is the normalized value. Then generated data by the model could be reversed to its true value for analysis and practical applications.

Each dataset sample is labeled with one of 18 predefined crane signaling gestures. Stringent quality control measures were applied to ensure the data's integrity, removing incomplete or corrupted samples. The labeled data is stored in CSV format for accessibility and structured analysis.

The dataset includes contributions from six volunteers, reflecting diverse crane signaling experiences. This diversity ensures the dataset represents realistic variations in hand gestures. The final dataset comprises 13,514 samples, providing a robust foundation for developing deep learning models capable of recognizing dynamic gestures accurately in real-world scenarios.

3.3 Deep learning Model and Implementation

In the present study, deep learning is employed to achieve two critical objectives: gesture recognition and gesture prediction. Based on the data from flex sensors and IMU mounted on gloves, deep learning model will extract the features and time sequence characteristics to recognize different hand signal and generate sensor data at time step t+1. The section provides the details of deep learning models and implementation used in the proposed paper.

3.3.1 Model architecture and implementation: gesture recognition

To achieve robust gesture recognition, a combined CNN-LSTM model is employed. This architecture takes advantage of the strengths of CNNs in spatial feature extraction and LSTM in modeling temporal dependencies. Below, the model components and their functionalities are outlined:

- 1D-CNN Module

The 1D-CNN layers are designed to extract spatial features from the input sensor data, capturing correlations among the mounted sensors at each time step. This means CNN can identify static patterns. For example, if G_x and G_y increase together, it may signal a clockwise rotation. Through convolutional and pooling operations, the module reduces the dimensionality of the data while retaining its essential features. The first convolutional layer employs 64 filters with a kernel size of 3, followed by batch normalization to stabilize and accelerate training, max-pooling for dimensionality reduction, and dropout for regularization. A second convolutional layer, consisting of 128 filters, further enhances the model's ability to extract complex spatial features.

- LSTM Module

The LSTM module processes the sequence of features extracted by the CNN module to capture temporal dependencies and dynamic patterns in the sensor data. A single LSTM layer with 100 units is implemented, allowing the model to learn both short-term and long-term dependencies across time steps. To mitigate overfitting and improve generalization, a dropout rate of 0.5 is applied.

- Implementation

A Global Max Pooling layer was used to retain the most critical temporal features, followed by a dense layer with 256 units and ReLU activation to capture high-level representations. The final output layer employed SoftMax activation to classify the data into 18 predefined gesture categories.

The model was trained using the Adam optimizer with a learning rate of 1×10^{-3} and categorical cross-entropy loss. An early stopping mechanism with a patience of 10 epochs was employed to prevent overfitting, and the best weights were restored after training. The dataset was split into 80% training and 20% testing subsets, with a batch size of 64 and a maximum of 300 epochs.

3.3.2 Model architecture and implementation: Gesture Prediction

The predictive model employs an LSTM-based architecture designed to forecast sensor data for the next time step. The model consists of a single LSTM layer with 50 units, followed by a dense layer to generate predictions. Optimized using the Adam optimizer with a learning rate of 1×10^{-5}, the model minimizes Mean Squared Error (MSE) during training. Early stopping and model checkpointing are implemented to prevent overfitting and ensure optimal performance, with the best model weights saved for future use.

The hyperparameters of both models are summarized in Table 1, including architecture, training, and regularization settings selected for optimal performance.

Table 1. Summary of parameters of model structure and training details for task recognition and task prediction.

Aspect	Gesture Recognition	Gesture Prediction
Model type	CNN-LSTM hybrid	LSTM
LSTM Units	100	50

LSTM layers	1 LSTM Layer	1 LSTM Layer
Dropout Rate	0.5	Not Applicable
Optimizer	Adam	Adam
Learning Rate	1e-3	1e-5
Loss Function	Categorial Cross-Entropy	Mean Squared Error (MSE)
Activation Function	ReLU (Dense Layer), SoftMax (Output Layer)	Linear (Output Layer)
Batch Size	64	32
Epochs	Maximum of 300	Maximum of 20000
Early stopping Patience	10	10
Metrics	Accuracy	MSE
Output	Gesture Classification (18 classes)	Predicted Sensor Data

4 Result

The results of this study demonstrate the effectiveness of the proposed deep learning models and dedicatedly designed data structure for gesture recognition and prediction.

4.1 Gesture recognition

The CNN-LSTM model achieved a high accuracy of 98% on the testing dataset, which was derived from a split of 20% (2703 samples) of the entire dataset, in classifying 18 predefined crane signal gestures.

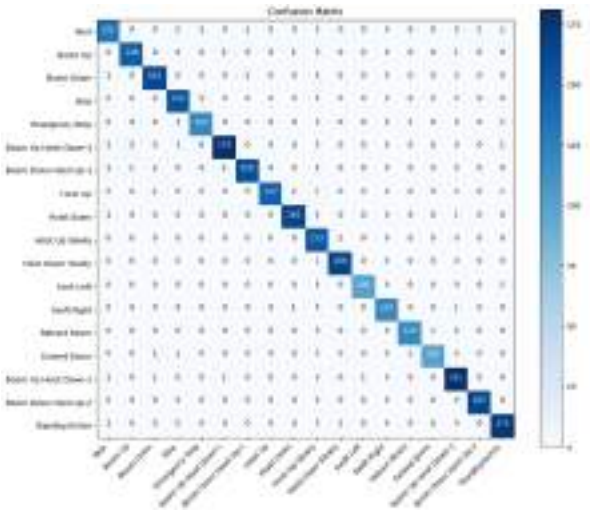


Figure 5. Confusion matrix of CNN-LSTM model for task gesture recognition.

In real situation test, a volunteer was instructed to wear the gloves and perform all 18 predefined gestures. The model successfully detected the correct signals with an accuracy of 95.72% (179 out of 187 samples), demonstrating its strong generalization ability and effectiveness in recognizing both dynamic and static gesture signals. In Figure 5, the confusion matrix provides a comprehensive view of the model's classification performance across all gestures. Figure 6 displays the detailed classification report. In the report, precision measures the proportion of correctly predicted positive observations out of all predicted positive observations, recall (or sensitivity) measures the proportion of correctly predicted positive observations out of all actual positive observations, F1-score is the harmonic mean of precision and recall, and support represents the total number of true instances for each class.

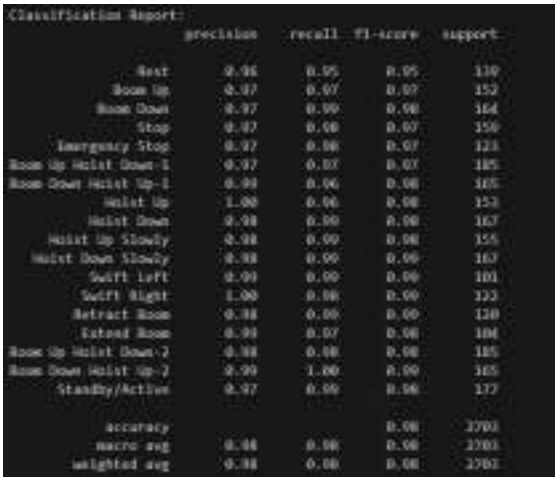


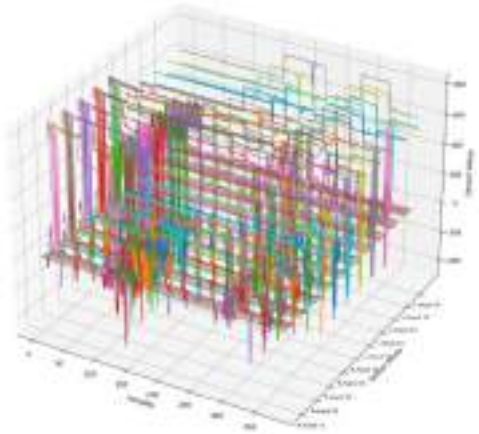
Figure 6. Classification report for gesture recognition

4.2 Gesture prediction

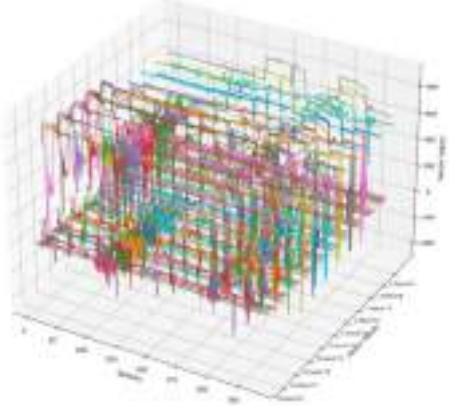
The LSTM model effectively predicted sensor data with an overall R^2 of 0.872, demonstrating reliable accuracy. Here, R^2 measures the proportion of variance in the data explained by the model. With values closer to 1 indicating better predictions. The mean squared error (MSE) quantifies the average squared difference between predicted and true values. The root mean squared error (RMSE) represents the average prediction error in the same unit as the data.

Table 2. Summary of Sensor Prediction Performance

	R^2	MSE	RMSE	Scope
Accelerator	0.924	0.030	0.172	-4.0 - 4.0
Gyroscope	0.789	759.019	27.550	-481.81 - 618.84
Flex sensor	0.903	2557.124	50.568	0 - 843



(a) True sensor values



(b) Predicted sensor values

Figure 7. Comparison between the true values and predicted values

For individual sensors, the accelerator achieved R^2 = 0.924, MSE = 0.030, and RMSE = 0.172 within a range of -4.0 to 4.0, indicating precise predictions. The gyroscope data showed R^2 = 0.789, MSE = 759.019, and RMSE = 27.550 over a broader range of -481.81 to 618.84. The flex sensor achieved R^2 = 0.903, MSE = 2557.124, and RMSE = 50.568, spanning 0.0 to 843.0. The detailed prediction performance indicators are shown in table 2. These results confirm the model's robust performance across diverse sensor types, reliably supporting crane operation safety and decision-making.

From the overall perspective, the prediction results essentially reflect the data characteristics, including the temporal dynamics and distribution patterns of the real data. The comparison between the true value and the predicted value is shown in Figure 7, (a) is the true value, (b) is the predicted value.

The predicted value will be reinput the CNN-LSTM model, the gesture recognition model, to get a predicted signal, so that the operator has an expectation of the next operation and more reaction time to perform correctly.

5 Conclusion and Future work

In the proposed system, the gloves, equipped with multiple sensors (accelerator, gyroscope, and flex sensors), are used as both data collection devices and personal protective equipment (PPE). A dedicated dynamic data structure, combining five consecutive time steps, ensures the preservation of temporal features in signal recognition. This is particularly important for differentiating dynamics signals like "*Boom up*" and "*Boom up hoist down (Option 1)*", where sequential motion is essential for correct classification.

The hybrid CNN-LSTM deep learning model successfully recognizes 18 predefined signals with 98% accuracy on test dataset and 95.72% accuracy in real situation test. Considering the high accuracy, the model performance is effective and robust. In addition, an independent LSTM model is employed for forecasting sensor data for the next time step. The predicted sensor data will be re-input the recognition model to provide crane operators with the oncoming signals. This enables operators to have an expectation for the oncoming instructions, which can help them have longer reaction time and increase the possibility of making correct operations. In turn, operation losses caused by miscommunication and accidents will be minimized.

Future work will focus on continuous data collection and online learning, allowing the system to adapt new data from different users and continuous improvement through real-world usage; enhancing the signal transmission process, which will ensure that correct signals are reliably passed between signalmen and crane operators; improving hardware reliability and standardization, which will prevent equipment damage during use and ensure consistent system performance. Finally, integrating the system with crane control systems, which will enable semi-autonomous or fully automated crane operations, further improving efficiency and safety.

Acknowledgement

This work was supported by China Scholarship Council (CSC), the author sincerely thanks CSC for their financial support.

References

[1] Q. Fang, H. Li, X. Luo, L. Ding, H. Luo, and C. Li. Computer vision aided inspection on falling prevention measures for steeplejacks in an aerial environment. *Autom. Constr.*, 93:148–164, 2018, doi: 10.1016/j.autcon.2018.05.022.

[2] Crane-related work deaths trended down from 1992 to 2017: The Economics Daily: U.S. Bureau of Labor Statistics, Bureau of Labor Statistics.

2024. On-line: https://www.bls.gov/opub/ted/2019/crane-related-work-deaths-trended-down-from-1992-to-2017.htm, Accessed: 17/12/2024.

[3] G. Raviv and A. Shapira. Systematic approach to crane-related near-miss analysis in the construction industry. *Int. J. Constr. Manag.*, 18(4): 310–320, 2018, doi: 10.1080/15623599.2017.1382067.

[4] P. R. Zekavat, S. Moon, and L. E. Bernold. Performance of short and long range wireless communication technologies in construction. *Autom. Constr.*, 47: 50–61, 2014, doi: 10.1016/j.autcon.2014.07.008.

[5] B. Gao, Z. Zhu, J. Zhao, and B. Huang. A Wireless Swing Angle Measurement Scheme Using Attitude Heading Reference System Sensing Units Based on Microelectromechanical Devices. *SENSORS*, 14(12): 22595–22612, 2014, doi: 10.3390/s141222595.

[6] X. Li, G. Gao, J. Na, and X. Chen. Design of an Automatic T-beam Erection System Based on NB-IoT for Bridge-Erecting Cranes. In *PROCEEDINGS OF THE 33RD CHINESE CONTROL AND DECISION CONFERENCE (CCDC 2021)*, in Chinese Control and Decision Conference. New York: IEEE, 2021, 684–689. doi: 10.1109/CCDC52312.2021.9602504.

[7] F. B. Ter Haar, F. Ruis, and B. T. van Manen. A 3D World Interpreter System for Safe Autonomous Crane Operation. *Robotics*, 13(2): 23, 2024, doi: 10.3390/robotics13020023.

[8] N. Singh, V. Gunjan, G. Chaudhary, R. Kaluri, N. Victor, and K. Lakshmanna. IoT enabled HELMET to safeguard the health of mine workers. *Comput. Commun.*, 193: 1–9, 2022, doi: 10.1016/j.comcom.2022.06.032.

[9] A. Mansoor, S. Liu, A. Bouferguene, and M. Al-Hussein. Crane Signalman Hand-Signal Classification Framework Using Sensor-Based Smart Construction Glove and Machine-Learning Algorithms. *J. Constr. Eng. Manag.*, 150(8): 04024094, 2024, doi: 10.1061/JCEMD4.COENG-14458.

[10] L. Wang *et al.* Snake optimizer LSTM-based UWB positioning method for unmanned crane. *PLOS ONE*, 18(11): e0293618, 2023, doi: 10.1371/journal.pone.0293618.

[11] S.-T. Liong, F.-W. Kuo, Y. S. Gan, Y.-T. Sheng, and S.-Y. Wang. Predicting trajectory of crane-lifted load using LSTM network: A comparative study of simulated and real-world scenarios. *Expert Syst. Appl.*, 228: 120215, 2023, doi: 10.1016/j.eswa.2023.120215.

[12] X. Wang and Z. Zhu. Vision-based hand signal recognition in construction: A feasibility study. *Autom. Constr.*, 125: 103625, 2021, doi: 10.1016/j.autcon.2021.103625.

[13] A. A. Neacsu, G. Cioroiu, A. Radoi, and C. Burileanu. Automatic EMG-based Hand Gesture Recognition System using Time-Domain Descriptors and Fully-Connected Neural Networks. In *2019 42nd International Conference on Telecommunications and Signal Processing (TSP)*, Jul. 2019, 232–235. doi: 10.1109/TSP.2019.8768831.

[14] Y. Hu and L. Wang. Hand gesture recognition based on convolutional neural networks. In *LIDAR IMAGING DETECTION AND TARGET RECOGNITION 2017*, D. Lv, Y. Lv, and W. Bao, Eds., Bellingham: Spie-Int Soc Optical Engineering, 2017, UNSP 106051S. doi: 10.1117/12.2291737.

[15] X. Wang, D. Veeramani, F. Dai, and Z. Zhu. Eye Gaze and Hand Gesture-Driven Human-Robot Interaction in Construction. 991–1000, 2024, doi: 10.1061/9780784485262.101.

[16] X. Wang, D. Veeramani, F. Dai, and Z. Zhu. Context-aware hand gesture interaction for human–robot collaboration in construction. *Comput.-Aided Civ. Infrastruct. Eng.*, 39(22): 3489–3504, 2024, doi: 10.1111/mice.13202.

[17] Nano 33 BLE | Arduino Documentation. On-line: https://docs.arduino.cc/hardware/nano-33-ble/, Accessed: 24/09/2024.

[18] sea jump 4pcs 3.7v 150mAH Lipo Battery with 4in1 Charger for H36 E010 Furibee f36 REALACC H36 Drone, Battery Packs - Amazon Canada. On-line: https://www.amazon.ca/jump-Battery-Charger-Eachine-Furibee/dp/B075642WRF, Accessed: 24/12/2024.

[19] H. Junker, P. Lukowicz, and G. Troster. Sampling frequency, signal resolution and the accuracy of wearable context recognition systems. In *Eighth International Symposium on Wearable Computers*, Oct. 2004, 176–177. doi: 10.1109/ISWC.2004.38.

[20] Y. Xiong and F. Quek. Hand Motion Gesture Frequency Properties and Multimodal Discourse Analysis. *Int. J. Comput. Vis.*, 69(3): 353–371, 2006, doi: 10.1007/s11263-006-8112-5.

[21] H. L. Cornman, J. Stenum, and R. T. Roemmich. Video-based quantification of human movement frequency using pose estimation: A pilot study. *PLOS ONE*, 16(12): e0261450, 2021, doi: 10.1371/journal.pone.0261450.

Adaptive Fractal Quadtree Approach for Efficient Drone-Based Construction Site Safety Inspection

Marvin Cheng[1], Ci-Jyun Liang[2], and Hugo Camargo[1]

[1]Division of Safety Research, National Institute for Occupational Safety and Health, USA
[2]Department of Civil Engineering, Stony Brook University, USA
MCheng@cdc.gov, Ci-Jyun.Liang@stonybrook.edu, HCamargo@cdc.gov

Abstract –

This study presents a novel approach for optimizing the flight paths of multiple drones during construction site inspection and monitoring by leveraging image compression techniques, specifically the fractal quadtree method. Traditional path planning methods often fail to account for varying levels of complexity across the construction site, leading to inefficiencies in energy consumption and scanning time. By applying the fractal quadtree approach, we dynamically adapt grid sizes based on the complexity of the initial captured site images, allowing drones to prioritize more intricate areas while minimizing focus on simpler regions. Simulation results demonstrate that this method efficiently reduces the number of required scans compared to conventional approaches, significantly improving the overall efficiency of construction site inspections. The integration of this adaptive path planning strategy provides a scalable and time-saving solution for hazard identification and site monitoring, ensuring more efficient use of drone resources.

Keywords –

Path planning; Safety; Drones; Fractal Algorithm

1 Introduction

Many construction companies are increasingly recognizing the benefits of incorporating drones and unmanned aerial vehicles (UAVs) into their operations [1, 2]. These devices are used to monitor construction progress [3] and inspect structural conditions of civil infrastructures [4] based on 3D reconstruction models. They can also be used to identify hazardous conditions on-site, such as missing personal protective equipment for work at height, and to generate safety warnings for workers [5]. Timely information about nearby hazards is essential for construction workers to ensure their safety. Therefore, drones and UAVs that can quickly and thoroughly scan the entire construction site are crucial [6]. However, deploying a fleet of drones and designing efficient flight paths across construction sites has posed challenges. Existing drone monitor and inspection methods are focused on how to identify the best flight path based on user's experience and 3D reconstruction model quality [3], [4]. Optimizing flight paths is crucial to saving time and energy; for example, Yang et al. [7] proposed an ant colony optimization approach to determine the flight path for digital terrain model reconstruction based on user-defined area, image overlap, and battery limit. In addition, existing methods usually do not consider multiple-drone scenarios in optimization process and assume using a single drone. Traditional flight pattern design methods, such as zigzag or grid-based approaches [8], apply uniform scanning across the entire site and do not account for the varying complexity of different areas. As a result, they can lead to inefficient resource utilization, redundant scans, and even missed critical safety hazards in more complex regions.

In contrast, applying image compression techniques, such as the fractal quadtree method [9], provides a more efficient solution for drone path planning during surface monitoring of construction sites. This approach has been successfully used in adaptive path planning for scanning microscopes [10], [11] robot path planning [12], and trajectories for aerial coverage [13], enabling effective scanning of both complex and simple geometric shapes with different resolutions. By dynamically adjusting the grid sizes based on the visual complexity of different regions, this approach focuses the drones' efforts on areas that require more detailed scanning. Consequently, by dynamically adjusting grid sizes based on the complexity of surface conditions, the proposed fractal quadtree approach (FQA) may allow for better resource allocation, faster coverage, and improved accuracy, as the drones can spend more time in high-risk or complex areas while minimizing redundant scans in simpler zones.

2 Safety Inspection with Drones

To effectively utilize drones to assess safety risks in construction, they must be equipped with high-resolution cameras, GPS for precise localization, and an optimized

angle of view to capture fine details across varying site conditions. The clarity of these images primarily depends on two factors: the resolution of the camera and the drone's Ground Sampling Distance (GSD) above the target area. These variables determine the number of pixels captured per unit length in site imagery, which may vary across different areas of interest. However, high-resolution footage is not always required uniformly across all zones; in areas with minimal construction activity, less detailed scrutiny is sufficient compared to zones with intense building operations.

2.1 Flight Altitude and Resolution of Camera

Drones used for hazard inspection on construction sites should be equipped with two cameras: a forward-facing camera for navigation and a downward-facing camera for capturing site images (Figure 1). These cameras can be combined using a gimbal mechanism to reduce the number of sensors the drone needs to carry.

To examine the relationship among the map resolution of the acquired images, the drone's altitude, or ground sampling distance (GSD), and the resolution of the camera used, a geometric relationship can be applied. As illustrated in Figure 1, the visible range of the camera can be easily derived from this geometric configuration. In this setup, the downward-facing camera's field of view (FOV) angle determines the number of pixels per unit length and area based on its resolution at various GSDs. The drone's visible range at a given GSD is determined by the camera's FOV angle, while the pixel density per unit length is governed by the camera's physical resolution as well as GSD. Given the camera resolution and flying altitude, the visible range of each image can be calculated, allowing the total coverage of the construction site to be determined and divided into a grid. Table 1 provides the visible ranges and image resolutions at specified GSDs.

Table 1 Covered visible ranges and image resolutions on the targeted area at specified GSDs.

Camera resolution			720p	1080p
FOV	GSD (m)	Visible range (m)	Pixels / meter	Pixels / meter
90°	10	20 m	36	54
	20	40 m	18	27
	30	60 m	12	18
	40	80 m	9	13.50
120°	10	34.64 m	20.78	31.18
	20	69.28 m	10.39	15.59
	30	103.92 m	6.93	10.39
	40	138.56 m	5.20	7.79

3 Resolution Adjustment and Flight Path Planning of Swarming Drones

Swarming drones involve multiple drones working collaboratively and autonomously in a coordinated manner to perform tasks. To ensure that drones gather useful information across an entire construction site, their visible range and flight pattern had to be defined. In this study, three approaches were examined. Two of the approaches required the drones to maintain a fixed resolution for the captured images, including: 1) random walk approach [14], and 2) zigzag flight path planning. The third approach used an adjustable resolution based on a fractal quadtree approach (FQA). This section explores each approach and outlines the corresponding flight paths.

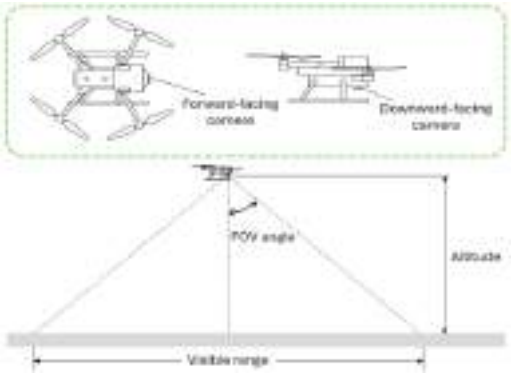


Figure 1. Geometric relationship between FOV angle, altitude, and visible range of the downward-facing camera used by a drone.

3.1 Random Walk Approach

One of the simplest methods for controlling drone behavior is to allow the drones to swarm randomly within a designated map, navigating in various directions across the targeted construction site. The primary advantage of the random walk approach is that the drones only need to communicate for collision avoidance, eliminating the need for coordinated communication or pre-programmed flight paths, which simplifies the overall operation. The images captured by the drones can be stitched together to identify potential hazards. However, this approach can be time-consuming for the drones to cover the entire area, as random flight patterns may result in insufficient coverage and overlap, leading to delays before the full site is scanned. Additionally, since the drones are not coordinated, some areas may receive more attention than necessary, while others might be underexplored. The starting location of the drones, whether random or fixed, does not significantly affect the total time required to complete the scan, but the overall efficiency remains lower compared to more structured flight patterns.

With the random walk theory, the flight pattern can be modeled as a two-dimensional random walk on a grid or a continuous space. Figure 2 demonstrates the possible movements using the random walk method. Each drone makes a series of steps, where each step is determined by

a random direction. We assume that the drone's movement direction follows a uniform distribution in the random walk approach, with random starting positions. Figure 3 illustrates the flight paths of two selected drones (out of a total of 50 drones in the simulation) navigating a 100 × 100 grid area during the process of a full scan.

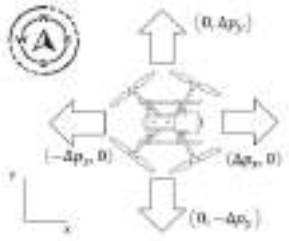


Figure 2. Flight pattern of random walk approach.

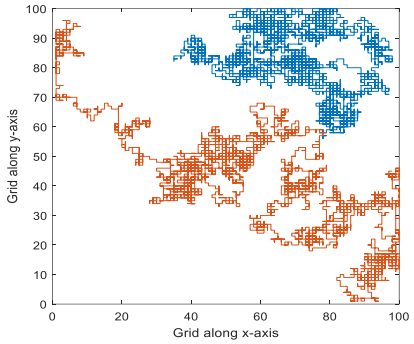


Figure 3. Flight trajectories of two selected drones using random walk approach (50 drones deployed in an area with 100 × 100 cells).

3.2 Zigzag Flight Path Planning

To ensure the image of the targeted construction site can be captured within a predictable timeframe, the zigzag approach can be a versatile option. In this approach, the area is divided proportionally among the total number of drones, allowing the time required to cover the entire site to be easily estimated. By following predefined zigzag paths, the drones systematically sweep through their assigned zones, ensuring full coverage.

In the zigzag approach, the construction site is divided into equal sections for each of the N_D drones, with each drone responsible for scanning a designated portion of the total area, A_S. The total area can be represented as a rectangle or grid of dimensions $L \times W$, where L is the length and W is the width of the site. Thus, the area each drone must cover can be calculated as

$$A_{drone} = \frac{A_S}{N_D} = \frac{L \times W}{N_D}. \quad (1)$$

In this case, the path for each drone is defined by a series of horizontal or vertical zigzag motions across their respective areas.

The path length for each drone depends on the shape of its assigned area. Assuming that the drone moves in a horizontal zigzag pattern, the path length per section for a drone covering an area of $L_{region} \times W_{region}$ is determined by the number of back-and-forth passes required to cover the width of the area. The total length of the path l_{drone} for each drone can be approximated as

$$l_{drone} = \left(L_{region} + 1\right) \cdot \frac{W_{region}}{\Delta p}. \quad (2)$$

where L_{region} is the length of the area each drone covers, W_{region} is the width of the area each drone covers, Δp (Δp_x or Δp_y) is the distance between consecutive passes in the zigzag path. One advantage of the zigzag method is its efficiency. With clearly defined paths, drones can avoid redundant overlap, minimizing wasted time and ensuring that each section of the site is inspected. This structured approach also allows for better time management and straightforward coordination between drones. Moreover, the predictable nature of this method simplifies post-processing, as the images can be stitched together more seamlessly. However, the effectiveness of the zigzag pattern is influenced by the starting positions of the drones. If the drones are not deployed from optimal locations, some may need more time to reach their designated zones, leading to uneven coverage and potential delays. Additionally, the rigid path planning might be less adaptable to irregularly shaped construction sites or areas with obstacles, reducing flexibility compared to more dynamic flight patterns. Overall, the zigzag approach offers a reliable and efficient solution for structured environments, making it a preferred choice when speed and predictability are critical.

3.3 Fractal Quadtree Approach

As noted in the previous section, the visible range of a drone changes with its altitude, affecting the grid sizes used for mapping the targeted area. Image processing techniques using FQA offer a way to dynamically adjust these grid sizes based on the complexity of the construction site surface. Instead of applying a uniform grid across the entire site, this technique adapts the grid resolution according to the visual details captured in lower-resolution images. In areas with lower complexity, such as inactive or open zones, larger grid sizes are used for faster scanning. In contrast, regions with higher complexity, such as those with active construction, are subdivided into smaller grids to capture finer details.

3.3.1 Iterated Function System

In a fractal process, the theory of iterated function systems (IFS) is applied, serving as the foundation for fractal processing techniques [9]. In the application for a construction site, IFS employs affine transformations to capture the relationships between different sections of the image S of the target construction site captured by the downward-facing camera. By using the transformations,

it can define and represent areas with complex structures efficiently. Affine transformations are combinations of rotation, scaling, and translation applied to coordinate axes in *n*-dimensional spaces. A transformation is denoted by q, with $q(S)$ representing the transformed set of points S. The general form of an affine transformation can be expressed as

$$q(p_x, p_y) = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} p_x \\ p_y \end{bmatrix} + \begin{bmatrix} e \\ f \end{bmatrix} = \begin{bmatrix} a \cdot p_x + b \cdot p_y + e \\ c \cdot p_x + d \cdot p_y + f \end{bmatrix}. \quad (3)$$

In this transformation, the coefficients a, b, c, d, e, and f are real numbers that define the parameters. If the translation, scaling, and rotation associated with the mapping q are predetermined, the values of these coefficients are given as

$$a = u\cos\theta,\ b = -v\sin\varphi,\quad c = u\sin\theta,\ d = v\cos\varphi, \quad (4)$$

In this context, u represents the scaling factor along the x-axis, v is the scaling factor along the y-axis, θ is the rotation angle about the x-axis, φ is the rotation angle about the *y*-axis, e and f represent the translations along the x- and y-axis, respectively.

To define an IFS, the concept of contractivity is introduced as following:

Definition 1 (Contractivity) [9] Let S be a metric space with metric f_q. A mapping $q: S \to S$ is said to be Lipschitz with a Lipschitz factor s if there exists a positive real value s such that

$$f_q\big(q(\mathbf{p}_1), q(\mathbf{p}_2)\big) \le s \cdot f_q(\mathbf{p}_1, \mathbf{p}_2) \quad (5)$$

for all $\mathbf{p}_1, \mathbf{p}_2 \in S$. If the Lipschitz factor satisfies $s < 1$, then q is considered contractive with contractivity s.

An IFS consists of a finite collection of contractive affine transformations. This collection of transformations W defines a mapping from the original image S to the transformed image $Q(S)$, which can be expressed as

$$Q(S) = \cup_{i=1}^{n} q_i(S_i). \quad (6)$$

It is important to note that the contractive affine transformation q_i operates on a subset of the original image S, referred to as block S_i. Consequently, $Q(S)$ is formed as a union of these individual contractive transformations. Each transformation q_i is associated with a predetermined property that determines the significance relative to the other transformations.

To establish a systematic method for identifying the affine transformation that generates the IFS encoding of a desired image, the following two theorems are introduced: the collage theorem and contractive-mapping fixed-point theorem.

Theorem 1 (Collage Theorem) [15] Let S be a complete metric space and $q: S \to S$ be a contractive mapping with a contractivity factor s for any pixel $\mathbf{p} \in S$. Then, the following inequality holds:

$$f_q\big(\mathbf{p}, q^n(\mathbf{p})\big) \le \tfrac{1}{1-s} f_q\big(\mathbf{p}, q(\mathbf{p})\big), \quad (7)$$

where $q^n(\cdot)$ represents the n^{th} iteration of the affine transformation.

Theorem 2 (Contractive-Mapping Fixed-Point Theorem) [15] Under the assumption of the collage theorem, there exists a unique point $\mathbf{p}_f \in S$ such that for any point $\mathbf{p} \in S$,

$$\mathbf{p}_f = q\big(\mathbf{p}_f\big) = \lim_{n\to\infty} q^n(\mathbf{p}). \quad (8)$$

This point $\mathbf{p}_f$ is referred to as the fixed point or the attractor of the mapping q.

For a contractive affine mapping, the distance $f_q(\mathbf{p}, q^n(\mathbf{p}))$ between the original points and their mapped counterparts decreases with each iteration. Similarly, for an input image S, it can be demonstrated that the error between the original point and the mapped point remains below a certain bound after n iterations. This is established through repeated application of the triangle inequality and the contractivity of q [9]. If the mapped image $q(S)$ exhibits lower complexity than the original image, then q can be viewed as a lossy encoding of S. We refer to such a transformation w within S that satisfies (5) and (6) as a fractal code for the original image S, indicating that S is self-transformable under q. The process exemplifies a common method for constructing fractal objects.

3.3.2 Partitioned Iterated Function System

In the fractal processing approach used in this study, a key requirement is that the width and length of the target area must be equal, with the side length of the scanned image being a power of 2. Once these conditions are met, the partitioned iteration function system (PIFS) can be applied, which generalizes the standard IFS method. In a PIFS, the original image is divided into numerous non-overlapping square blocks of two different sizes, creating a two-level square partition, commonly referred to as quadtrees. The large blocks of the original image, measuring $L_p \times L_p$, are designated as parent blocks, while the smaller ones, measuring $L_p/2 \times L_p/2$, are termed child blocks. Each parent block can be partitioned into four non-overlapping child blocks that correspond to its four quadrants.

During the transformation process, a segment of the original image S, denoted as S_i^q, is transformed into a corresponding section of the reproduced image, represented as R_i, through the mapping q_i. Here, S_i^q is referred to as a domain block and R_i is called a range block. The collection of domain blocks $\{S_i^q\}_{0\le i<n}$ is termed the domain pool. The decisions to split a parent block occurs during the encoding of individual image blocks. A metric that measures the difference between the original block and the transformed block is essential for determining whether a split is necessary. This process is repeated until the difference between the original and the approximated image is below a predefined threshold.

Based on this recursive approach, given an image S of the construction site and a collection of n contractive

mappings q_1, q_2, …, q_n, the representation of the transformed image of the site can be written as

$$Q(S) = q_1(S) \cup q_2(S) \cup \cdots \cup q_n(S) = \cup_{i=1}^{n} Q_i(S). \quad (9)$$

The partition created in this manner is dependent on the specific image taken from the construction site being processed. This technique enables the encoder to utilize larger blocks in less complex regions of the image while employing smaller blocks to capture finer details in more complex areas.

3.3.3 Fractal Encoding Procedure of FQA

The proposed FQA partitioning algorithm assumes that the images captured from construction sites exhibit fractal characteristics. A key aspect of fractal image processing is minimizing the difference between the original image and its processed result. The goal PIFS encoding process is to identify a set of mappings that reduces this difference between partitioned results and original site conditions with a predefined threshold. The encoding procedure is governed by two main parameters: the tolerable deviation e_t between the original image and the approximated result and the depth (or the finest resolution) r_{min} of the quadtree. The tolerable deviation e_t defines the acceptable deviation between the encoded partitions and the actual image captured at the construction site. The depth of the quadtree determines how much detail is preserved in the transformed partitions, allowing for adaptive grid size based on the visual complexity of different areas.

The coarse image, captured from the construction site, is initially divided into several sections based on a predefined minimum number of grid cells. This quadtree partitioning method represents the image as a hierarchical structure, where each node corresponds to a square section (parent block) of the image, which can be further subdivided into four sub-sections (child blocks). The encoding process begins at the root node, which represents the entire initial image. The relationships among these nodes are structured in a tree-like format, where each node in the branches corresponds to a part of the processed image, or a range block.

After partitioning the image, the range blocks (nodes) are compared with domain blocks from a predefined domain pool $\{S_i^q\}_{0 \le i < n}$. An affine transformation is applied to find the best match that minimizes the difference between the range block and its corresponding domain block. Each domain block, consisting of four non-overlapping square blocks, is four times the size of the range block. To compare them, the pixels in the domain block are averaged in groups of four, reducing the domain block to match the size of the range block.

If the difference between the range block and the domain block exceeds the preset error threshold e_t and the distance between adjacent pixels is larger than the defined minimum resolution r_{min}, the range block is further subdivided into four child blocks, and the process repeats. If the difference is within the acceptable threshold, the affine transformation q_i is stored for that domain block. This collection of all such transformations $Q = \cup q_i$ forms the encoding process of the targeted area, allowing for adaptive grid resolution across different sections of the site based on visual complexity.

Once the area is partitioned into blocks, drones can be deployed accordingly: larger grid blocks correspond to higher flight GSDs, while smaller grids represent lower flight altitudes. Drones are assigned to these grids based on altitude requirements. The blocks, varying in size, are sorted by their proximity to optimize the drones' flight paths. As a result, the flight trajectories at different altitudes can be determined. Taking this into account, the number of drones N_D can be determined and allocated to the appropriate altitudes. Figure 4 illustrates the deployment of drones at different altitudes, tailored to the resolution needed for the partitioned blocks. This approach significantly reduces the number of high-resolution images required, thereby speeding up the overall scanning process for the targeted construction site.

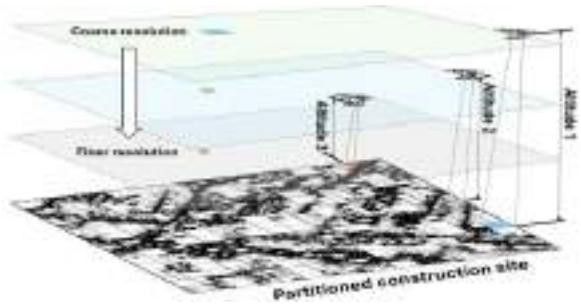


Figure 4. Partitioned construction site and its corresponding drone flight altitudes.

4 Simulation of Different Approaches

To evaluate the performance of the three approaches, a real construction site image covering an area of 64 m × 64 m, as shown in Figure 5, was used to develop a simulated test case. The target resolution was set to a 2 m visible range, based on the downward-facing camera equipped on the drone. This resolution allowed the construction site to be divided into a grid of 32 × 32 cells for both the random walk and zigzag methods.

4.1 Random Walk Approach

In the random walk approach, the number of steps required to fully scan the entire targeted area can vary significantly due to the drone's unpredictable movements. To assess the variability, simulations were conducted using both one drone and two drones. Each scenario was simulated over 200 trials to ensure that the results

exhibited Gaussian distribution. The simulated outcomes are presented in Table 2. The random walk method proved to be highly inefficient. A single drone required over 26,000 steps to cover all 1,024 cells of the 32 × 32 grid at least once. Even with two drones working in tandem, the required steps remained high, exceeding 13,000 to ensure full coverage of the grid. These results demonstrate that, while multiple drones can reduce the total steps, the random walk approach is still suboptimal and time-consuming for large-scale site inspections.

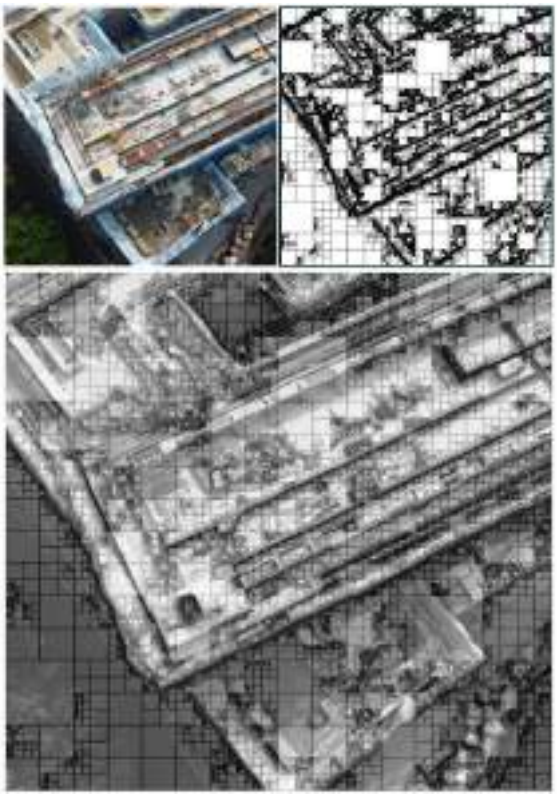


Figure 5. The acquired image of the construction site and the partitioned result (e_t = 10% of the maximum and minimum grayscale values).

4.2 Zigzag Flight Path Planning

Of the simulated approaches, the zigzag approach was the most efficient method for completing a full scan with a fixed resolution. A single drone covered the entire 32 × 32 grid in 1,024 steps, much fewer in comparison to the random walk method. When two drones were deployed, the required steps were halved to 512 for full coverage (with the drones positioned at specific initial locations within the partitioned blocks). This positioning ensures that the zigzag trajectories in each block can be completed systematically and without overlap. Table 3 presents the required steps for areas partitioned into varying grid sizes, highlighting the efficiency and predictability of the zigzag approach compared to the random walk method. This structured deployment helps to ensure optimal utilization of drone resources while maintaining comprehensive coverage.

Table 2 Required steps for a complete scan on the targeted area using random walk method (32× 32 cells).

	Number of drones			
	1		2	
Completed area	Average steps	Standard deviation	Average steps	Standard deviation
50%	2479.6	716.8	1110.1	412.9
70%	4371.8	1066.1	1987.7	678.0
95%	10612	2493	5594.8	1596.7
100%	26502	8901	13585	4479

Table 3 Required steps for a complete scan on the targeted area using zigzag method.

	Number of drones		
	1	2	4
Gridded cells	Required steps	Required steps	Required steps
32 × 32	1024	512	256
64 × 64	4096	2048	1024

4.3 Quadtree Approach

In FQA simulated approach, the number of cells that required drones to visit and acquire images varied, depending on the complexity on the site surface.

4.3.1 Parameters of Fractal Quadtree Approach

To evaluate the performance of the proposed fractal-based path planning approach, images of the College of Liberal Arts Building construction at the National Taiwan University were again used. The area, measuring approximately 64 m × 64 m, was captured, as shown in Figure 5. Before deploying the drones to scan the site and assigning specific flight paths, a preliminary image of the construction site was taken to determine how the area should be partitioned. Once the image was obtained, it was first converted from color to grayscale, a necessary step for FQA. For this test, the construction site was initially divided into 64 blocks, each measuring 8 m × 8 m, to ensure that no portions of the site were scanned without adequate pixel-per-meter resolution. The tolerable deviation e_t was set at 10% of the maximum and minimum grayscale values in the image. Using this threshold, the fractal quadtree approach was applied to generate the partitioned grids for flight scanning, ensuring an optimized balance between resolution and scan efficiency.

4.3.2 Partitioned Blocks Using FQA

The results showed that a simulated construction site could be efficiently partitioned into grids to plan drone flight paths, as demonstrated in Figure 5 and Figure 6.

Initially, the site was divided into very fine grids with a 0.25 m visible range. The original image of the construction site had a resolution of 8 pixels per meter. When partitioned into 8 m × 8 m blocks, the resolution increased to 64 pixels per meter. Doubling the resolution to 128 pixels per meter requires the blocks to be 4 m × 4 m. For this study, the blocks were divided into 2 m × 2 m to achieve the desired resolution of 256 pixels per meter. After merging, the final partitioning resulted in six 8 m × 8 m blocks, 44 blocks sized 4 m × 4 m, and 784 blocks of 2 m × 2 m. This totaled 834 blocks that needed to be scanned and imaged by the drones. In comparison, a traditional zigzag approach for the same area would require scanning 1,024 locations with a single drone, meaning the FQA reduced the total imaging time (no flight time) by approximately 18.75%, requiring only 81.25% of the zigzag approach's locations to be visited.

If the visible range is further refined to 1 m × 1 m blocks, the zigzag approach would require scanning 4,096 locations, while FQA reduces this number to 2,647, or 64.6% of the total grids. At even higher resolutions, 0.5 m × 0.5 m blocks, the FQA covers only 47.91% of the total grids, significantly reducing the imaging effort compared to the zigzag flight path planning method.

Table 4 provides a comparison of the required number of blocks for different visible ranges. Results show that as the resolution becomes finer, FQA substantially reduces the number of locations that need to be scanned, showcasing its efficiency over the zigzag approach. This reduction in the number of required blocks is the key advantage of the FQA, particularly when fine visible range is essential.

Table 4 Comparison of visited blocks and ratios across different tolerable deviations and visible ranges.

e_t	10%			15%		
Visible range	# of blocks	Ratio (%)	PSNR	# of blocks	Ratio (%)	PSNR
2 m × 2 m	832	81.3	24.08	778	76.0	22.84
1 m × 1 m	2647	64.6	25.59	2305	56.3	25.85
0.5 m × 0.5 m	7849	47.9	28.44	6283	38.4	26.04

The quality of the acquired images at different GSDs can be assessed using peak signal-to-noise ratio (PSNR) during the reconstruction of the site image by stitching all partitioned images. To ensure acceptable image fidelity, the PSNR value should typically remain above 25 dB [10]. In this study, the PSNR exceeds 25.59 dB when the compression ratio is kept below 65%, indicating minimal loss of critical visual details.

4.3.3 Synthesis of Flight Paths Using FQA

In FQA, the construction site was partitioned into grids of varying sizes based on the complexity of the acquired images, with smaller grids indicating more detailed areas and larger grids representing fewer complex regions. Once the partitioning was completed, the flight paths for the drones were generated by assigning drones to fly at corresponding altitudes, where the altitude was correlated with the size of the grids. These results indicate that drones flying at higher altitudes can cover larger, less complex areas represented by the bigger grids, while drones flying at lower altitudes can be tasked with scanning smaller, more detailed areas. This adaptive use of altitude ensures that drones are optimally deployed according to the resolution requirements of each specific region within the site.

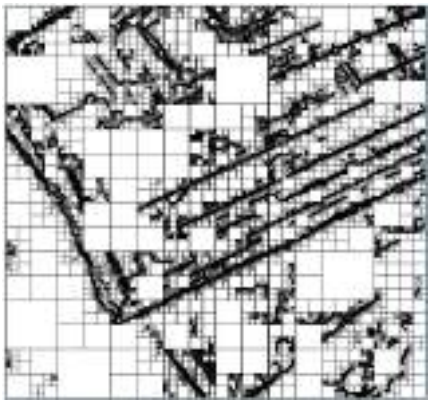


Figure 6. The partitioned result of the construction site ($e_t = 15\%$ of the maximum and minimum grayscale values).

The key advantage of the FQA approach is its efficiency. Drones at higher altitudes can scan large sections of the site in less time, as the lower level of visual detail in these areas requires fewer images. By covering more ground per flight, these drones reduce the total number of paths needed to fully scan the site. Meanwhile, drones operating at lower altitudes can focus on areas that require a higher resolution due to their complexity, ensuring that detailed information is captured where necessary. This distribution of tasks allows for a balanced and scalable operation, where drones can be deployed in the most efficient manner possible based on visual complexity of different zones.

By leveraging the flexibility of flying at different altitudes, FQA may reduce the overall time needed for site scans compared to the two traditional methods. Drones do not need to uniformly fly at low altitudes across the entire site, which can be time-consuming and inefficient. Instead, they can be selectively allocated to specific altitudes, aligning their flight paths with the partitioned grid sizes. This not only reduces the total number of flight paths but also minimizes redundant scanning, allowing for faster coverage of the site without compromising the quality of the images in critical areas. As a result, the overall scan time for the construction site can potentially be reduced, maximizing both the efficiency and precision of the hazard inspection process.

5 Conclusion

The proposed FQA approach offers an efficient and scalable solution for drone-based construction site inspections. By dynamically adjusting grid sizes based on the complexity of acquired images, the FQA approach allows for optimized flight path planning and reduces the time required to scan large areas without sacrificing image quality in critical regions. The adaptive nature of the fractal quadtree technique also enables drones to be deployed at varying GSDs, further enhancing the efficiency of the scanning process by aligning drone altitude with the complexity of the targeted area. This approach not only outperforms traditional approaches such as zigzag and random walk flight patterns but also provides greater flexibility in handling diverse site conditions. Ultimately, the fractal-based grid partitioning method presents a robust and practical solution for improving hazard identification and monitoring on construction sites, facilitating efficient inspections.

6 Acknowledgement and Disclaimer

The authors would like to extend their gratitude to Dr. Jacob J. Lin and Ms. Phuong Linh Le in the National Taiwan University for their valuable support in providing the drone and UAV photographs of the construction sites. The findings and conclusions in this manuscript are those of the authors and do not necessarily represent the official position of NIOSH, CDC. Mention of any company or product does not constitute endorsement by NIOSH, CDC. This activity was reviewed by the CDC, deemed research not involving human subjects, and was conducted consistent with applicable federal law and CDC policy (§See, e.g., 45 C.F.R. part 46; 21 C.F.R. part 56; 42 U.S.C. §241(d), 5 U.S.C. §552a, 44 U.S.C. §3501 et seq).

References

[1] Elghaish F., Matarneh S., Talebi S., Kagioglou M., Hosseini M. R., and Abrishami S. Toward digitalization in the construction industry with immersive and drones technologies: a critical literature review. *Smart and Sustainable Built Environment*, 10(3): 345-363, 2021.

[2] Liang C. J., Le T. H., Ham Y., Mantha B. R., Cheng M. H., and Lin J. J. Ethics of artificial intelligence and robotics in the architecture, engineering, and construction industry. *Automation in Construction*, 162: 105369, 2024.

[3] Lin J. J., Han K. K., and Golparvar-Fard M. A framework for model-driven acquisition and analytics of visual data using UAVs for automated construction progress monitoring. In *Computing in civil engineering 2015*. 156-164, Austin, TX, 2015.

[4] Lin J. J., Ibrahim A., Sarwade S., and Golparvar-Fard M. Bridge inspection with aerial robots: Automating the entire pipeline of visual data capture, 3D mapping, defect detection, analysis, and reporting. *Journal of Computing in Civil Engineering*, 35(2): 04020064, 2021.

[5] Shanti M. Z., Cho C. S., de Soto B. G., Byon Y. J., Yeun C. Y., and Kim T. Y. Real-time monitoring of work-at-height safety hazards in construction sites using drones and deep learning. *Journal of safety research*, 83: 364-370, 2022.

[6] Irizarry J., Gheisari M., and Walker B. N. Usability assessment of drone technology as safety inspection tools. *Journal of Information Technology in Construction*, 17(12): 194-212, 2012.

[7] Yang C. H., Tsai M. H., Kang S. C., and Hung C. Y. UAV path planning method for digital terrain model reconstruction–A debris fan example. *Automation in Construction*, 93:214-230, 2018.

[8] Saeed R. A., Omri M., Abdel-Khalek S., Al E. S., and Alotaibi M. F. Optimal path planning for drones based on swarm intelligence algorithm. *Neural Compu. and Appl.*, 34(12): 10133-10155, 2022.

[9] Fisher Y. *Fractal Image Compression: Theory and Application.* Springer-Verlag, New York, 1995.

[10] Cheng M. H. M., Chiu G. T. C., and Reifenberger R. Fractal compression and adaptive sampling: reducing the image acquisition time in scanning probe microscopy. *Scanning: The Journal of Scanning Microscopies*, 30(6): 463-473, 2008.

[11] Cheng M. H. M. and Bakhoum E. G. Fractal compression and adaptive sampling with HV partitioning: accelerating the scanning process in scanning probe microscopy. In *ASME International Mechanical Engineering Congress and Exposition*, pages 195–201, St. Louis, MO, 2009.

[12] Pinzi M., Galvan S., and Rodriguez y Baena F. The Adaptive Hermite Fractal Tree (AHFT): a novel surgical 3D path planning approach with curvature and heading constraints. *Int. J. of Computer Assisted Radiology and Surgery*, 14: 659-670, 2019.

[13] Sadat S. A., Wawerla J., and Vaughan R. Fractal trajectories for online non-uniform aerial coverage. In *2015 IEEE international conference on robotics and automation (ICRA)*. 2971-2976, Seatle, WA, 2015.

[14] Tai D. and Altschuld J. *Drone Technology in Architecture, Engineering and Construction: A strategic Guide to Unmanned Aerial Vehicle Operation and Implementation*. John Wiley & Sons, New Jersy, 2021.

[15] Jacquin A. E. A novel fractal block-coding technique for digital images. In *IEEE ICASSP Proceeding*, 2225–2228, Albuquerque, NM, 1990.

Demonstration of a robotic system for remote control of the end-hose during concrete pumping

Mathias Haage[1] and Robert Larsson[2]

[1]Department of Computer Science, Lund University, Sweden
[2]Department of Environmental and Building Technology, Lund University, Sweden

mathias.haage@cs.lth.se, robert.larsson@kstr.lth.se

Abstract

Placing concrete using a truck-mounted pump is a common construction method, enabling pouring of large volumes of concrete in a highly efficient way. However, the method is associated with unfavorable working conditions due to the need for manual handling of the end-hose. The hose-man are exposed to hazardous and poor working conditions. This paper reports on efforts to develop a robotic system for wireless remote control of a concrete pump's end-hose. The idea is to replace the need for manual handling of the end-hose. The robotic system is comprised of four pneumatic actuators (so called pneumatic muscles) attached to the pump's end-hose by two light-weight fixtures, and a control system. The proposed system offers improved safety and working conditions for concrete crew members. The robotic device was tested both in laboratory and in real conditions during concrete pumping.

Keywords – Construction robotics, concrete pumping, end-hose, robotic system, pneumatic actuator.

1 Introduction

Concrete is one of the most used materials in construction projects worldwide. The material properties provide important structural functionality in a wide range of applications such as in buildings, bridges, dams, tunnels etc. About 30 billion tons of concrete are used globally each year and the consumption is estimated to increase to about 44 billion tons in 2050 [1].

Due to the heavy weight of concrete (about 2.4 tons per cubic meter) in combination with large volumes normally used in a construction project, there is a great need for an efficient distribution method on the construction site. The idea of transporting concrete using a pump was introduced almost a century ago. Since then, concrete pumps have technically evolved and today, it is an obvious choice of placement method in many construction projects [2]. The pump enables pouring of large volumes of concrete in a shorter time compared to other existing methods such as using a crane-lifted concrete skip [3]. Today, in some countries more than 50% of annual concrete volume is poured using truck-mounted pumps [4].

A truck-mounted concrete pump consists basically of a piston pump, a foldable boom arm, and a flexible end-hose. The boom arm consists of multiple articulated elements whose movements are controlled by hydraulic actuators. The elements are unfolded at the construction site enabling vertical and horizontal reach.

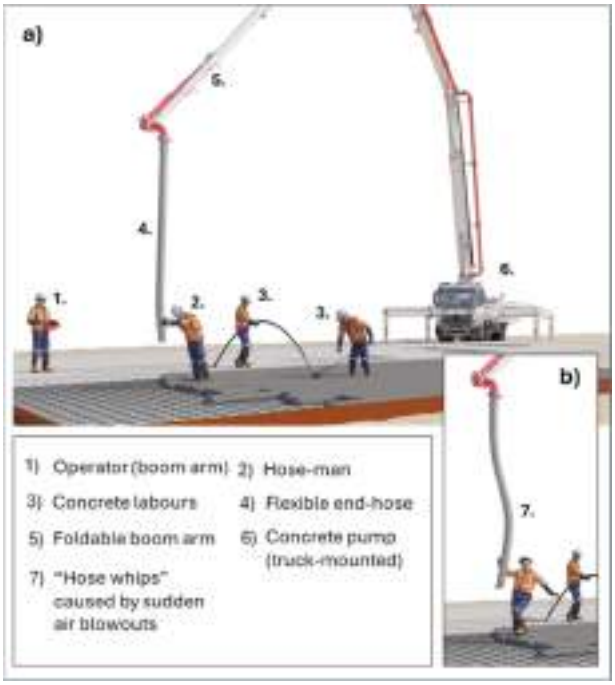


Figure 1a) Truck-mounted pump and concrete crew during placement of concrete. 1b) Sudden blow-outs may cause serious injuries to the hose-man and other workers close to the end-hose.

Concrete is pumped under high pressure (up to 85 bar) through a conveying pipe attached along the boom arm and is sprayed out at the end of the flexible hose [5]. The overall position of the boom arm is controlled remotely by an operator, but the precise position of the end-hose is manually directed by concrete laborer (so called hose-man). The hose-man directs the flexible hose close to the point where concrete is sprayed into the formwork. This is heavy work, but also associated with personal risk due to risk of unexpected buildup of compressed air voids in the conveying pipe systems, resulting in sudden blowouts forcing the hose into uncontrolled movements. This can cause severe injuries to the hose-man or anyone else that are close to the swinging end-hose. To eliminate risk of severe injuries to concrete laborers, this paper describes the efforts to develop a robotic system for remote control of the flexible end-hose. The paper focuses on describing the proposed robotic system and the main findings from a field test under real conditions at a construction site during pumping.

The paper is structured as follows; first a review of previous research is described, then work methodology and the robotic system is presented. Thereafter, results of the tests are provided followed by conclusions and suggestions on future work.

2 State of the art review

2.1 Concrete pumping method

A typical truck-mounted pump and the concrete crew involved during concrete pumping is presented in figure 1 (a and b).

The operator controls the movement of the foldable boom arm using a radio remote control. Another important task for the operator is to control pumping pressure and position of the truck. During concrete pumping, the hose-man handles the flexible end-hose during pumping and the other concrete crew members perform vibration and levelling of the pumped concrete. Handling of the end-hose is a heavy task due to the weight of the flowing concrete, but also hazardous due to the risk of blowouts. Blowouts are caused when air is sucked in the piston pump and compressed in the pipe system. When the compressed air reaches the end-hose it can cause strong and sudden movements of the flexible hose. These “hose whips” can cause serious injuries for laborers who are hit by the end-hose. Typical injuries are broken bones, head injuries, eye damages, but also more severe injuries have been reported resulting in death of the labor worker [6]. The risk of serious accidents increases when pumping is performed at elevated working areas, e.g. pouring slabs in multistorey buildings.

2.2 Review of previous research

During the last two decades, boom arms have been designed to be as light as possible to enable an increased operating reach [7]. However, light-weight boom arms are more vulnerable to vibrations which make them more difficult to control. Accurate control of the boom arm is essential to enable a safe and efficient pumping process, but also to enhance the level of automation of the concrete boom arm. Understanding the dynamic properties of modern boom arms are important. Accordingly, several research attempts have focused on the mechanical and dynamic behavior of the boom arm during concrete pumping.

Zhang et al. used dynamic modeling to study the behavior of a boom arm [8]. The model described the boom arm as a system of several linked sub-structures accounting specifically for the stiffness and dampening of the hydraulic cylinders. The model was compared with a test rig and the results showed that the model could predict displacements and vibrations. Ren et al. studied the displacements and forces subjected to a concrete pump boom arm. Rigid and flexible simulation models were developed and compared with a test rig. It was concluded that a model accounting for flexible behavior is necessary [9]. Schramm et al described a multibody simulation model along with the use of transmission elements to simulate the dynamic behavior of a concrete pump boom arm [10]. It was shown that the model can generate a motion control algorithm that reduce the elastic effects of the boom arm during a typical maneuver. Other researchers have focused on developing a scheme to enable automatic spray of concrete using the boom arm [11]. A simulation model that accounted for the flexible behavior of the boom arm was used to develop the algorithm for automatic control of the end-hose. Simulation results confirmed that the proposed algorithm could control the movement of the boom arm along a narrow pathway. Fan et al reported on their efforts on the design of a new 3-DoF concrete pouring robot along with an autonomous path planning system [12]. The robot was designed to be stationary at a construction site but with a flexible operational reach due to a moveable truss arm. The path planning system included a simplified obstacle algorithm based on reinforced learning. The feasibility of the obstacle avoidance algorithm was demonstrated by simulation and prototype experiments.

Another approach was introduced by Hudelmaier where a mechanical device was attached onto the flexible hose and by using muscle-like actuators (cylinders), the hose could be forced into a desired direction [13]. The cylinders were suggested to be controlled by a handheld remote control eliminating the need for manual handling of the end-hose. The invention was described in a patent document that has expired, and any documentation of realization could not be found.

2.3 Summary of state of the art

Several research attempts have focused on developing control algorithms for precise control and path planning of the boom arm to enable automatic placement of concrete. However, precise control of the boom arm is very difficult to achieve due to the inherent mechanical properties of the boom arm (combination of rigid and flexible parts), vibrations caused by the pulsating flow of concrete, and from varying wind loads acting on a fully stretched boom arm. Moreover, these attempts have not addressed how to control the flexible end-hose where the concrete is sprayed. Therefore, this paper reports on the efforts to develop and test a robotic system for remote control of the end-hose. The system is based on the idea introduced by Hudelmeier but has been updated and modified with recent technologies when it comes to pneumatic actuators and control systems. This paper presents findings of a first-in-kind prototype of a robotic controlled flexible end-hose tested in real environment during concrete pumping.

3 Methodology and setup

The work was initiated by conducting a pre-study, followed by laboratory tests, and finally a field test. The pre-study started with a review of previous research, review of industry reports, and a search of patent databases. A workshop with experts in concrete pumping was also conducted. The purpose of the workshop was to identify functional requirements of a robotic system for controlling the end-hose. Based on findings from the workshop, a design proposal of the robotic system was developed involving experts in robotics and mechanical engineering. After some modifications, a physical prototype was finally manufactured. The system is described in sections 3.1-3.3.

3.1 Description of robotic system

A schematic illustration of the robotic system is outlined in figure 2. The system consists of four pneumatic actuators (so called air muscles) which are attached to a flexible rubber hose by two light-weight fixture plates (top and bottom). The pneumatic cylinders can be manipulated using compressed air which are supplied or released from the air muscles forcing them to change in length. Depending on which cylinders that are filled or released with air, it will force the hose to flex in a certain direction. A valve connected to each cylinder controls the supply and release of air. Compressed air is supplied by the compressor unit of the truck-mounted concrete pump. However, in this version of the robotic system, an external mobile compressor unit is used.

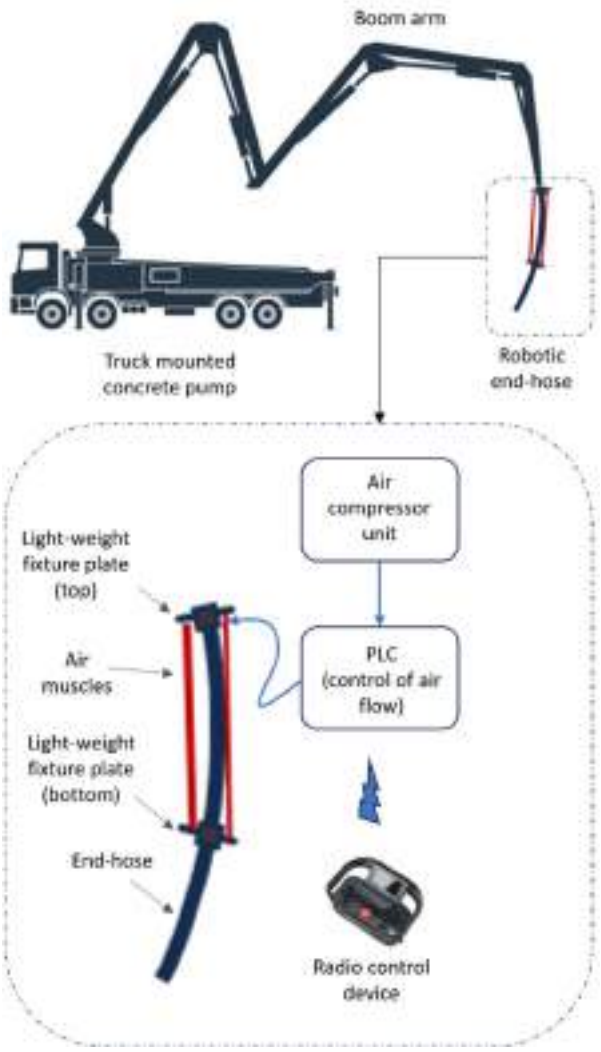


Figure 2. Schematic layout of the robotic system.

A control system comprised of a PLC unit and a wireless radio control device is used to control the movements of the end-hose.

3.2 Mechanical components

The mechanical components of the robotic system are the flexible end-hose, four air muscles (pneumatic actuators), and two light-weight fixture plates. Figure 3 shows a layout of the mechanical components.

The end-hose is a standard concrete hose made of rubber-based material. The hose is available in different standard dimensions, e.g. 3-5 inches in diameter. In this configuration, the end-hose is 4 inches in diameter with a total length of 4 meters.

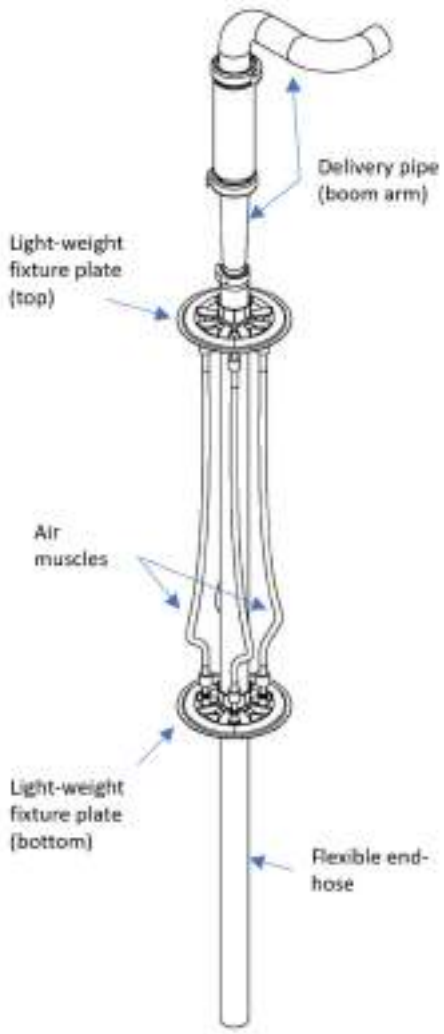


Figure 3. Mechanical components of the robotic end-hose device.

The air muscles are tensile actuators from Festo[1]. Each air muscle consists of a contractible tubing made of a rubber sheath and appropriate connectors. The muscle expands lengthways when it is subjected to an external force, but when pressurized, the muscle contracts, i.e. its length decreases. The air muscles are available in different sizes, e.g. diameter and length.

The air muscles are attached at each end to light-weight fixture plates made of carbon fiber. The upper fixture plate is attached to the end of the boom arm using a standard clamp.

The bottom fixture plate is attached to the flexible end-hose using a standard clamp. This allows for manipulation of the end-hose by controlling the air flow to and from the air muscles.



Figure 4. The wireless joystick control device used for remote control of the end-hose[2].

3.3 Control system

The air muscles are controlled by a programmable logic controller (PLC), as recommended by Festo. The PLC is integrated with programmable pneumatic valves with built-in pressure sensing and equipped with analog and digital I/O for integration with external sensors and control units.

For the field test a wireless industrial joystick was integrated as control unit with the PLC, see Figure 4. Steering programs for muscle control using joysticks were then developed in preparation for the field test.

The joystick device enables control of the end-hose based on two different principles, 1) position-based, or 2) velocity-based. Position-based mode means that when the operator releases the grip of the joystick, the end-hose will return to its neutral position, i.e. straight vertical direction. On the other hand, when the operator releases the grip of the joystick, the end-hose will stay fixed in its current position, e.g. in a position where the end-hose is bent pointing towards a specific direction.

3.4 Laboratory tests

Prior to the field test, prototype development and testing took place in laboratory settings, kindly hosted by facilities at the company Composite Design AB[3] and the Center for Construction Robotics at Lund University[4].

The robotic end-hose prototype was mounted vertically with pneumatic muscles connected to a control PLC through 10 meters of 10 mm air hoses, one for each

[1] www.festo.com
[2] https://akerstroms.se/
[3] https://www.composite-design-sweden.com/
[4] https://robotics.lth.se/research/center-for-construction-robotics/

muscle. This setup enabled systematic experiments where different sub-systems could be tested and evaluated in a controlled environment. The aim was to perform assembly of the prototype and to evaluate the remote control of the pneumatic muscles and the corresponding response of the end-hose. The target was as easy and fast remote handling of the hose as during manual handling, to be evaluated at the field test.

3.5 Field test

A field test was carried out where the robotic system was tested under real conditions, i.e. during concrete pumping. Details about the field test are given in table 1.

Table 1 Details about the field test.

Field test items	Specifications
Construction type and dimensions	Foundation slab (8x12x0.2 m)
Total concrete volume	19 m^3
Concrete type	C35/45, slump range: 100-150 mm
Truck-mounted concrete pump	Putzmeister M36
Boom arm reach	36 m

The aim was to evaluate the performance of the robotic system. More specifically, the evaluation focused on two areas: 1) Remote control, 2) On-site handling and performance of the robotic end-hose. An experienced concrete pump operator from Swerock[5] was used as a test pilot for remote control of the end-hose during concrete pumping. It was considered important to select a test operator with experience of the concrete pumping procedure as well as being used to handle remote controls similar to the one used in this setup. Vibration, levelling, and surface treatment were done by three concrete labor workers. Prior to the start of concrete pumping, the test pilot person was instructed in how to use the remote control. The evaluation was carried out by analysis of video recordings and an interview with the test operator after completion of the pouring. The results were also presented in a workshop after completion of the field test. The workshop involved 9 experts in concrete pumping technology and the aim was to discuss and evaluate the robotic system and the findings from the field test. Special attention was on safety aspects when introducing a robotized pumping method.

[5] www.swerock.se (Swedish ready-mix concrete producer).

4 Results from field test

Here, the findings from the field test are presented. In Figure 5, a schematic illustration of the pouring scheme is shown. The pour units denoted S1-S18 indicate the approximate working area of the robotic end-hose. The boom arm with its end-hose was positioned close to the center of the first pour unit (S1). From here, the end-hose was remotely directed to enable distribution of concrete over the whole area of the pour unit. Thereafter, the boom arm is moved to the next position (center of S2) where the end-hose is used to spray concrete into this area. This procedure was repeated for all successive pour units. In parallel to the concrete placement, labor workers performed tasks such as vibration, level control, and

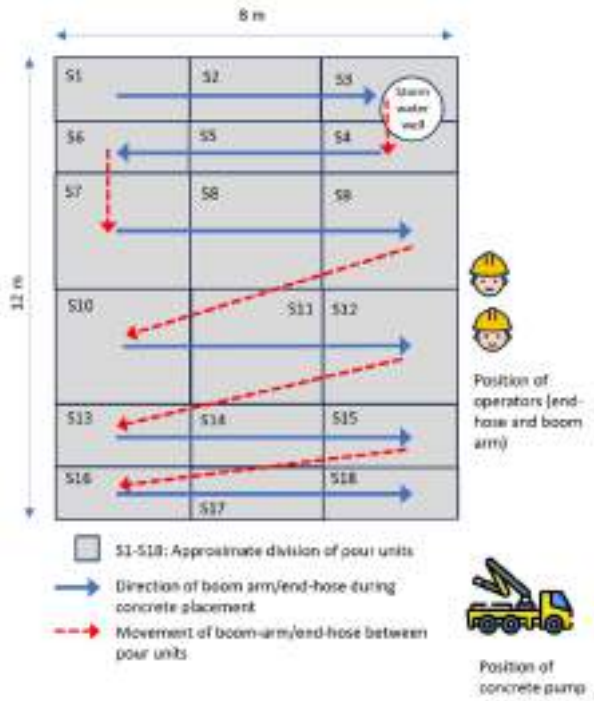


Figure 5. Schematic layout of the pour units and direction of concrete placement.

surface levelling. In Figure 5, the blue arrows indicate the principal direction of the boom arm and end-hose during concrete placement.

The red dotted arrows indicate the movement of the boom arm with its end-hose between concrete pours. Within pour units S3 and S4, a well for storm water is located requiring careful placement of concrete to prevent the well from being filled with concrete.

The truck-mounted concrete pump was positioned near the slab allowing for the boom arm to reach the whole slab without moving the truck.

Table 2. Lead Times (LT) for remote control and manual control of end-hose during concrete placement.

Pour unit	LT Remote placing [sec]	LT Manual placing [sec]	LT Total [sec]
S1	125	0	125
S2	100	0	100
S3	50	85	135
S4	0	67	67
S5	35	18	53
S6	50	0	50
S7	111	0	111
S8	122	5	127
S9	133	0	133
S10	127	3	130
S11	65	0	65
S12	70	0	70
S13	67	0	67
S14	85	5	90
S15	85	0	85
S16	185	0	185
S17	102	3	105
S18	120	0	120

The test operator was standing along the long side of the slab together with the crane operator controlling the movements of the boom arm.

In table 2, lead times (LT) for each pour unit is presented based on video recording analysis. The table presents LT for remote and manual control respectively. Also, total LT is presented. Note that lead time is here defined as the time when concrete flows through the end-hose. Pausing of concrete pumping is not included in lead time.

As seen in table 2, LT:s vary between 50-185 seconds. Note that the sizes of the pour units are not exactly the same, even though it seems to be the case according to Figure 5. Other factors that influence LT are operator performance, need for more accurate control of spraying, interactions with labor workers performing parallel tasks.

The end-hose was remotely controlled for most pour units. In 11 of 18 pour units, concrete was poured using the remote control. Manual control of the end-hose was essentially needed in pour units S3 and S4 due to a more accurate placing of concrete around the storm water well. Also, in pour unit S5, the end-hose was manually handled even though it wasn't necessary. At pour units S8, S10, S14, and S17, only very limited manual handling was necessary to make quick adjustments of the end-hose.



Figure 6. a) manual control of end-hose, b) remote control of end-hose, c) end-hose in bended position, d) overview of robotic end-hose connected to boom arm.

In figure 6, photos from the field test are shown illustrating both manual and remote control of the robotic end-hose (figures 6a and b-c).

The total concrete volume poured was 19 m^3. The time for pouring, vibration, and surface finish was recorded to 60 minutes in total. With four labor workers (three plus one operator), the productivity rate was calculated to be 0.21 man-hours per cubic meter of concrete.

This is equal to the productivity rate used in Sweden for time estimation of manual concrete pumping of foundation slabs [14]. The test operator also confirmed, based on his experience, that the robotic method was equal to the manual method when it comes to concrete placement productivity. This clearly demonstrates that the proposed system provides added value when it comes to risk elimination and improved safety without lowering on-site productivity. In fact, this first test indicates that there is room for significant improvements in both safety and productivity aspects.

5 Discussion

Several aspects in the dimensioning of the robotic equipment turned out to be important. Specific requirements for mounting robotic equipment on an end-hose mounted on the boom arm of a concrete pump were specified by Swerock early on. End-hoses together with the concrete material flowing through them are subject to a tight weight budget. The weight allowance determined by the boom arm used. A typical number for the truck that was used in the field test is 160 kg. A standard 4-inch hose as utilized in our experiments weighs around 8 kg per meter, resulting in 30-35 kg payload for a 4-meter hose. The volume that can be filled with concrete in such a hose is approximately 0.03 m^3 weighing around 90 kg assuming a standard concrete density around 2.400 kg per m^3. Additional equipment, such as adaptors for inner diameter adjustment, limits weight allowance for the robotic equipment to around 20 kg to avoid implications on the hose's applicability.

Another reason for limiting the weight allowance of mounted robotic equipment was risk assessment. Preferably, a robotic end-hose should have a similar assessment as a regular hose to allow for manual handling without the need to dismount the equipment during a pumping session. During the field test it was seen that concrete laborers simply grabbed the hose and pushed it when deemed necessary. To enable ease of manual handling, the lower fixture plate was manufactured in composite material to keep the weight low.

To actuate the hose, four air muscles from Festo were selected. Air muscles are light-weight and therefore fit the weight requirement well. The Festo muscles need to be rather long for effective hose actuation. A typical muscle can contract around 20% of its total length. Some of this length needs to be reserved for slack when the opposite muscle is contracting, leaving an effective actuation of around 10-15% during operation. The muscles selected for the experiments were two meters in length. Considering the actuations of the hose in relation to the allowed 400 mm curvature radius of the hose, the length of muscles was appropriately chosen. Bending the hose has implications for its expected service life, where small bends are preferred over large bends.

An aspect to consider is that the air muscles are pull-only, with return motion dependent on factors such as permitted return air flow and the existence of actuating return forces on the muscle. In this setup, hose bends will act as a return force together with stretched muscle skin pressure. A hampering factor for the evacuation of pressure from a muscle is the length and surface area of the connection between the air muscle and the PLC valve units. During the field test the test operator observed that the equipment was slightly slower in motion than was preferred. Therefore, after the field test, our setup was simulated by Festo using their simulation software. The simulation results show that the tested field setup could get an increased return motion rate by a factor of two by mounting the valves close to the muscles. It is believed that such modification would meet the motion speed range preferred by the operator.

Another determining factor is the availability of air with a flow rate capable of sustaining continuous operation of the air muscles. A typical middle range air pump delivers around 200 liters per minute of 6 bar air, and this also seems to be what is mounted on the pump trucks and available at the tip of the boom arm. The current muscle setup requires more than this amount for sustained operation as noted during the field test. This issue could be solved by reducing the air volume needed by choosing smaller muscles (still strong enough to actuate the hose) or by adding silicon material inside muscles to decrease volume, as suggested by Festo.

During the field test the experienced operator made several observations. The first thing that was commented was alignment of the joystick with the perceived hose actuation directions for intuitive steering. During the field test this was achieved by switching air muscle to valve connections. In the future an alignment functionality in the steering system should be incorporated. There seem to be a learning curve involved in operating the hose remotely, as the operation of the hose became smoother towards the end of the field test. It should be noted though that there was no prior exposure of joystick remote operation to the test operator besides a short introduction right before pouring started. The slight lack of air flow for sustained operation was noted and seen as an inconvenience during the test. It was possible to steer the hose but with less capability as time progressed. When the boom arm was moving the hose between pouring units, the air volume recuperated though. A source of complaint was that the hose behaved differently due to the variation of mass in the hose. This variation was a result of varying amount of concrete flowing in the hose during pumping, e.g. full flow, partial flow, and no flow. The effect was not compensated for in the steering system used in the field test. Of the two

steering modes, the hold function in the velocity mode was preferred.

Overall, the general impression of the remote-control functionality was good. Communication between the boom arm operator, hose operator and concrete crew worked well during the field test, probably due to the experience of all involved. The remote control did not hamper the operation.

Assembly and disassembly of the robotic end-hose onto the boom arm of the concrete pump were done within five minutes by two labor workers. Disassembly and cleaning of end-hose was also done quickly and without any obstacles. In general, the on-site handling of the end-hose was conducted without any significant obstacles.

6 Conclusions and future work

The conclusion from the field test is that the setup worked well, especially considering this was a first of a kind test. There are considerations exposed from the field test, but those seem manageable to handle in a revision of the current setup. The remote steering did not hamper the operation, and the setup did not hamper end-hose handling and manual operation when perceived necessary. Though not performed in formal manner, there were no raised concerns on introduction of risk in the tested scenario.

For future work, we aim to do a revision of the current setup to address considerations exposed during the field test. A longer-term issue is the removal of the remote joystick for the end-hose, so that boom arm and hose operation are controlled through one joystick only.

7 Acknowledgement

This project has been carried out with financial support from SBUF[6], Formas[7] (coordinated by Smart Built Environment), and Heidelberg Materials Cement Sweden[8]. This financial support is kindly acknowledged. A special thanks to Swerock for their valuable support providing test resources and practical know-how.

References

[1] GCCA. Concrete Future - The GCCA 2050 Cement and Concrete Industry Roadmap. On-line: https://gccassociation.org/concretefuture, Accessed: 21/08/2024.

[2] Feys D., De Schutter G., Fataei S., Martys N.S., and Mechtcherine V. Pumping of concrete: Understanding a common placement method with lots of challenges. *Cement and Concrete Research*, 154, 2022.

[3] Zhou, S-W., Guo L-X., and Zhang S-Q. Study on Automatic Spray of Distribution Boom System of Truck-Mounted Concrete Pump, *The Open Mechanical Engineering Journal,* 84-92, 2008.

[4] Swerock. Betongpumpning, Concrete pumping safety manual (In Swedish), pp. 20, 2018, https://swerock.se/siteassets/pdfer/betongpumpning_2018-06_lu.pdf, accessed: 01/10/2024.

[5] Putzmeister, Truck-mounted concrete pump M36, Technical data sheet, pp. 9, https://putzmeister.com, Accessed: 01/10/2024.

[6] Ministry of Business, Innovation and Employment, Concrete Pumping: Health and Safety Guidelines, ISBN 978-0-478-39101-5, pp. 75, 2013.

[7] Henikl, J., Kemmetmüller, W., and Kugi, A. Modeling and Control of a Mobile Concrete Pump, *Proceedings of the International Federation of Automatic Control,* p. 91-98, 2013.

[8] Zhang, Z.W., Wu, Y.X, Liu, J.G., Ren, W., and Cao, M.H. Research on the rigid-flexible multibody dynamics of concrete placing boom, *Automation in Construction,* 67, p. 22-30, 2016.

[9] Ren, W., Li, Z., Bi, Y., Zhao, S., Peng, B., and Zhou, L. Modeling and Analysis of Truck Mounted Concrete Pump Boom by Virtual Prototyping, *J. of Robotics*, Article ID 9182143, pp. 10, 2017.

[10] Schramm, D., Franitza, D., and Lalo, W. Virtual Prototyping and Real-Time Simulation of Heavy Equipment Manipulators including Elastic Deformations and Hydraulics, *Int. Symp. On Automation and Robotics in Construction*, p. 11-14, 2005.

[11] Zhou, S.-W., Guo, L.-X., and Zhang, S.-Q., Study on Automatic Spray of Distribution Boom System of Truck-Mounted Concrete Pump, *The Open Mechanical Engineering Journal,* 2, p. 84-92, 2008.

[12] Fan, S, Li, W., and Yin, R. Research on Path Planning of Concrete Pouring Construction Robot Based on Online Obstacle Avoidance Algorithm, *The 8th Int. Conf. on Advances in Construction Machinery and Vehicle Engineering*, p. 1013-1023, 2024.

[13] Hudelmaier, G. Vorrichtung zum Steuern der Bewegungen eines Betonverteilerschlauches, *German Patent document DE 10 2004 014 949 (In German)*, pp. 12, 2005.

[14] Byggföretagen. Ny- och Ombyggnadslistan 2023, *Productivity rates for construction works* (In Swedish), pp. 224, 2023. https://www.byggnads.se, Accessed:08/10/2024.

[6] Development fund of the Swedish Construction Industry, www.sbuf.se, project number 14308.

[7] www.formas.se, diary number 2023-00260

[8] www.cement.heidelbergmaterials.se

Adaptive Hazard Zone Model for Improved Safety in Highway Work Zones

Ayenew Yihune Demeke[1], Moein Younesi Heravi[1], Israt Sharmin Dola[2], Inbae Jeong[2], and Youjin Jang[1]

[1]Dept. of Civil, Construction and Environmental Engineering, North Dakota State University, Fargo, ND, USA
[2]Dept. of Mechanical Engineering, North Dakota State University, Fargo, ND, USA
ayenew.demeke@ndsu.edu, moein.younesiheravi@ndsu.edu, israt.dola@ndsu.edu, inbae.jeong@ndsu.edu, y.jang@ndsu.edu

Abstract

Highway work zones present hazardous conditions where worker safety is at constant risk due to interactions with construction equipment. In such environments, accurately defining hazard zones around equipment is vital for timely detection and prevention of incidents. Traditional static hazard zone models, often relying on simple geometric shapes, fall short as they cannot adapt to changes in equipment speed and movement, leading to either inadequate safety coverage or frequent false alarms. This study proposes a dynamic hazard zone model that leverages a probabilistic approach based on bivariate Gaussian distribution to generate adaptive and realistic hazard zones. Real-time data from GPS and inertial measurement unit (IMU) sensors enable the model to adjust hazard zone dimensions dynamically in response to equipment behavior, improving both accuracy and responsiveness. Simulations conducted in Webots software demonstrated that the proposed model achieved 93% accuracy in hazard detection, significantly outperforming the 60% accuracy of static circular zones tested in the same environment. The findings underscore the potential of dynamic hazard zone modeling to enhance safety in highway work zones, providing a practical and effective approach for reducing risks.

Keywords –

Highway work zones, construction safety, probabilistic modeling, real-time tracking, dynamic hazard zones

1 Introduction

Highway work zones are inherently complex and hazardous environments where the safety of construction workers and motorists is constantly challenged by active construction and traffic flows. The increased number of accidents and fatalities associated with these zones has led researchers and practitioners to focus on developing more effective safety measures [1], [2]. Accidents in these areas can result from a variety of causes, including but not limited to vehicular collisions, worker-equipment interactions, and improper management of traffic flow [3]. A significant number of these accidents happen due to worker-equipment interactions, making the prevention of near-miss incidents and collisions a critical issue [4].

Effectively mapping the hazard zone around equipment plays a pivotal role in this context, serving as a crucial mechanism for incident detection [5], [6]. Traditional approaches to hazard zone mapping typically employ static models, which use simple geometric shapes, such as circles, to define areas of danger around construction equipment [7]. These models could, however, fall short in adequately providing the necessary protection due to two main reasons. First, these hazard zones have fixed sizes and provide the same hazard area, even when the equipment moves at different speeds [8]. This would lead to either an insufficient zone that could not timely detect hazards for equipment moving at a higher speed, or too large zone, which would lead to unnecessary alarms for equipment moving at a low speed or a stationary piece of equipment [9], [10]. The second reason is related to the shape of the hazard zone. The use of simple shapes like circles will lead to the marking of relatively safer areas, such as the sides of a moving vehicle, as hazardous, which could lead to false alarms [11], [12]. These issues are especially relevant in congested areas such as highway work zones where workers on foot and equipment work side by side in a relatively small area [13]. In such situations, if the hazard zones are not carefully designed to efficiently detect true hazardous situations and avoid any false alarm scenarios, it will lead to either insufficient protection for workers or repetitive false alarms, which could result in work disruption and desensitization [8], [14].

To address these problems, this study proposes an adaptive hazard zone model that dynamically adjusts the

hazard boundaries in real-time based on equipment movement. The proposed model addresses the existing issues by creating a dynamic and more realistic hazard zone around the equipment, effectively mapping the real hazardous areas and avoiding safer areas. The study proposes a rigorous probabilistic approach to developing a mathematical hazard zone model that relies on equipment movement parameters to produce a hazard map with varying levels of safety around the equipment. Real time tracking of equipment movement is achieved by utilizing GPS and inertial measurement unit (IMU) sensors. The proposed model is validated in a simulated environment in Webots software, where several scenarios of realistic interactions between workers and construction equipment are evaluated.

2 Literature Review

Properly modelling hazard zones around construction equipment is essential for preventing near-miss and collision incidents between equipment and workers on foot. These zones must effectively detect and alert potential incidents, allowing both workers and equipment operators sufficient time to react safely. Additionally, the zones should be precise enough to avoid unnecessary alarms that could disrupt work or desensitize workers to genuine hazards. Various hazard zoning approaches exist in practice and in literature, each with its own advantages and limitations. The following sections discuss some of these methods.

2.1 Hazard Zone Models

Hazard zone modelling in construction and industrial environments traditionally relies on static models with fixed sizes, and typically utilizes simple geometric shapes to demarcate areas of danger around equipment. These static hazard zones have predefined danger areas around equipment that do not change shape or size in response to the equipment's real-time movement. Because static zones remain constant regardless of vehicle speed or direction, they can either overestimate or underestimate the true hazardous area in different scenarios.

Among the various static hazard zone approaches, the use of circular hazard zones is common in both practice and literature, with variations in how the circle's size and the number of hazard levels are determined. Wang & Razavi [6] proposed circular hazard and warning zones around a construction equipment. The warning zone represents the area where a warning should be issued when a worker breaches it while the hazard zone represents the actual dangerous area. Although the formulation in the study considered the speed and deceleration of the equipment for determining the braking distance, only typical values for a specific equipment were finally used to determine the dimensions of the hazard zone. Similarly, circular and rectangular hazard zones were explored in studies by Teizer et al. [7] and Luo et al. [15]. Jo et al. [14] also utilized circular hazard zones to represent hazardous areas around an excavator.

In an attempt to make circular zones more representative, Teizer & Cheng [11] proposed a hazard zone that considers only part of a circle, in the direction of the equipment's movement. The proposed hazard zone is also dynamic, adjusting its size based on equipment speed. Although the proposed hazard zone worked generally well, it was pointed out that there were missed hazards and false alarms during the experiment. A study by Shen et al. [5] proposed a standard procedure to come up with hazard zones with different shapes and sizes specific to the type and size of equipment. The procedure also uses typical equipment speed and deceleration values to determine the size of the hazard zone.

Although the studies mentioned above tried to consider several parameters to develop effective hazard zones, there is still a gap in providing a hazard zone model which both has a realistic representation of the actual hazardous area and adapts its size based on real-time equipment movement information. Park et al. [8] affirms this by stating that previous research has insufficiently investigated the variations in proximity sensing caused by different settings such as equipment types and approaching speeds. The study further recommends future research to focus on defining dynamic hazard zones based on factors such as approaching speeds.

Another key aspect to consider is the level of hazard that these zones represent. Many proposed methods provide a single hazard zone, which defines whether a worker is either inside or outside a hazardous area [11], [14]. In contrast, some methods suggest a multi-level hazard zoning approach, which establishes warning and alert areas to offer an additional layer of safety by distinguishing between moderate and critical risk levels [7]. Teizer [16] used three levels of hazard, corresponding to "warn," "slow," and "stop", during field experiment of a proposed magnetic field-based proximity detection and alert system.

Overall, previous literature highlights the need for a hazard zoning methodology where the hazard zone adjusts its size and shape in real-time based on equipment behavior. Such a model would significantly enhance the accuracy of detecting and preventing near-miss incidents. In addition, provision of multiple zones corresponding to varying levels of hazard would be beneficial to give early warnings before a more dangerous situation happens.

Probabilistic approaches are increasingly gaining traction in hazard zoning to improve risk estimation and adaptability in construction and industrial settings,

though there is limited research in applying them for modeling hazard zones around a vehicle. Studies have explored probabilistic risk assessments in construction and industrial environments to enhance safety performance. Park et al. [17] developed a sensor-based safety risk model that quantifies hazard exposure using probabilistic evaluation methods. Zhang et al. [18] applied stochastic processes to model safety risk in construction, demonstrating how probability-driven methods can improve accident prediction. Similarly, Abdrakhmanova et al. [19] proposed a probabilistic hazard analysis framework for industrial operations, incorporating dynamic exposure zones. While these studies demonstrate the benefits of probabilistic approaches in various domains, their application to highway work zones and to model vehicle hazard zones remains largely unexplored.

2.2 Recent Advances in Real-Time Location Systems

Implementing an adaptive hazard zone requires real-time tracking of equipment speed and acceleration. Recent technological advances, particularly in IoT-based applications, have significantly improved the feasibility of real-time data collection in construction environments. Real-time location systems (RTLS) are now frequently deployed in various industries to track the positions of workers, vehicles, and equipment. In construction, these systems could offer substantial benefits in various areas including safety improvements by monitoring proximity and movement in hazardous areas [20], [21].

RTLS employs a range of technologies, including GPS, IMU, radio frequency identification (RFID), Ultra-Wideband (UWB), and Bluetooth, to achieve precise location tracking. These technologies allow for real-time monitoring of assets and workers, enhancing situational awareness in complex, dynamic environments like highway work zones [22], [23], [24]. The use of these sensors allows accurate real-time tracking, which could be crucial for creating adaptive safety zones that adjust based on equipment speed and worker proximity [25].

With real-time data from IoT-based systems, it is now feasible to create hazard zones that expand or contract in response to movement, ensuring more precise safety measures than traditional static models. Such systems have been shown to improve both safety and efficiency on sites where real-time tracking of workers and equipment is crucial, as noted by Teizer et al. [26] in their study on LoRaWAN applications for tracking in construction environments.

3 Methodology

The aim of this study is developing an adaptive hazard zone model for construction equipment with a more realistic hazard zone area using a probabilistic approach. The use of probabilistic method enabled mapping the area around the equipment as a continuous zone with varying levels of hazard risk. The methodology used in this study generally included two main steps: theoretical derivation of a hazard zone model and testing in a simulated environment. In the theoretical derivation, the issues discussed in the previous sections, namely, having a more realistic hazard zone geometry, making it adaptive to equipment speed and direction, and providing zones for varying hazard levels, were addressed by utilizing a probabilistic approach.

3.1 Hazard Zone Model Development

A probabilistic approach with a bivariate Gaussian distribution is used to derive a hazard zone model. This resulted in a 3D graph that not only provides more representative elliptical hazard zones with adaptive major and minor axes but also allows to determine the probability of hazard at any location from the position of the equipment. A series of elliptical projections with the required probability of hazard can be extracted from the graph to represent varying levels of safety around the equipment (Figure 1). The colors of the ellipses in the figure represent an increasing level of hazard in moving towards the equipment, with the red color representing the highest level of hazard and the green one representing the lowest.

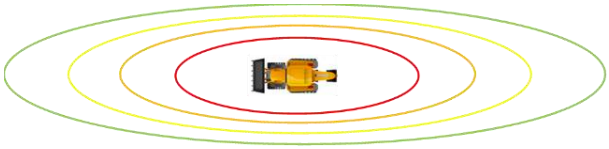


Figure 1. Elliptical projections from the proposed 2D Gaussian graph representing different hazard levels

The equation for a bivariate Gaussian distribution is given in Eq. 1, for independent variables x and y. The expression mainly involves four parameters: the mean and standard deviation values in both directions. The center of the distribution is taken to be at the center of the equipment and in line with this the mean values in both directions are taken to be zero. The standard deviation σ_x controls the spread along the x-axis while σ_y controls the spread along the y-axis. These standard deviations are made to be functions of the equipment's speed.

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y}\exp\left(-\frac{1}{2}\left[\frac{(x-\mu_x)^2}{{\sigma_x}^2}+\frac{(y-\mu_y)^2}{{\sigma_y}^2}\right]\right) \quad (1)$$

This study employs a six-step procedure to draw hazard zones around various construction equipment, as suggested by Shen et al. [5]. These procedures are generally followed to determine the necessary variables

used in the derivation, except for the last step, where a different formulation is used in this study.

1. **Equipment Footprint:** This parameter represents the overall dimensions of the equipment, which depend on the type of equipment. For this derivation, the length (L) and width (W) of the equipment are considered as variables, allowing for flexibility in applying the model to different types of equipment.
2. **Initial Safety Boundary:** The initial safety boundary, represented by S in this study, defines the hazard zone when the equipment is stationary and is also treated as a variable in this derivation.
3. **Equipment Function:** This step involves determining the required safety distance based on the function and type of movement involved in the equipment. For example, the hazard zone of an excavator will have a circular shape with a radius governed by the size of the rotating arm. This study focused on moving equipment such as wheel loaders and dump trucks, with a hazard zone governed by the speed of the equipment. Choosing the type and function of the equipment forms the basis for computing parameters that influence the size of the hazard zone such as the distance required for the equipment to stop once a hazard is perceived by the operator, which is addressed in steps 4 and 5.
4. **Operator Reaction Distance (D_R):** The reaction distance is the distance covered by the equipment between the moment the operator perceives the hazard and takes measure (Eq. 2). It is computed as:
$$D_R = vt \quad (2)$$
Where v is the speed of the equipment in m/s, and t is the reaction time. The reaction time t is recommended to be 2.5 seconds [6], [27]. Thus, the equation simplifies to:
$$D_R = 2.5v \quad (3)$$
5. **Braking Distance (D_B):** The braking distance is the distance the equipment travels between the moment the operator takes action, and the equipment comes to a full stop, as given in Eq. 4 [5].
$$D_B = \frac{v^2}{2a} \quad (4)$$
Where v is the speed of the equipment in m/s, and a is the deceleration of the equipment in m/s². The deceleration could be different depending on the type of equipment. For example, a deceleration value of 1.9 m/s²is taken in Wang & Razavi [6] for forklift. In this study, it is considered as a variable, so that appropriate value could be assigned for a specific type of equipment later.
6. **Determine Hazard Zone:** The final step involves determining the resulting hazard zone based on steps 1 through 5. As mentioned above, the hazard zone is taken to be elliptical in shape, following the projection of a bivariate Gaussian distribution with the major axis representing the direction of movement of the equipment. The lengths of the major and minor axis of the ellipse are computed based on the values in the previous steps as described in Eq. 5 to Eq. 7.

Major axis:
$$Major\ axis = L + 2S + 2D_R + 2D_B \quad (5)$$
Where L is the length of the equipment, S is the initial safety boundary, D_R is the reaction distance, and D_B is the breaking distance. Therefore,
Major axis:
$$Major\ axis = L + 2S + 5v + \frac{v^2}{a} \quad (6)$$
Minor axis:
$$Minor\ axis = W + 2S \quad (7)$$

To determine σx and σy, these values are equated with the equation of an ellipse for a typical projection of the proposed bivariate Gaussian probabilistic plot, taking the hazard zone to be the ellipse with a 70% probability of hazard. This could be done for different probability values to derive different safety boundaries for different levels of hazard.

For a bivariate Gaussian distribution, the equation for an ellipse representing a specific probability contour is given in Eq. 8.
$$\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 = k \quad (8)$$
Where k is a constant corresponding to a specific confidence level. For a 70% confidence level, the value is obtained from the chi-squared distribution table to be approximately 0.713.

For a bivariate Gaussian distribution, the major and minor axes of the ellipse can be expressed in terms of the standard deviations as in Eq. 9 and Eq. 10.
Major axis:
$$Major\ axis = 2\sigma_x\sqrt{k} \quad (9)$$
Minor axis:
$$Minor\ axis = 2\sigma_y\sqrt{k} \quad (10)$$

Equating Eq. (9) and (10) with the corresponding expressions obtained earlier for the major and minor axes will result in an equation for the standard deviations in terms of equipment parameters.
Major axis:
$$2\sigma_x\sqrt{k} = L + 2S + 5v + \frac{v^2}{a} \quad (11)$$
$$\sigma_x = C_1L + C_2S + C_3v + C_4\frac{v^2}{a} \quad (12)$$
Minor axis:
$$2\sigma_x\sqrt{k} = W + 2S \quad (13)$$

$$\sigma_y = C_5 W + C_6 S \quad (14)$$

Where C_1 = 0.592, C_2 = C_6 = 1.183, C_3 = 2.959, and C_4 = C_5 = 0.592 are the coefficients.

The final expressions for the standard deviations show that the size of the hazard zone along the major axis of the ellipse depends on the real-time speed and the deceleration rate of the equipment, while the size in the cross-sectional direction remains constant. If the equipment turns, the major axis will change direction accordingly keeping the hazard zone aligned with the equipment's direction of movement.

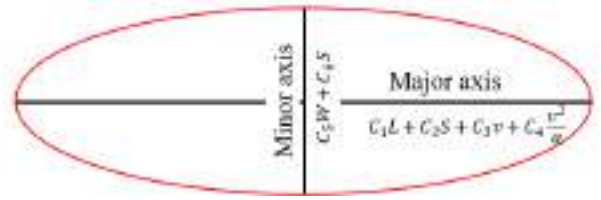


Figure 2. Hazard zone dimensions

In practical applications, real-time GPS/IMU sensor data will be used to continuously track the equipment's location, ensuring that the center of the hazard zone remains properly aligned with its current position. Additionally, these sensors will provide velocity and acceleration measurements, enabling the dynamic adjustment of the hazard zone size using Equations (12) and (14). To enhance accuracy and mitigate sensor noise, efficient sensor fusion techniques should be employed to effectively integrate GPS and IMU readings.

3.2 Simulation Setup

The primary objective of the simulation was to evaluate the effectiveness of the proposed hazard zone model in correctly detecting near-miss incidents at the right time. The simulation was conducted in Webots R2023b, an open-source desktop application from Cyberbotics, primarily used for robot simulations. The software was chosen for its strengths in effectively creating realistic road and traffic environments [28].

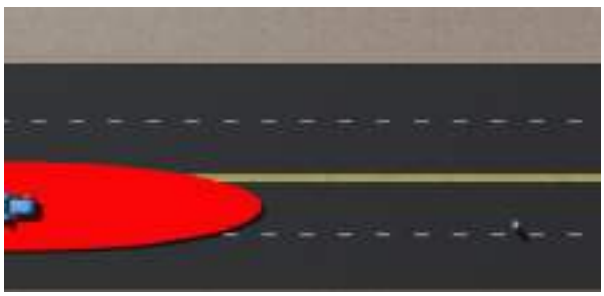
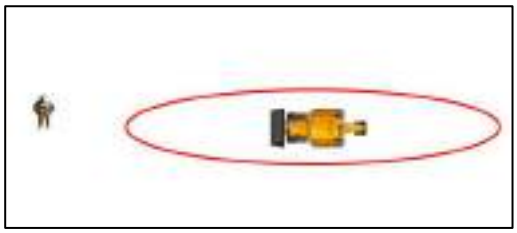


Figure 3. Simulation setup in Webots

The strategy involved driving equipment, equipped with GPS and IMU sensors, towards a worker positioned in front of it. A hazard zone was generated around the equipment, and when the worker entered the hazardous area, the equipment began decelerating, mimicking an operator's response upon being alerted (Figure 3). The equipment continued at its current speed for 2.5 seconds to simulate operator reaction time, then decelerated at the typical rate for the equipment in the simulation environment. The stopping position was evaluated to determine whether the hazard zone provided sufficient time for the equipment to stop. The simulation was conducted using the proposed dynamic hazard zone, and a static circular hazard zone for comparison (Figure 4).



(a)



(b)

Figure 4. Model evaluation strategy (a) Proposed hazard zone (b) Static circular hazard zone

A total of 100 simulations were conducted for both hazard zones, with varying worker locations and equipment speeds. The worker's location was randomly varied both in the longitudinal and cross-sectional directions. To keep the location within a reasonable range, the location is kept within a 30 m range across the road, i.e. 15 m from the centerline of the equipment in both directions, and 100 m range along the road. The equipment speed was sampled from a normal distribution with a mean of 10 kph and a standard deviation of 5 kph. To ensure realistic speeds for construction equipment, the minimum and maximum speed values were set at 5 kph and 20 kph, respectively.

The proposed model has variables including equipment footprint, initial safety boundary, the speed of equipment, and deceleration rate of the equipment. These variables were set before running the simulation, except

for the speed of equipment which is variable. The equipment footprint was directly measured for the specific equipment used in the simulation to be 3 meters long and 1.7 meters wide. The initial safety boundary is taken to be 2 meters as is recommended by Shen et al. [5] and Wang et al. [6]. The deceleration rate, which is the rate at which the equipment decreases its speed when the brake is applied, was obtained to be 1.05 m/s^2 for the specific equipment used in the simulation.

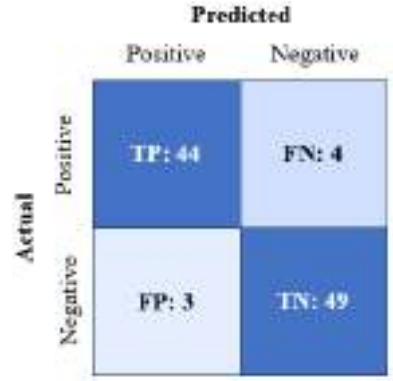
3.3 Evaluation Criteria

After completing the necessary setup, the simulation was then run, and results were categorized into true positives, true negatives, false positives, and false negatives, with the following meanings:

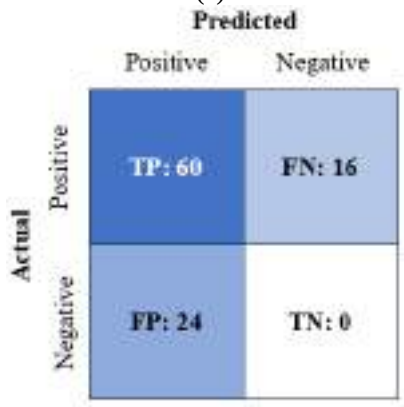
- **False Negative (FN):** The equipment hits or dangerously approaches the worker, indicating a failure to alert in time. This indicates that the model couldn't timely recognize a positive case and gave a false safety, leading to a dangerous situation. A trial is classified as false negative if the equipment hits the worker or the worker comes within a 1-meter distance around the equipment footprint at any time during the equipment's movement or when the equipment finally stops.
- **False Positive (FP):** The equipment stops too far from the worker or passes safely, yet the system triggers an unnecessary alert. Previous literature provides little guidance on the appropriate distance between the worker and equipment to classify it as "too far". In a field experiment by Teizer et al. [7], equipment stopping at least 10 meters away from the worker was considered safe. This value is adopted as the criterion in this study to differentiate false positives from true positives, along the direction of equipment movement. However, applying the 10-meter threshold to the width of the equipment is deemed unreasonable, as it would be overly conservative and result in many false alarms being classified as true positives. Therefore, the width threshold is set at one standard lane width, i.e., 3.6 meters, from the equipment. Accordingly, a false positive occurs when the equipment stops while the worker is outside the 10 meters by 3.6 meters boundary.
- **True Positive (TP):** The equipment stops safely with the worker at a safe distance, indicating that the system correctly detected the danger.
- **True Negative (TN):** The worker is located at a safe distance from the equipment and no alert is issued, as expected.

4 Results and Discussion

Following the methodology described, a total of 100 simulations were conducted for both the proposed dynamic hazard zone model and a conventional static circular hazard zone. The simulations accounted for varying worker locations and equipment speeds, ensuring diverse scenarios for evaluation. The simulation results showed that the proposed hazard zone model classified 93% of trials correctly while the static circular hazard zone correctly classified 60% of trials. The total number of cases classified as true positives, true negatives, false positives, and false negatives are presented in Figure 5.



(a)



(b)

Figure 5. Confusion matrix (a) Proposed hazard zone (b) Static circular hazard zone

It can be clearly seen that the circular hazard zone struggled to identify negative cases and classified them as false positives. In addition, the static circular hazard zone was less efficient in detecting true positive cases, as it classified more cases as false negatives than the proposed hazard zone. This occurred because the static circular hazard zone used a fixed radius for varying equipment speed values, leading to false negatives at higher speeds and false positives at lower speeds. To further understand the performance of the proposed dynamic hazard zone model, the results were analyzed based on different speed ranges of construction equipment. The findings are summarized in Table 1.

Table 1. Breakdown of model performance across

different speed ranges

Speed Range	Hazard Zone	TP	TN	FP	FN	ACC
5–10 kph	Proposed	20	22	1	1	95.5%
	Static	37	0	12	0	75.5%
10–15 kph	Proposed	19	20	1	2	92.9%
	Static	17	0	9	12	44.7%
15–20 kph	Proposed	5	7	1	1	85.7%
	Static	6	0	3	4	46.2%

The results highlight the following key points:

- **Low Speeds (5–10 kph):** The proposed model demonstrated high accuracy (95.5%) with minimal false positives (1) and false negatives (1), effectively distinguishing between hazardous and safe scenarios. In comparison, the static model achieved lower accuracy (75.5%) due to a higher number of false positives (12).
- **Medium Speeds (11–15 kph):** At medium speeds, the proposed model maintained strong performance with an accuracy of 92.9%. The static model, however, exhibited significant performance degradation, with an accuracy of only 44.7% due to large number of false positives (9) and false negatives (12). This decline in performance is primarily because, despite the increase in speed, the static hazard zone remains unchanged, treating equipment moving at higher speeds the same as slower-moving equipment. As a result, the static model fails to account for the increased stopping distance and reaction time, leading to false negatives, where a real hazard is incorrectly classified as non-hazardous. This limitation highlights the advantage of the proposed model, which dynamically adjusts the hazard zone based on speed variations, ensuring more accurate hazard identification.
- **High Speeds (16–20 kph):** The proposed model achieved an accuracy of 85.7%, with a slight decrease compared to lower speeds. This indicates that the model still remains effective even at higher speeds. In contrast, the static model struggled significantly, achieving only 46.2% accuracy. These results reinforce the adaptability of the proposed model compared to static hazard zones.

While the simulation results demonstrate the proposed model's effectiveness, its validation in a controlled virtual environment limits generalizability to real-world conditions, where factors like unpredictable worker behavior, sensor inaccuracies, and environmental variables may influence performance. To address these challenges, field testing is essential to assess the model's accuracy and robustness under diverse, dynamic scenarios involving varying equipment types and operational conditions.

5 Limitations

While the proposed probabilistic hazard zoning model performs well in the simulated environment used in this study, several real-world challenges may affect its practical implementation. Real-world sensors are subject to noise and drift, which can introduce inaccuracies in position tracking and cause deviations in hazard zone boundaries. To mitigate this, effective sensor fusion techniques should be employed to enhance location reliability. Additionally, in highly dynamic environments such as highway work zones, unpredictable worker behavior and complex interactions between workers and construction equipment present further challenges. Adapting the proposed model for such scenarios will require future work focused on incorporating predictive techniques for worker movement patterns and worker-equipment interactions. Field validation through real-world deployments will also be essential to refine the model and optimize its parameters for practical applications.

6 Conclusion

This study developed a dynamic hazard zone model to address the limitations of traditional static circular hazard zones in highway work zones. By employing a probabilistic approach using a bivariate Gaussian distribution, the proposed model generates adaptive hazard zones that account for real-time changes in equipment speed and movement. The proposed model's effectiveness was validated through simulations, where it demonstrated superior performance compared to static circular hazard zones, achieving 93% accuracy in correctly classifying hazard scenarios. The findings highlight the significant potential of dynamic hazard zones in enhancing worker safety by accurately detecting true hazards and minimizing false alarms, even in highly dynamic environments. This approach not only reduces the risk of near-miss incidents and collisions but also minimizes disruptions caused by unnecessary alerts. Future work could focus on validation through field experiments and refinement of the model for more accuracy.

Acknowledgements

The authors would like to acknowledge the funding (Grant No. 1036338) provided by the Minnesota Department of Transportation (MNDOT) to conduct this study.

References

[1] M. Osman, R. Paleti, and S. Mishra, "Analysis of passenger-car crash injury severity in different work zone configurations," *Accid Anal Prev*, vol. 111, pp. 161–172, Feb. 2018, doi: 10.1016/J.AAP.2017.11.026.

[2] N. S. S. Al-Bdairi, "Does time of day matter at highway work zone crashes?," *J Safety Res*, vol. 73, pp. 47–56, Jun. 2020, doi: 10.1016/J.JSR.2020.02.013.

[3] J. M. Wong, M. C. Arico, and B. Ravani, "Factors influencing injury severity to highway workers in work zone intrusion accidents," *Traffic Inj Prev*, vol. 12, no. 1, pp. 31–38, Feb. 2011, doi: 10.1080/15389588.2010.525569.

[4] K. Elelu, T. Le, and C. Le, "Collision Hazard Detection for Construction Worker Safety Using Audio Surveillance," *J Constr Eng Manag*, vol. 149, no. 1, p. 04022159, Nov. 2022, doi: 10.1061/JCEMD4.COENG-12561.

[5] X. Shen, E. Marks, N. Pradhananga, and T. Cheng, "Hazardous Proximity Zone Design for Heavy Construction Excavation Equipment," *J Constr Eng Manag*, vol. 142, no. 6, p. 05016001, Jan. 2016, doi: 10.1061/(ASCE)CO.1943-7862.0001108.

[6] J. Wang and S. N. Razavi, "Low False Alarm Rate Model for Unsafe-Proximity Detection in Construction," *Journal of Computing in Civil Engineering*, vol. 30, no. 2, Mar. 2016, doi: 10.1061/(asce)cp.1943-5487.0000470.

[7] J. Teizer, B. S. Allread, C. E. Fullerton, and J. Hinze, "Autonomous pro-active real-time construction worker and equipment operator proximity safety alert system," *Autom Constr*, vol. 19, no. 5, pp. 630–640, Aug. 2010, doi: 10.1016/j.autcon.2010.02.009.

[8] J. W. Park, X. Yang, Y. K. Cho, and J. Seo, "Improving dynamic proximity sensing and processing for smart work-zone safety," *Autom Constr*, vol. 84, pp. 111–120, Dec. 2017, doi: 10.1016/J.AUTCON.2017.08.025.

[9] I. Abotaleb, K. Nassar, and O. Hosny, "Layout optimization of construction site facilities with dynamic freeform geometric representations," *Autom Constr*, vol. 66, pp. 15–28, Jun. 2016, doi: 10.1016/J.AUTCON.2016.02.007.

[10] J. Wang and S. Razavi, "Two 4D Models Effective in Reducing False Alarms for Struck-by-Equipment Hazard Prevention," *Journal of Computing in Civil Engineering*, vol. 30, no. 6, p. 04016031, Nov. 2016, doi: 10.1061/(ASCE)CP.1943-5487.0000589/ASSET/194B8CA0-62E0-4FA5-99AB-0273D5464CAF/ASSETS/IMAGES/LARGE/FIGURE12.JPG.

[11] J. Teizer and T. Cheng, "Proximity hazard indicator for workers-on-foot near miss interactions with construction equipment and geo-referenced hazard areas," *Autom Constr*, vol. 60, pp. 58–73, Dec. 2015, doi: 10.1016/j.autcon.2015.09.003.

[12] I. G. Awolusi, E. D. Marks, N. Pradhananga, and T. Cheng, "Hazardous proximity zone design for heavy construction equipment," 2015, doi: 10.14288/1.0076403.

[13] I. Awolusi and E. D. Marks, "Active work zone safety: Preventing accidents using intrusion sensing technologies," *Front Built Environ*, vol. 5, p. 431578, Mar. 2019, doi: 10.3389/FBUIL.2019.00021/BIBTEX.

[14] B. W. Jo, Y. S. Lee, J. H. Kim, D. K. Kim, and P. H. Choi, "Proximity Warning and Excavator Control System for Prevention of Collision Accidents," *Sustainability 2017, Vol. 9, Page 1488*, vol. 9, no. 8, p. 1488, Aug. 2017, doi: 10.3390/SU9081488.

[15] X. Luo, H. Li, T. Huang, and T. Rose, "A field experiment of workers' responses to proximity warnings of static safety hazards on construction sites," *Saf Sci*, vol. 84, pp. 216–224, Apr. 2016, doi: 10.1016/J.SSCI.2015.12.026.

[16] J. Teizer, "Magnetic Field Proximity Detection and Alert Technology for Safe Heavy Construction Equipment Operation," 2015.

[17] J. Park, Y. K. Cho, and A. Khodabandelu, "Sensor-Based Safety Performance Assessment of Individual Construction Workers," *Sensors 2018, Vol. 18, Page 3897*, vol. 18, no. 11, p. 3897, Nov. 2018, doi: 10.3390/S18113897.

[18] Z. Zhang, W. Li, and J. Yang, "Analysis of stochastic process to model safety risk in construction industry," *Journal of Civil Engineering and Management*, vol. 27, no. 2, pp. 87–99, Feb. 2021, doi: 10.3846/JCEM.2021.14108.

[19] K. N. Abdrakhmanova, A. V. Fedosov, K. R. Idrisova, N. K. Abdrakhmanov, and R. R. Valeeva, "Review of modern software complexes and digital twin concept for forecasting emergency situations in oil and gas industry," *IOP Conf Ser Mater Sci Eng*, vol. 862, no. 3, p. 032078, May 2020, doi: 10.1088/1757-899X/862/3/032078.

[20] W. Umer and M. K. Siddiqui, "Use of Ultra Wide Band Real-Time Location System on Construction Jobsites: Feasibility Study and Deployment Alternatives," *Int J Environ Res*

Public Health, vol. 17, no. 7, Apr. 2020, doi: 10.3390/IJERPH17072219.

[21] A. Y. Demeke *et al.*, "Advancing Highway Work Zone Safety: A Comprehensive Review of Sensor Technologies for Intrusion and Proximity Hazards," *https://doi.org/10.1177/03611981241265845*, Aug. 2024, doi: 10.1177/03611981241265845.

[22] N. Soltanmohammadlou, S. Sadeghi, C. K. H. Hon, and F. Mokhtarpour-Khanghah, "Real-time locating systems and safety in construction sites: A literature review," *Saf Sci*, vol. 117, pp. 229–242, Aug. 2019, doi: 10.1016/J.SSCI.2019.04.025.

[23] M. Younesi Heravi, I. S. Dola, Y. Jang, and I. Jeong, "Edge AI-Enabled Road Fixture Monitoring System," *Buildings 2024, Vol. 14, Page 1220*, vol. 14, no. 5, p. 1220, Apr. 2024, doi: 10.3390/BUILDINGS14051220.

[24] I. Jeong, Y. Jang, I. Sharmin Dola, and M. Younesi Heravi, "A Framework for Remote Road Furniture Monitoring System Using Smart IoT Dashcams and Digital Twin," *Computing in Civil Engineering*, pp. 1080–1088, 2024, doi: 10.1061/9780784485248.129.

[25] H. Kim and S. Han, "Accuracy Improvement of Real-Time Location Tracking for Construction Workers," *Sustainability 2018, Vol. 10, Page 1488*, vol. 10, no. 5, p. 1488, May 2018, doi: 10.3390/SU10051488.

[26] J. Teizer *et al.*, "Construction resource efficiency improvement by Long Range Wide Area Network tracking and monitoring," *Autom Constr*, vol. 116, p. 103245, Aug. 2020, doi: 10.1016/J.AUTCON.2020.103245.

[27] MUTCD, "Chapter 2C - MUTCD 2009 Edition - FHWA." Accessed: Jun. 03, 2024. [Online]. Available: https://mutcd.fhwa.dot.gov/htm/2009/part2/part2c.htm

[28] "Cyberbotics: Robotics simulation with Webots." Accessed: Oct. 21, 2024. [Online]. Available: https://cyberbotics.com/#webots

Advanced Sensor Integration for Enhanced Flight Control in UAV-Based Construction Automation

Tianyu Ren[1], Xiayu Zhao[1] and Houtan Jebelli[1]

[1] Department of Civil and Environmental Engineering, University of Illinois Urbana-Champaign, USA
tianyur2@illinois.edu, xiayuz3@illinois.edu, hjebelli@illinois.edu

Abstract

In construction automation, several tasks such as surface finishing, material transport, and inspections typically rely on labor-intensive manual methods, which are inefficient, susceptible to human error, and expose workers to hazardous conditions. Drones have emerged as a solution to automate these processes, but maintaining precise navigation and stable attitude control in dynamic construction environments remains challenging. This paper presents a multi-sensor fusion and control framework to enhance UAV performance in dynamic and complex construction environments. The framework integrates multi-dimensional data from diverse sensing modalities, including geometric, visual, and inertial information, to achieve robust situational awareness and adaptive control. A Simultaneous Localization and Mapping (SLAM) model fuses multi-modal sensor inputs, including LiDAR point clouds, camera visual features, and inertial measurements, to generate precise environmental maps and perform accurate localization. The SLAM framework employs advanced loop closure techniques to minimize trajectory drift and optimize navigation paths in real-time. Additionally, flight stability under external disturbances is managed by a Deep Reinforcement Learning (DRL) algorithm, which models the UAV's control system as a Markov Decision Process (MDP). The DRL agent dynamically adjusts the UAV's flight parameters, such as thrust, roll, pitch, and yaw, based on real-time feedback from fused sensor data. The control strategy leverages a stability-based reward function to minimize deviations in position and orientation, ensuring rapid recovery from external disturbances. Simulation tests conducted in a Robotic Operation System and Gazebo simulation environment demonstrated mapping accuracy of 95.4%, localization errors under 2 cm, and rapid recovery from disturbances within 0.8 seconds.

Keywords – Construction Robotics, Reinforcement Learning, Unmanned Aerial Vehicle, Machine Learning

1 Introduction

The construction industry encompasses a wide range of tasks that are labor-intensive, repetitive, and often expose workers to hazardous conditions. Tasks such as material transport, structural inspections, and equipment monitoring demand precision and thorough coverage to meet stringent project standards and ensure site safety [1]. Material transport, for instance, frequently involves maneuvering heavy loads across uneven and challenging terrain, posing physical risks to workers and increasing logistical complexity [2]. Similarly, structural and visual inspections are critical for maintaining safety and quality but often require access to elevated or confined spaces that are difficult or unsafe for human workers. As construction projects grow in scale and complexity, there is an increasing demand for automation to enhance worker safety, improve operational efficiency, and uphold the high-quality standards required by large-scale projects [3].

To address these challenges, drones have emerged as a promising solution, automating tasks that traditionally require significant manual effort [4]. Drones offer numerous advantages, including rapid data collection, the ability to navigate hard-to-reach areas, and operational flexibility across diverse scenarios. Despite these benefits, achieving precise navigation and stable flight in dynamic construction environments remains a significant challenge. Construction sites are often cluttered and ever-changing, with complex layouts, sudden obstacles, and environmental disturbances such as wind gusts, debris, or equipment vibrations. Traditional drone systems frequently struggle to maintain steady flight paths, consistent altitude, and reliable navigation under these conditions, resulting in task inaccuracies, inefficiencies, and potential safety risks.

Advanced sensor integration in Unmanned Aerial Vehicle (UAV) systems has shown great potential to address these limitations by enhancing stability, navigation precision, and situational awareness [5]. Sensors such as Light Detection and Ranging (LiDAR), GPS, high-resolution cameras, and Inertial Measurement

Units (IMU) allow drones to collect comprehensive data about their surroundings. LiDAR provides precise distance measurements and environmental mapping, essential for spatial awareness in complex site layouts [6]. GPS offers high-level navigation strategies, while high-resolution cameras capture visual details critical for inspections and texture analysis. IMU sensors deliver real-time orientation and motion data to support stability during flight. By combining data from these sensors, drones can respond adaptively to real-time changes, improving their accuracy and operational efficiency across a range of construction tasks.

[7]However, current approaches to sensor integration often fail to fully utilize the potential of these technologies. Many systems process sensor data independently, leading to fragmented information, delayed responses, and limited adaptability to dynamic environmental changes [7]. Additionally, conventional flight control systems typically rely on basic algorithms, which struggle to capture the complex, non-linear relationships between diverse sensor inputs [8]. This results in suboptimal navigation, reduced stability, and inefficiencies in unpredictable construction settings.

To overcome these challenges, this paper proposes a multi-sensor fusion and control framework that integrates LiDAR, GPS, high-resolution cameras, and IMU sensors to significantly enhance UAV performance in construction automation. The proposed system employs a Simultaneous Localization and Mapping (SLAM) model to fuse multi-modal sensor data, generating accurate environmental maps and enabling precise localization. Additionally, a Deep Reinforcement Learning (DRL) algorithm refines flight control strategies, dynamically adjusting the UAV's responses to external disturbances such as wind gusts, vibrations, and near-miss collisions. Simulation tests conducted in a ROS-Gazebo environment validate the system's performance, and test the robustness of the designed system in various external disturbances.

2 Current Developments in Multi-Sensory Drone Flight Control for Automated Construction Tasks

2.1 Advancements in UAV Automation for Construction

In recent years, the integration of Unmanned Aerial Vehicles (UAVs) into construction workflows has accelerated, driven by the need to improve efficiency, safety, and precision across various operational tasks [9]. UAVs provide significant advantages over traditional manual methods, including faster data collection, the ability to reach hazardous or inaccessible areas [10], and reduced physical risk to human workers. These benefits have made UAVs invaluable tools in tasks such as material transport, aerial surveying, progress tracking, and structural inspection, where timely and accurate data collection is essential for effective project management [11]. Current applications primarily employ high-resolution cameras and LiDAR sensors to capture comprehensive visual and spatial data of construction sites [12], enabling tasks such as topographical mapping, structural monitoring, and site progress tracking. However, the potential of UAVs in construction automation remains largely untapped, with current systems often limited to basic applications due to challenges in achieving the level of control required for more complex, precision-oriented tasks.

2.2 Challenges in UAV Flight Control in Dynamic Construction Environments

Despite their advantages, UAVs face considerable challenges in maintaining stable navigation and precise flight control within dynamic and cluttered construction environments. Construction sites are characterized by rapidly changing conditions, including variable layouts, the presence of obstacles, and environmental disturbances such as wind and dust. These factors complicate UAV navigation, potentially causing navigation errors and inefficient task performance. In complex tasks that require close-range inspections or consistent altitude maintenance, such as detailed structural inspections, UAV stability becomes even more critical. However, most existing UAV control systems are built on basic sensor setups and simple algorithms, which struggle to adapt to environmental changes and cannot provide the required precision. As a result, UAVs may lack the stability and adaptability needed for more intricate construction tasks, where accuracy is paramount.

2.3 Sensor Integration and Advanced Control Algorithms in UAV Systems for Construction

To enhance UAV functionality in complex construction environments, the integration of multiple sensors has become crucial. This multi-sensor approach allows UAVs to collect detailed, varied data, increasing their situational awareness and enabling adaptive responses to real-time changes [13]. Key sensors such as LiDAR, GPS, high-resolution cameras, and infrared sensors each contribute uniquely to UAV operations [14]. LiDAR sensors generate 3D environmental maps and measure distances, aiding obstacle avoidance and altitude maintenance, while GPS supports broader site navigation, though it often lacks the precision required for dense or cluttered construction spaces. High-resolution cameras capture critical visual information for texture analysis,

defect detection, and adherence to specifications, providing UAVs with the ability to identify quality issues in structural elements. This diverse sensor data creates a richer environmental model, allowing UAVs to operate with improved accuracy, safety, and effectiveness.

The integration of such varied data sources has also driven advances in sensor fusion and control algorithms, addressing the limitations of traditional UAV systems that often process sensor data separately, leading to delays and lower precision in dynamic construction sites. Recent innovations in sensor fusion techniques unify data from multiple sensors to create a comprehensive environmental model, enabling more precise UAV control [15]. Deep Neural Network (DNN)-based models are particularly effective in this context, as they capture non-linear relationships between sensor inputs, enhancing UAV environmental mapping and control accuracy [16]. Techniques such as Simultaneous Localization and Mapping (SLAM), especially when combined with neural networks, further improve localization and trajectory planning by dynamically adjusting flight paths in real time [17]. Additionally, Deep Reinforcement Learning (DRL) algorithms allow UAVs to refine control strategies by continuously learning from environmental feedback, adapting to unexpected obstacles, wind disturbances, or other sudden changes while maintaining stable flight [18]. Together, these advancements provide a framework for UAVs to achieve greater accuracy, adaptability, and resilience, particularly in the unpredictable conditions of construction sites where conventional systems may struggle.

2.4 This Paper's Contribution for Enhanced UAV Control in Construction Automation

Despite notable advancements in sensor integration and flight control technologies, the application of these innovations in construction automation remains limited. Current research predominantly emphasizes UAV navigation and data collection, often neglecting the critical need for precise and stable flight control in scenarios requiring close-range accuracy, such as detailed structural inspections or dynamic obstacle avoidance. This gap underscores the potential of advanced sensor fusion, SLAM, and reinforcement learning techniques to address the unique challenges of UAV deployment in construction environments.

Therefore, this paper seeks to advance the state of UAV control in construction by presenting a multi-sensor fusion framework that integrates LiDAR, GPS, high-resolution cameras, and IMU sensors with advanced control algorithms. The SLAM system fuses multi-modal sensor data to generate accurate environmental maps and enable precise localization. Additionally, a Deep Reinforcement Learning (DRL) algorithm is employed to dynamically adjust flight parameters, allowing the UAV to maintain stability under external disturbances such as wind gusts, vibrations, and near-miss collisions.

3 Multi-Sensor Fusion and Control Framework

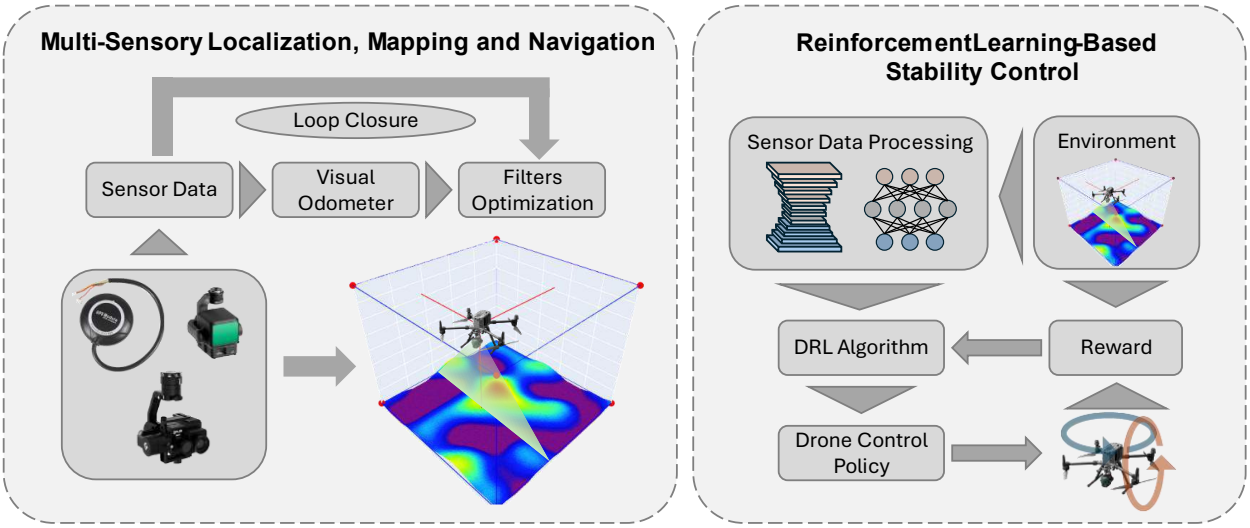


Figure 1. Multi-Sensor Fusion and Control Framework Method Overview

This framework is designed to enhance UAV flight control and stability for high-precision construction tasks by integrating advanced sensor fusion and control strategies as shown in Figure 1. The system consists of

two distinct modules: Multi-Sensory Localization, Mapping, and Navigation and Reinforcement Learning-Based Stability Control. Together, these modules enable the UAV to navigate complex environments, maintain balance under disturbances, and dynamically adapt to real-time changes. Each module is built on a combination of sensor data processing, algorithmic optimization, and machine learning techniques.

3.1 Multi-Sensory Localization, Mapping, and Navigation

Accurate environmental understanding and navigation are fundamental for UAVs operating in complex construction sites. This system integrates data from LiDAR, cameras, GPS, and IMU to construct a real-time, 3D map of the environment and provide reliable localization and navigation control.

3.1.1 Sensor Data Acquisition and Pre-Processing

The UAV is equipped with a suite of sensors that provide complementary data. LiDAR captures 3D point clouds, offering spatial awareness critical for obstacle detection and surface mapping. High-resolution cameras deliver RGB images for texture analysis and visual feature detection. The IMU provides angular velocity and acceleration for motion estimation, while GPS supplies coarse-grain global positioning for initial localization.

Raw data undergoes pre-processing to enhance accuracy and consistency. For LiDAR, a moving average filter is applied to reduce noise:

$$\widetilde{D}(i) = \frac{1}{N} \sum_{j=i-\frac{N}{2}}^{i+\frac{N}{2}} D(j) \quad (1)$$

where $D(j)$ represents the distance measured at point j and N is the window size. Given the need for rapid data processing and flight adjustments, the moving average filter presents a balanced trade-off between efficiency and accuracy.

3.1.2 Visual Odometry

Using pre-processed image data, the system detects and tracks visual features between consecutive frames. Pose estimation is calculated based on matched feature points, and IMU data is fused to refine these estimates:

$$\mathbf{T}_{k+1} = \mathbf{T}_k \cdot \Delta\mathbf{T}_{\text{vision}} \cdot \Delta\mathbf{T}_{\text{IMU}} \quad (2)$$

where $\mathbf{T}_k$ is the pose at time k, $\Delta\mathbf{T}_{\text{vision}}$ is the incremental pose estimated from visual odometry, and $\Delta\mathbf{T}_{\text{IMU}}$ accounts for short-term motion from IMU readings.

3.1.3 SLAM Framework

A Simultaneous Localization and Mapping (SLAM) algorithm optimizes the UAV's trajectory and generates a globally consistent map. The SLAM framework integrates LiDAR, visual odometry, and GPS data, estimating the UAV's position $P'(t)$ and orientation $\theta'(t)$ by minimizing localization error:

$$L_{\text{SLAM}} = \|P(t) - P'(t)\|^2 + \|\theta(t) - \theta'(t)\|^2 \quad (3)$$

where $P(t)$ and $\theta(t)$ are ground truth values. Loop closure constraints are incorporated to correct drift over time, ensuring trajectory accuracy.

The outputs from the SLAM framework guide the UAV's navigation system. Real-time path planning is performed using the fused environmental map, enabling obstacle avoidance and dynamic adjustments to environmental changes. The result is precise trajectory execution, even in highly cluttered and variable construction sites.

3.2 Reinforcement Learning-Based Stability Control

In addition to navigation, maintaining flight stability under disturbances such as wind gusts, airborne debris, or dynamic obstacles is critical for UAV operation. This system employs a Deep Reinforcement Learning (DRL) agent to adaptively adjust the UAV's flight parameters in real-time.

3.2.1 Sensor Data Processing

The stability control system uses fused data from the IMU, LiDAR, and cameras to define the UAV's state. The IMU provides orientation measurements (pitch, roll, yaw) and acceleration for disturbance detection, while LiDAR monitors altitude and proximity to surfaces. Cameras contribute visual feedback, which is particularly useful for fine-grained stability adjustments during precision tasks.

3.2.2 Markov Decision Process (MDP) Framework

The UAV's control system is modeled as an MDP, defined by the tuple (S, A, R, P), where:

State S: Fused sensor data, including position $P(t)$, orientation $\theta(t)$, and environmental features.

Action A: Control commands for thrust, roll, pitch, and yaw adjustments.

Reward R : A stability-based reward function penalizing deviations from reference position P_{ref} and orientation θ_{ref}:

$$R = -\alpha\|P - P_{\text{ref}}\|^2 - \beta\|\theta - \theta_{\text{ref}}\|^2 \quad (4)$$

where α and β are weighting factors.

Transition Probability P : The probability of

transitioning to a new state given the current state and action.

3.2.3 Deep Reinforcement Learning Algorithm

A policy gradient method optimizes the control policy $\pi(a \mid s)$ to maximize the expected cumulative reward:

$$J(\pi) = \mathbb{E}\left[\sum_{t=0}^{T} R(t)\right] \quad (5)$$

Through iterative training in simulated environments, the DRL agent learns to adapt to disturbances and maintain stable flight under various conditions.

In real-world deployment, the trained DRL agent dynamically adjusts flight parameters based on environmental feedback. For example, in the presence of wind gusts, the agent compensates by modulating thrust and adjusting orientation to minimize deviations from the intended trajectory. The system achieves rapid recovery from disturbances, maintaining consistent altitude and balance for precision tasks.

4 Implementations

The proposed framework was implemented and tested in a simulated ROS environment using Gazebo as the simulation platform. Gazebo provided a realistic simulation of construction site environments, including dynamic objects, complex terrains, and environmental physics. The testing process was divided into two stages: evaluating the UAV's environmental mapping and navigation capabilities, and assessing its stability control under various external disturbances.

4.1 Environmental Mapping and Navigation Testing

In the first stage, the UAV was tasked with collecting environmental data and navigating autonomously in a simulated construction site. The site included obstacles such as scaffolding, equipment, and uneven surfaces, emulating real-world conditions. The UAV's primary objectives were:

Environmental Mapping: Using LiDAR, cameras, and IMU sensors, the UAV generated a 3D map of the environment. The map's accuracy was evaluated using:

Mapping Accuracy: Intersection-over-Union (IoU) between the generated map and the ground truth.

Localization Error: Root Mean Square Error (RMSE) between estimated and true positions.

Navigation Performance: The UAV navigated predefined waypoints while avoiding obstacles. Navigation metrics included:

Path Efficiency: Ratio of the shortest path length to the actual path length.

Obstacle Avoidance Success: Percentage of successfully avoided obstacles.

Gazebo provided realistic physics for the environment, such as collisions, lighting effects for cameras, and terrain irregularities, enhancing the fidelity of the simulation.

4.2 Stability Control Testing Under External Disturbances

In the second stage, the UAV's ability to maintain stability and recover from disturbances was evaluated.

External disturbances relevant to construction sites were simulated in Gazebo, including: **Wind Gusts**: Random lateral forces were applied to mimic the effect of wind blowing through open or partially enclosed construction sites. The UAV was tested on its ability to maintain its trajectory and recover from sudden deviations caused by these forces. **Vibrations from Heavy Machinery**: Periodic oscillatory forces were introduced to replicate the effect of operating heavy equipment such as drills, concrete mixers, or excavators near the UAV. These vibrations tested the system's robustness to continuous destabilizing inputs. **Dynamic Load Changes**: External weights were dynamically added and removed from the UAV during flight, simulating scenarios where it might carry tools, sensors, or materials. This tested the system's ability to adapt to changes in payload and maintain stability. **Torsional Forces**: Rotational disturbances were applied to simulate uneven thrust distribution or sudden torque imbalances caused by dynamic construction operations. This tested the UAV's ability to quickly stabilize its orientation under unexpected rotational forces. **Near-Miss Collisions**: Fast-moving objects (e.g., scaffolding, cranes) were simulated to narrowly pass by the UAV. When encountering such a near-miss situation, the UAV was required to immediately halt its movement, stabilize itself, and correct any resulting disturbances. This scenario tested the drone's ability to rapidly switch to emergency stop procedures while ensuring stability post-disruption. **Dynamic Obstacles**: Moving objects, such as simulated workers or operational machinery (e.g., cranes), introduced additional challenges for stability and trajectory adjustment. These obstacles required the UAV to maintain stability and dynamically adjust its flight path in real-time to avoid collisions.

Gazebo's physics engine enabled precise simulation of these disturbances, ensuring realistic force dynamics and environmental interactions.

4.3 Metrics for Evaluation

The UAV's performance under disturbances was

assessed using: **Recovery Time**: The time taken for the UAV to return to stable flight after experiencing a disturbance. **Stability Deviation**: Standard deviation of orientation angles (pitch, roll, yaw) during and after disturbances

$$\sigma_\theta = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\theta_i - \bar{\theta})^2} \quad (6)$$

where θ_i is the orientation angle at time i, and $\bar{\theta}$ is the mean orientation angle. **Trajectory Accuracy**: Deviation of the UAV's actual path from the planned trajectory. **Force Compensation**: Percentage of external forces effectively neutralized by the stability control system.

5 Results

The proposed framework was tested for 100 rounds in the ROS-Gazebo environment to evaluate its performance each in mapping/navigation tasks and stability control under simulated construction conditions. The results demonstrated the framework's capability to achieve accurate localization, robust navigation, and reliable stability in the presence of external disturbances.

5.1 Mapping and Navigation Performance

During the mapping and navigation tests, the UAV demonstrated its ability to collect detailed environmental maps and execute precise navigation tasks in the simulated construction site. The system achieved a high mapping accuracy, with an Intersection-over-Union (IoU) score of 95.4%, indicating substantial overlap between the generated map and the ground truth. Localization was also highly accurate, with a Root Mean Square Error (RMSE) of only 1.95 cm between the UAV's estimated and actual positions. In terms of navigation performance, the UAV showed efficient path planning, with a path efficiency ratio of 1.08, meaning the UAV's actual flight path was only slightly longer than the shortest possible path. Additionally, the system excelled in obstacle avoidance, successfully avoiding 96% of obstacles in its path. These results validate the effectiveness of the multi-sensory fusion system in enabling the UAV to navigate complex construction environments with precision and reliability.

For comparative evaluation, we benchmarked our SLAM-integrated fusion framework against conventional sensor fusion methods that independently process LiDAR, IMU, and visual data. Our results show that the proposed method improved localization accuracy by 22.5% and mapping accuracy by 17.8%, attributed to the integration of loop closure techniques that effectively mitigate trajectory drift. This enhancement demonstrates the robustness of our method in generating precise environmental maps under dynamic conditions.

5.2 Stability Control Performance

The stability control system was evaluated across six different disturbance scenarios, representing common challenges in construction environments. Each scenario was tested multiple times, and the UAV's ability to respond, recover, and maintain stability was measured using metrics such as recovery time, orientation deviation, and trajectory accuracy. Table 1 summarizes the performance results.

Table 1. Summary of Performance Metrics Under Different Disturbance Conditions

Disturbance	Recovery Time (s)	Orientation Deviation (°)	Trajectory Deviation (%)	Force Compensation (%)
Wind Gusts	1.3	2.8	3.40%	95%
Vibrations from Machinery	1.4	2.5	**0.60%**	97%
Dynamic Load Changes	1.8	3.1	1.10%	86%
Torsional Forces	1.5	2.9	2.60%	85%
Near-Miss Collisions	**0.8**	**2**	4.80%	97%
Dynamic Obstacles	1.6	3	8.20%	**98%**

The stability control system exhibited strong performance across various disturbance scenarios, effectively maintaining flight stability and compensating for external forces. Among the tests, vibrations from machinery demonstrated the best performance, with the UAV achieving a recovery time of 1.4 seconds, an orientation deviation of only 2.5°, and a trajectory accuracy deviation of just 0.60%, the lowest across all scenarios. This highlights the system's robustness against continuous oscillatory forces commonly encountered near heavy equipment.

Conversely, the dynamic obstacle scenario posed the greatest challenge. While the UAV achieved a high force compensation rate of 98%, its trajectory accuracy deviation reached 8.20%, the highest among all tests, due to the complex task of avoiding fast-moving objects

while maintaining stability. Despite this, the system's recovery time of 1.6 seconds and orientation deviation of 3.0° indicate its adaptability to highly dynamic environments.

The near-miss collision test highlighted the system's effectiveness in emergency scenarios. The UAV executed an immediate stop with a recovery time of just 0.8 seconds, the fastest among all disturbances, and managed a low orientation deviation of 2.0°. Although the trajectory deviation of 4.80% was higher due to the abrupt braking, the system compensated 97% of the external forces, demonstrating its reliability in handling rapid, high-stress situations.

To highlight the advantages of our reinforcement learning-enhanced flight control system, we compared its performance with a conventional PID-based control strategy. The proposed system demonstrated superior performance in maintaining stability under dynamic disturbances, reducing trajectory deviation by 35.2% and improving recovery time by 21.7%. These improvements are attributed to the reinforcement learning module's ability to adaptively adjust control parameters in response to real-time environmental changes, unlike PID controllers that rely on predefined gain tuning.

In addition to evaluating stability performance, we assessed the computational efficiency of the proposed framework to ensure its feasibility for real-time UAV applications. The system achieved a feedback rate of 10 Hz on an NVIDIA Jetson AGX Xavier, a commonly used edge computing device. This result highlights the framework's capacity to maintain real-time responsiveness in dynamic construction environments.

Overall, the UAV showed consistent recovery across most scenarios, with recovery times ranging from 0.8 to 1.8 seconds and orientation deviations remaining under 3.1°. The framework effectively balanced rapid stabilization and precision, making it suitable for complex construction environments.

6 Conclusions

This paper proposed a multi-sensor fusion and control framework to enhance UAV performance in dynamic construction environments. The integration of LiDAR, GPS, high-resolution cameras, and IMU sensors enabled precise environmental mapping, accurate localization, and robust stability control. The SLAM system provided reliable trajectory optimization and mapping, while a Deep Reinforcement Learning algorithm ensured adaptive responses to disturbances such as wind gusts and near-miss collisions. Simulation results demonstrated the system's effectiveness, achieving mapping accuracy of 95.4%, localization errors under 2 cm, and rapid recovery from disturbances within 0.8 seconds.

Despite these advancements, the framework's evaluation was limited to simulation environments, and real-world implementation is needed to assess its robustness under unpredictable site conditions. Future work will focus on addressing these challenges by optimizing computational efficiency, validating the framework in field conditions, and exploring the integration of collaborative multi-UAV systems to further enhance construction automation.

References

[1] H. Liang, S. Zhang, and Y. Su, “The structure and emerging trends of construction safety management research: a bibliometric review,” *International Journal of Occupational Safety and Ergonomics*, vol. 26, no. 3, pp. 469–488, Jul. 2020, doi: 10.1080/10803548.2018.1444565.

[2] C. N. Anil Kumar, M. Sakthivel, R. K. Elangovan, and M. Arularasu, “Analysis of Material Handling Safety in Construction Sites and Countermeasures for Effective Enhancement,” *The Scientific World Journal*, vol. 2015, no. 1, p. 742084, Jan. 2015, doi: 10.1155/2015/742084.

[3] M. Tafazzoli, “Construction Automation and Sustainable Development,” *Automation and Robotics in the Architecture, Engineering, and Construction Industry*, pp. 73–95, 2022, doi: 10.1007/978-3-030-77163-8_4.

[4] H. W. Choi, H. J. Kim, S. K. Kim, and W. S. Na, “An Overview of Drone Applications in the Construction Industry,” *Drones 2023, Vol. 7, Page 515*, vol. 7, no. 8, p. 515, Aug. 2023, doi: 10.3390/DRONES7080515.

[5] T. Ren and H. Jebelli, “Developing an Advanced Perception System for Robotic Attic Insulation Using Large Vision-Language Models and LiDAR,” in *2024 International Conference on Computing in Civil Engineering (i3CE)*, Pittsburgh, 2025.

[6] T. Ren and H. Jebelli, “Robotic Navigation and Obstacle Avoidance System for Advancing Attic Insulation,” in *2024 International Conference on Computing in Civil Engineering (i3CE)*, Pittsburgh, 2025.

[7] N. Dovzhenko, Y. Ivanichenko, P. Skladannyi, and O. Zhyltsov, “Research of UAV and sensor network integration features for routing optimization and energy consumption reduction,” *elibrary.kubg.edu.uaN Dovzhenko, Y Ivanichenko, P Skladannyi, O ZhyltsovCybersecurity Providing in Information and Telecommunication, 2024•elibrary.kubg.edu.ua*, 2024, Accessed:

Mar. 10, 2025. [Online]. Available: https://elibrary.kubg.edu.ua/id/eprint/50160/

[8] Z. Zuo, C. Liu, Q. L. Han, and J. Song, "Unmanned Aerial Vehicles: Control Methods and Future Challenges," *IEEE/CAA Journal of Automatica Sinica*, vol. PP, no. 99, pp. 1–14, Jan. 2022, doi: 10.1109/JAS.2022.105410.

[9] D. D. York, A. J. Al-Bayati, and Z. Y. Al-Shabbani, "Potential Applications of UAV within the Construction Industry and the Challenges Limiting Implementation," *Construction Research Congress 2020: Project Management and Controls, Materials, and Contracts - Selected Papers from the Construction Research Congress 2020*, pp. 31–39, 2020, doi: 10.1061/9780784482889.004.

[10] T. Ren and H. Jebelli, "Efficient 3D robotic mapping and navigation method in complex construction environments," *Computer-Aided Civil and Infrastructure Engineering*, 2024, doi: 10.1111/MICE.13353.

[11] S. Jordan *et al.*, "State-of-the-art technologies for UAV inspections," *IET Radar, Sonar & Navigation*, vol. 12, no. 2, pp. 151–164, Feb. 2018, doi: 10.1049/IET-RSN.2017.0251.

[12] J. G. Martinez, G. Albeaino, M. Gheisari, R. R. A. Issa, and L. F. Alarcón, "iSafeUAS: An unmanned aerial system for construction safety inspection," *Autom Constr*, vol. 125, p. 103595, May 2021, doi: 10.1016/J.AUTCON.2021.103595.

[13] T. H. H. Chan, "Real-time multi-sensory monitoring by automated routing of UAV in civil/building construction site and cavern development: to improve safety, health and productivity," 2017, Accessed: Nov. 16, 2024. [Online]. Available: http://dspace.cityu.edu.hk/handle/2031/87

[14] F. Z. Saadaoui, N. Cheggaga, and N. E. H. Djabri, "Multi-sensory system for UAVs detection using Bayesian inference," *Applied Intelligence*, vol. 53, no. 24, pp. 29818–29844, Dec. 2023, doi: 10.1007/S10489-023-05027-Z/TABLES/4.

[15] C. Nguyen, V. Sagan, S. Bhadra, and S. Moose, "UAV Multisensory Data Fusion and Multi-Task Deep Learning for High-Throughput Maize Phenotyping," *Sensors 2023, Vol. 23, Page 1827*, vol. 23, no. 4, p. 1827, Feb. 2023, doi: 10.3390/S23041827.

[16] F. Corradi, G. Adriaans, and S. Stuijk, "Gyro: A digital spiking neural network architecture for multi-sensory data analytics," *ACM International Conference Proceeding Series*, pp. 9–15, Jan. 2021, doi: 10.1145/3444950.3444951.

[17] A. Steenbeek and F. Nex, "CNN-Based Dense Monocular Visual SLAM for Real-Time UAV Exploration in Emergency Conditions," *Drones 2022, Vol. 6, Page 79*, vol. 6, no. 3, p. 79, Mar. 2022, doi: 10.3390/DRONES6030079.

[18] J. Wu, Y. Ye, and J. Du, "Multi-objective reinforcement learning for autonomous drone navigation in urban areas with wind zones," *Autom Constr*, vol. 158, p. 105253, Feb. 2024, doi: 10.1016/J.AUTCON.2023.105253.

BIM-Driven Robotic Disassembly for Resource Recovery in End-of-Life Buildings

Jeehoon Kim[1] and Christopher Rausch[1]

[1]Fariborz Maseeh Department of Civil, Architectural, and Environmental Engineering, The University of Texas at Austin, USA
E-mail: jeehoon@utexas.edu and c.rausch@utexas.edu

Abstract –

Deconstruction for resource recovery offers a sustainable alternative to traditional demolition practices. However, deconstruction remains labor-intensive and time-consuming compared to demolition of buildings. This study proposes a novel framework that integrates Building Information Modeling (BIM) with robotic systems to automate the recovery of building components. The BIM file is converted into a robot simulation environment, enabling precise disassembly planning by providing location and size information of recoverable resources. A case study on a BIM model of an exposed interior residential wall assembly unit demonstrates the methodology's effectiveness in automating resource recovery of end-of-life building components. The results validate the potential of the proposed framework to reduce manual labor and contribute to the circular economy of the building construction sector.

Keywords –

Building Construction; Building Information Modeling; Circular Economy; Deconstruction; Resource Recovery; Robotic Disassembly; Sustainable Construction;

1 Introduction

The building construction sector plays a critical role in contributing to climate change. According to the United Nations Environment Programme (UNEP), the sector is responsible for approximately 37% of global energy-related carbon emissions [1]. Mitigating emissions within the construction supply chain requires transitioning from a linear economy to a circular economy. A circular economy in building construction emphasizes increasing reuse, recycle, reclaim of resources, thereby minimizing waste and resource production throughout the supply chain (Fig. 1).

Resource recovery from buildings at the end-of-life stage has garnered attention as a strategy to reduce the environmental impact of the building sector. As of 2022, the production of construction materials alone contributed 10% of global energy and process carbon emissions [2]. Additionally, the United States Environmental Protection Agency (EPA) reported that demolition activities in 2018 generated over 500 million tons of landfill debris in the U.S. [3]. By recovering resources through reuse, recycling, and reclamation, the volume of waste directed to landfills and the need for virgin materials can be markedly reduced.

Despite its environmental benefits, the deconstruction of buildings for resource recovery tends to be labor-intensive and time-consuming, requiring up to ten times longer processes compared to traditional demolition methods [4]. Improving the efficiency of resource recovery through automation offers a promising pathway to enhance the feasibility of a deconstruction process.

This study proposes a framework for planning automated resource recovery from end-of-life building assemblies using robotic systems. The approach leverages Building Information Modeling (BIM) enriched with resource information for each component. BIM files are converted into a simulation environment while retaining the resource information. This information is then used to plan robotic disassembly operations which can also be applied to physical world environments.

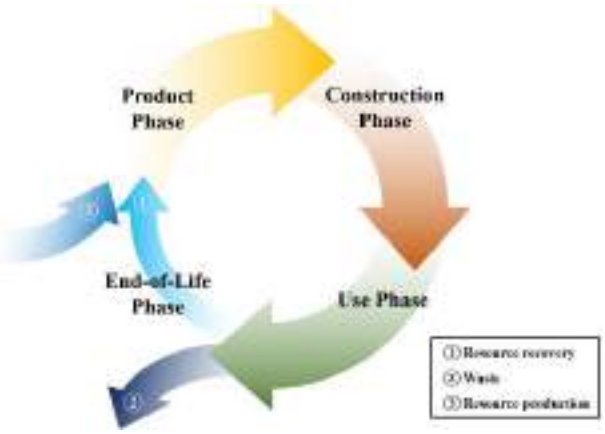


Fig. 1. Circular economy of building construction

2 Literature Review

Efforts to deconstruct buildings and effectively retrieve reusable resources have been widely studied. These approaches can generally be categorized into two main approaches: deconstruction planning up front during the design phase and planning once a building has reached its end-of-life phase.

Deconstruction planning during the design phase is commonly referred to as Design for Deconstruction (DfD). DfD involves incorporating strategies during the early design stages to simplify future resource recovery. These strategies include measures such as using bolted connections instead of adhesives [5], avoiding the use of toxic materials [6], and designing for offsite construction [7], among others. The primary goal of DfD is to streamline deconstruction efforts while minimizing the labor, cost, and environmental impact of the process.

Notable advancements in this field include the integration of BIM to simulate environmental and cost impacts throughout a building's lifecycle, enabling the optimization of designs to minimize these effects. For example, studies have developed BIM-integrated frameworks that enhance construction material management in the early design phase [8, 9]. These frameworks facilitate informed material selection and improve interoperability, ultimately supporting more efficient deconstruction and resource reuse planning at a building's end-of-life.

However, most modern buildings are not designed with deconstruction in mind [6]. Consequently, deconstruction planning often defers until the building reaches its end-of-life stage. In such cases, BIM serves as a critical tool for assessing a building's condition, resource properties, and reuse potential. Mollaei et al. [10] demonstrated the use of BIM's quantity take-off tools to extract material quantities and connection configurations. This data was then integrated into a multi-objective optimization model to evaluate various deconstruction strategies for different materials. Similarly, Sanchez et al. [11] proposed a semi-automated approach for selective deconstruction programming using BIM. Their approach optimized the deconstruction and reuse strategies by incorporating cost and environmental impact analysis.

While existing methods use BIM for deconstruction planning in the design phase or at the end-of-life stage, they often rely on manual processes that are labor-intensive and time-consuming. To improve the efficiency and scalability of deconstruction, researchers have explored the integration of automation and robotic systems. For instance, Lee and Brell-Cokcan [12] proposed a Human-in-the-Loop framework in which an operator defines high-level objectives, allowing robotic systems to execute deconstruction tasks with greater precision and efficiency.

However, deconstruction planning tailored for robotic systems remains largely unexplored. This study addresses this gap by introducing a novel approach that bridges BIM-based deconstruction planning with robotic execution. This method enables practitioners to directly extract information from a BIM model to plan and execute robotic resource recovery commands, enhancing efficiency in the deconstruction process.

3 Methodology

The proposed methodology involves converting a conventional BIM file (e.g., an Autodesk Revit file) into a Universal Scene Description (USD) format, which is compatible with robotic simulation environments (e.g., specifically Isaac SIM [13]). A USD file retains the resource information and hierarchy of the components, enabling the identification of recoverable resources in the simulation. This information is then used as input for robotic disassembly planning. Fig. 2 shows the proposed framework for robotic disassembly. Details of each step are explained in the following sections.

3.1 Converting BIM to USD

The BIM file containing information about the target structure is converted into a USD file. USD is a 3D scene representation file format designed to support complex 3D workflows including robotic simulations. This conversion is facilitated by the Nvidia Omniverse Revit Connector [14], which ensures that all subcategory information from the BIM file is preserved during the transition to the simulation environment. The resulting USD file retains critical resource specifications and hierarchical structural data necessary for accurate robotic simulations.

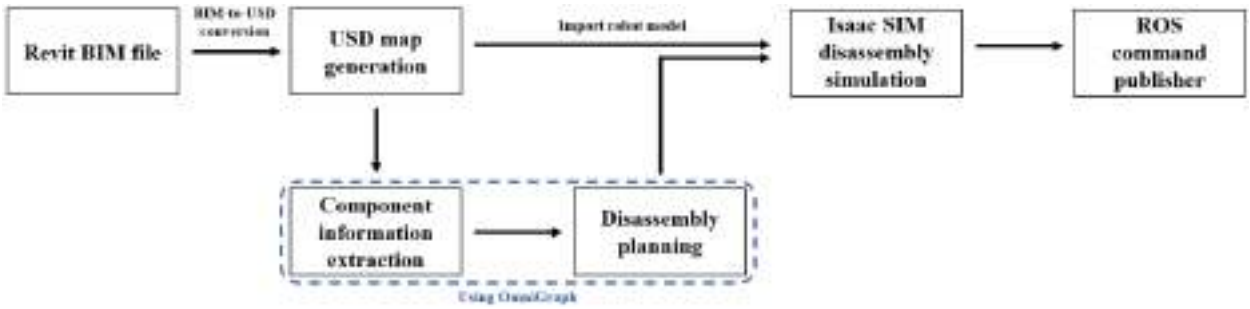


Fig. 2. Proposed framework for robotic disassembly using BIM information

3.2 Using Component Information for Disassembly Planning

The converted USD file is structured with elements defined as *prims* (short for primitives). A prim serves as a container for various attributes within any component in the USD scene, such as cameras, lights, objects, or robots. These prims are organized in a hierarchical structure, including the location and size information of specific components in their subcategories. Each object prim contains transformation coordinates and extent values. The transformation coordinate corresponds to the center of the component and the extent values define the min/max coordinates of the bounding box enclosing the component (Fig. 3).

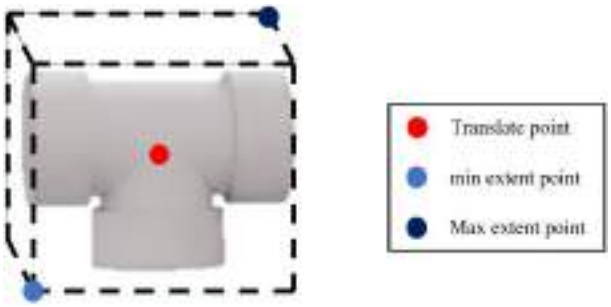


Fig. 3. Visualized example of component information in a prim

To plan and execute robotic disassembly, OmniGraph is used within the Isaac SIM environment. OmniGraph is a visual scripting language that enables dynamic control of robots [15]. It operates by connecting nodes, which perform specific tasks and contain input, output, and state attributes. These nodes are linked through connections, forming a network that controls robotic behaviors and interactions in the simulation. Action Graphs, a type of OmniGraph, define the execution of specific tasks, allowing for precise and flexible robot control.

Additionally, OmniGraph commands can be translated into ROS2 commands for physical world implementation. Isaac ROS2 nodes within the action graph enable the publishing of ROS2 messages, allowing the simulated actions to be executed by physical robots in real-world experiments.

4 Experiments

To demonstrate how the BIM-to-USD conversion process can support robotic disassembly planning, a small-scale experiment was conducted using an exposed interior residential wall assembly unit (Fig. 4(a)). The BIM file for the wall assembly unit was developed with an approximate Level of Detail (LoD) 300, providing detailed information about each component's size, shape, and location. The BIM file was then converted into a USD format and imported into Isaac SIM [13], a robotic simulation platform (Fig. 4(b)). A Franka FR3 [16] robot manipulator was incorporated into the simulation, mounted on a static platform to enhance its reachability during the disassembly process.

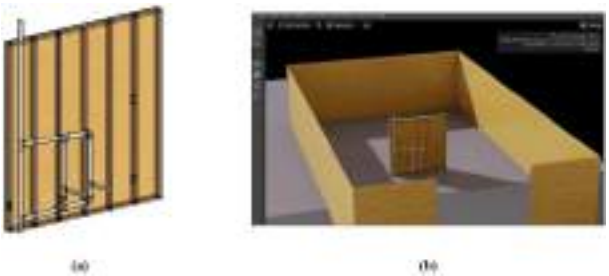


Fig. 4. Wall assembly unit: (a) LoD 300 BIM file and (b) USD file imported into Isaac SIM

Table 1. Recoverable resource attributes in wall assembly

Name	Mesh	Extent [min, max]	Translate	Material
Pipes 302099		[(-1.905,-1.905,-14.943), (1.905,1.905,14.943)]	(1213.926,252.986,38.882)	Plastic
Pipe Fittings 302081		[(-6.826,-6.826,-3.532), (6.826,3.532,3.532)]	(1236.335,252.930,105.643)	PVC
Electrical Fixtures 305293		[(-2.699,-5.080,-4.763), (2.699,5.080,5.414e^{-})]	(1233.557,141.838,33.020)	Steel

For this experiment, the primary objective was to utilize the resource information from the BIM model to demonstrate automated identification and recovery of reusable components. Table 1 shows an example of specific resources within the given wall assembly unit. As depicted in Fig. 3, the translate point indicates the center point of an object.

The disassembly task was executed using a combination of five OmniGraph action graph nodes (Fig. 5). The 'Pick and Place controller' node received inputs of the robot manipulator's joint movements, a target resource translate point, and a predefined placing point from the other four nodes. This combined algorithm enabled the robot manipulator to accurately pick up the center point of an identified resource and place it in a specified location.

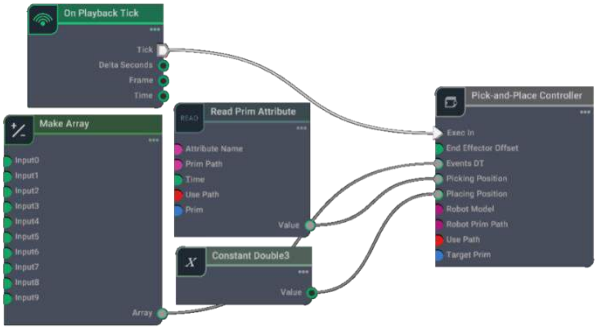


Fig. 5. OmniGraph node combination for resource pick and place task

The node combination automatically leverages the detailed positional data to perform precise disassembly operations (Fig. 6). The results show the potential of the proposed methodology in advancing resource recovery of end-of-life building components with a robotic system.

5 Conclusions

This study presents a new framework for automated resource recovery using a robotic system, addressing key challenges associated with deconstruction in the building construction sector. The methodology integrates BIM with robotic simulation tools by converting BIM files into USD format. The enriched data from the BIM model enables the detailed identification and retrieval of recoverable resources through robotic disassembly planning. A small-scale experiment demonstrates the potential of the approach, outlining the capability to utilize precise size and location information of components to retrieve target resources in the wall assembly unit. The results underscore how BIM-to-USD data conversion can directly support the potential of robotic systems as advanced disassembly tools. The proposed framework enhances the efficiency and feasibility of resource recovery and contribute to a more sustainable construction industry.

5.1 Discussion

While the RosBridge nodes of OmniGraph facilitate the conversion of simulated action graphs into ROS2 commands, significant challenges persist in replicating the simulation to physical world applications. A critical issue lies in the precise registration of simulated environments with their physical world counterparts. Addressing this challenge requires advanced machine

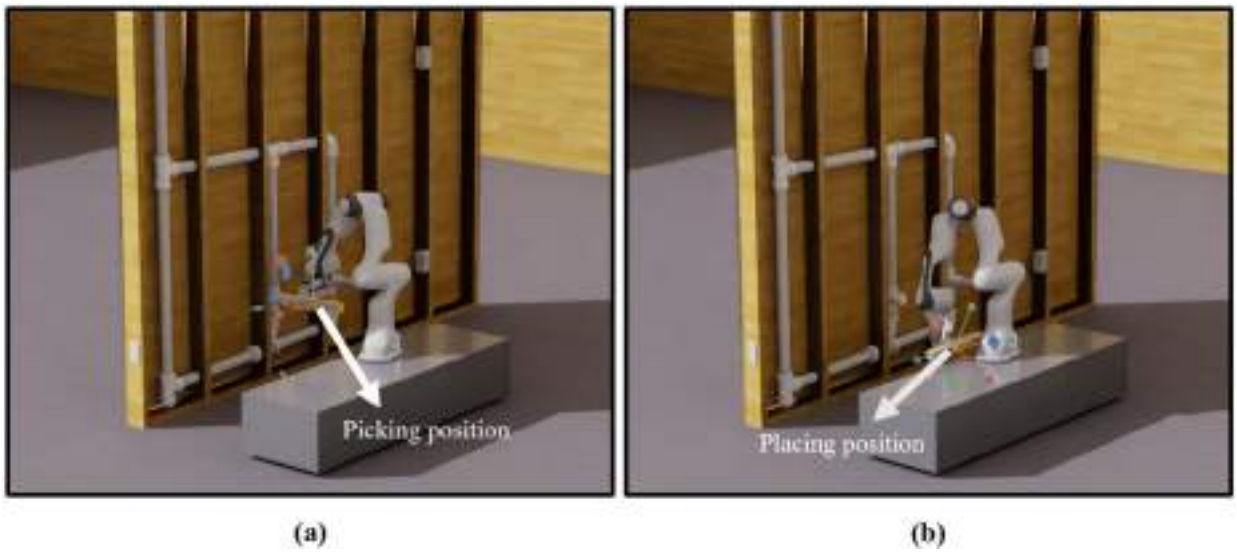


Fig. 6. Results of the robotic disassembly task: (a) picking the target object's translate point, and (b) placing the object at a specified point

vision systems capable of detecting fiducial markers (e.g., AprilTag [17] and QR code) or physical features of the building components to accurately align the simulation and physical world environments.

Future research should focus on determining the specific types and formats of building data that should be incorporated into BIM models to optimize robotic deconstruction processes. Accurate information modeling within BIM is essential for ensuring a seamless conversion to USD format using the Nvidia Omniverse Revit Connector. For example, it is critical to include compatible and detailed information identifying resources that are suitable and intended for robotic disassembly. Enhancing the fidelity of BIM-to-USD conversions should also be prioritized to ensure the accurate and comprehensive transfer of data into robotic simulation environments.

The current case study is limited to a small wall assembly unit in a simulation environment. Scaling this approach to larger, real-world applications will require addressing the complexities of diverse building materials with varying conditions. Unlike controlled simulation environments where material properties are predefined, physical structures comprise materials having different structural integrity, attachment types, and degradation levels. Developing adaptive robotic manipulation techniques capable of dynamically recognizing and adjusting disassembly strategies based on these variables is essential for maintaining efficiency. Furthermore, expanding the versatility of robotic systems is critical for addressing practical constraints (e.g., robot torque, reach, payload, and spatial maneuverability) [18]. Advancements that overcome these limitations will significantly enhance the effectiveness of the proposed framework.

Acknowledgments

The authors would like to acknowledge the financial support from the U.S. Department of Housing and Urban Development (HUD). The findings and conclusions in this manuscript are those of the authors and do not necessarily represent the official position or views of the U.S. Department of Housing and Urban Development.

References

[1] A. Dyson, N. Keena, M. Lokko, B. K. Reck, and C. Ciardullo, "Building Materials and the Climate: Constructing a New Future," *United Nations Environment Programme*, 2023.

[2] Global Alliance for Buildings and Construction (GlobalABC), *2023 Global Status Report for Buildings and Construction: Beyond foundations - Mainstreaming sustainable solutions to cut emissions from the buildings sector*. United Nations Environment Programme, 2024. doi: 10.59117/20.500.11822/45095.

[3] U.S. Environmental Protection Agency (EPA), "Advancing Sustainable Materials Management: 2018 Fact Sheet Assessing Trends in Materials Generation and Management in the United States," 2020. Accessed: Nov. 27, 2024. [Online]. Available: https://www.epa.gov/facts-and-figures-about-materials-waste-and-recycling/advancing-sustainable-materials-management

[4] A. R. Chini and S. F. Bruening, "Deconstruction and Materials Reuse in the United States," in *The*

future of sustainable construction 14, Oct. 2003, pp. 1–22. [Online]. Available: https://www.researchgate.net/publication/228759250

[5] A. Akbarnezhad, K. C. G. Ong, and L. R. Chandra, "Economic and environmental assessment of deconstruction strategies using building information modeling," *Autom Constr*, vol. 37, pp. 131–144, Jan. 2014, doi: 10.1016/j.autcon.2013.10.017.

[6] B. Guy, S. Shell, and H. Esherick, "Design for Deconstruction and Materials Reuse," in *Proceedings of the CIB Task Group 39.4*, 2006, pp. 189–209.

[7] M. Hussein, A. E. E. Eltoukhy, A. Karam, I. A. Shaban, and T. Zayed, "Modelling in off-site construction supply chain management: A review and future directions for sustainable modular integrated construction," *J Clean Prod*, vol. 310, p. 127503, Aug. 2021, doi: 10.1016/j.jclepro.2021.127503.

[8] S. Sivashanmugam, S. Rodriguez Trejo, and F. Rahimian, "BIM-integrated semantic framework for construction waste quantification and optimisation," *Autom Constr*, vol. 168, p. 105842, Dec. 2024, doi: 10.1016/j.autcon.2024.105842.

[9] O. O. Akinade *et al.*, "Waste minimisation through deconstruction: A BIM based Deconstructability Assessment Score (BIM-DAS)," *Resour Conserv Recycl*, vol. 105, pp. 167–176, Dec. 2015, doi: 10.1016/j.resconrec.2015.10.018.

[10] A. Mollaei *et al.*, "'A global perspective on building material recovery incorporating the impact of regional factors,'" *J Clean Prod*, vol. 429, p. 139525, Dec. 2023, doi: 10.1016/j.jclepro.2023.139525.

[11] B. Sanchez, C. Rausch, and C. Haas, "'Deconstruction programming for adaptive reuse of buildings,'" *Autom Constr*, vol. 107, p. 102921, Nov. 2019, doi: 10.1016/j.autcon.2019.102921.

[12] H. J. Lee and S. Brell-Cokcan, "Towards controlled semi-autonomous deconstruction," *Construction Robotics*, vol. 7, no. 3–4, pp. 253–263, Dec. 2023, doi: 10.1007/s41693-023-00111-9.

[13] NVIDIA, "Isaac SIM." Accessed: Dec. 02, 2024. [Online]. Available: https://developer.nvidia.com/isaac/sim

[14] NVIDIA, "Omniverse Revit Connector." Accessed: Dec. 02, 2024. [Online]. Available: https://docs.omniverse.nvidia.com/connect/latest/revit.html

[15] NVIDIA, "Core Concepts." Accessed: Dec. 22, 2024. [Online]. Available: https://docs.omniverse.nvidia.com/kit/docs/omni.graph.docs/latest/CoreConcepts.html

[16] FRANKA ROBOTICS, "FRANKA RESEARCH 3." Accessed: Dec. 22, 2024. [Online]. Available: https://franka.de/products

[17] E. Olson, "AprilTag: A robust and flexible visual fiducial system," in *2011 IEEE International Conference on Robotics and Automation*, IEEE, May 2011, pp. 3400–3407. doi: 10.1109/ICRA.2011.5979561.

[18] C. Rausch, S. Suh, N. Repousis, N. Titterington, and H. Nguyen, "Assessing the Viability of Robotic Disassembly of Building Components for Resource Recovery," Jun. 2024. doi: 10.22260/ISARC2024/0136.

The Development of a Low Code-based Planning Interface System for Coordinated Work of Multiple Construction Machines

Sheng Lian[1], Fumiya Matsushita[1], and Ryosuke Yajima[2]

[1]Civil Technology Division, Shimizu Corporation, Japan
[2]School of Engineering, University of Tokyo, Japan

s_lian@shimz.co.jp, f_matsushita@shimz.co.jp, yajima@i-con.t.u-tokyo.ac.jp

Abstract -

The construction industry faces labor shortages due to an aging workforce, driving the need for construction automation to boost productivity. Despite recent advances in construction robotics in Japan, the complexity of construction tasks requires coordinated operations among multiple robots. This necessitates not only advanced automation but also a sophisticated planning system to align workflows with construction plans. This study introduces a low-code Planning Interface System (PIS) using the Node-RED platform, enabling civil engineers to design and manage robot workflows without extensive programming expertise. The PIS translates construction plans into executable operational logic, leveraging Robot Operating System (ROS) middleware and the *rclnodejs* package for seamless communication. A prototype demonstrates the ability of the PIS to coordinate an excavator and dump truck in a simulated environment, performing tasks such as navigation, excavation, and dumping through ROS topics and actions. This study highlights the potential of the PIS to facilitate coordinated operations between multiple construction machines, paving the way for more autonomous and streamlined construction workflows.

Keywords -

ROS; Construction automation; Multi-robot coordination; Low-code platform

1 Introduction

The traditional construction industry has long been characterized by its labor-intensive nature, where the majority of construction activities heavily rely on human labor and manual operations [1]. However, the industry is now facing a significant labor shortage due to an aging workforce and difficulties in attracting younger workers [2]. To address this issue, the Ministry of Land, Infrastructure, Transport, and Tourism (MLIT) in Japan launched the i-Construction initiative in 2016, aiming to fully leverage information and communication technology to enhance efficiency across all stages of the construction process, from survey and design to construction and maintenance [3]. Building on this foundation, i-Construction 2.0, which was announced in April 2024, sets an ambitious goal to achieve automation of construction sites and increase productivity by 30% by 2040, a critical target considering the projected 20% decline in the working-age population by that year [4].

One of the key solutions to achieve site automation is the adoption of automated and autonomous construction machinery. In Japan, both construction machinery manufacturers and general contractors have made significant progress in developing their own construction robots, with a primary focus on earthwork projects. Examples include automated hydraulic excavators [5], dump trucks [6], and vibration rollers [7].

However, in actual construction sites, the complexity of tasks often necessitates the coordinated operation of multiple construction machines, as no single robot can handle the entire workload independently. Effective collaboration between various construction machines is therefore essential for smooth project delivery. Achieving this requires not only advanced automation for individual machinery, but also a robust planning system that is capable of designing and managing workflows in alignment with the construction plan.

This study addresses the need by focusing on the integration of construction planning with multi-robot coordination, aiming to provide civil engineers with an intuitive interface to efficiently plan and execute coordinated operations between construction machinery.

2 Research background

2.1 The Role of Construction Planning

Construction planning involves organizing and coordinating all activities and resources needed to complete a project effectively, on time, and within budget. A well-structured plan serves as the foundation for successful project delivery, considering constraints such as contract requirements, site conditions, project duration, and financial limitations [8]. Several key decisions are made dur-

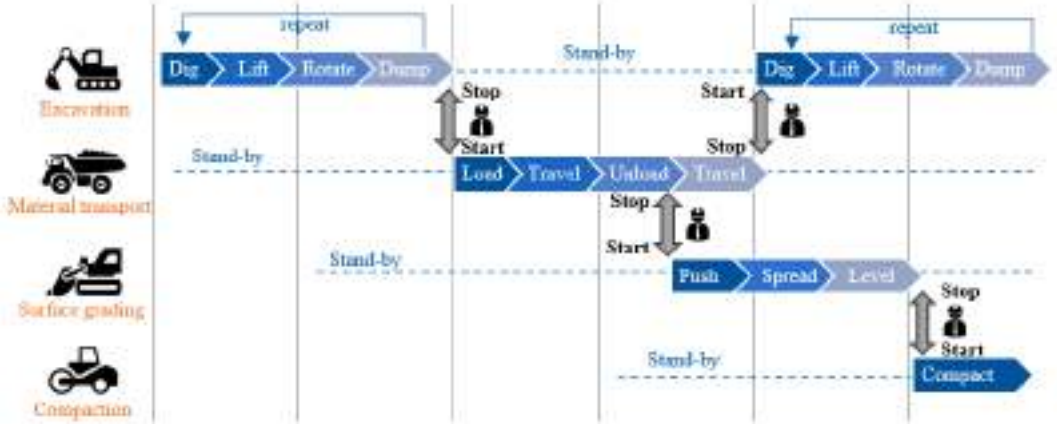


Figure 1. Conventional coordination of robot operations involving human intervention

ing this phase, including identifying specific construction tasks that need to be performed, determining the optimal number of construction machines required, and devising strategies for their arrangement and operation timings onsite. While construction robots have advanced in performing automated tasks such as excavation and material transportation, their operations often remain isolated rather than as part of an integrated construction workflow [9]. A construction plan, however, provides a framework for coordinating these robots to achieve more efficient task execution. By aligning construction operations with the overarching construction plan, isolated automated construction tasks can be transformed into a cohesive and collaborative workflow. This integration not only optimizes task execution but also facilitates smoother and more efficient project delivery.

2.2 Challenges in Multi-Robot Coordination

However, coordinating multiple robots poses significant challenges, as individual machines often lack awareness of others' positions and progress, making it difficult to determine their optimal operation timings. This limitation necessitates human intervention to monitor the overall status of construction and thus decide when to initiate or halt operations. As is demonstrated in Figure 1, in earthworks projects, construction machines such as excavators execute pre-programmed construction works (e.g., digging, lifting, dumping), but human intervention is still required to send "start" or "stop" signals for cooperative operation [10]. To achieve fully autonomous construction, robots must be able to exchange information and independently determine their operation timings. Existing methods often require complex code modifications involving conditionals, loops, and operators, demanding advanced programming expertise that civil engineers may not possess [11]. Therefore, there is a pressing need for an intuitive, accessible solution that allows engineers to define operational logic in a user-friendly manner, translating abstract ideas of collaborative robot operations into concrete, autonomous construction processes.

2.3 Objectives of This Study

The objective of this study is therefore to develop and prototype a Planning Interface System (PIS) that bridges the gap between construction planning and coordination of automated construction robots. The proposed system will allow civil engineers to translate construction plans into intuitive operational logic without requiring extensive programming skills. In addition, the PIS will provide comprehensive data visualization tools to enable civil engineers to monitor the status of construction activities and robot operations in real-time. Ultimately, this study aims to bring together all isolated automated construction tasks under a collaborative framework that aligns with the overarching construction plan, as well as to facilitate construction management process through real-time data visualization of the operational status of the site.

3 Specifications for PIS

3.1 System Requirements

The Public Works Research Institute of Japan has been advancing the Open Platform for Earthwork with Robotics and Autonomy (OPERA), an initiative designed to promote the development and social adoption of autonomous construction machinery [12]. OPERA achieves this by standardizing control signals for construction machinery, leveraging the widely adopted Robot Operating System (ROS) middleware to manage and transmit these signals. In light of the increasing trend among construction robots in Japan to adopt ROS-based control signals, it is essential that the proposed PIS is fully compatible with the ROS

framework to ensure seamless interoperability and integration. The following is a list of requirements that the proposed PIS should meet -

1. *Low-Code Visual Programming Interface:* The PIS should provide a low-code, flow-based visual programming environment to enable intuitive configuration of operational logic reflective of construction plans.
2. *ROS Integration:* The PIS must ensure compatibility with ROS middleware, enabling seamless integration with existing ROS-based robotic systems. This functionality must support the exchange of control commands and feedback data between the PIS and construction robots on-site.
3. *Data Visualization:* The PIS should provide comprehensive data visualization tools to help users monitor the status of construction activities and robot operations. Real-time dashboards and reports can assist in making informed decisions and optimizing workflows for construction management purpose.
4. *Scalability and Flexibility:* The system must be scalable to handle projects of varying sizes and complexities, while also offering flexibility to accommodate different types and manufacturers of construction equipment. It should allow users to control construction equipment in various combinations to suit specific tasks, ensuring adaptability to diverse workflows and site conditions.

3.2 Overall System Architecture

Figure 2 illustrates the overall system architecture, which comprises a construction planning system for defining detailed construction tasks, robot selection and operational instructions. This system interfaces with a centralized PIS that monitors overall vehicle statuses and issues control commands accordingly. Moreover, a group of automated construction robots are deployed at the construction site (i.e., the "edge"), where the operational data is generated and processed locally during task execution. Civil engineers utilize the PIS to translate construction plan details into operational logic that governs the collaborative behaviors of robots. By continuously acquiring and processing real-time data from edge robots regarding vehicle positions, conditions, and work progress, the PIS is able to gain a comprehensive understanding of the operational status of the site. This real-time insight allows the PIS to determine the optimal timings for each robot's tasks based on predefined logic. As a result, the PIS outputs a series of operational commands to individual robots, specifying task activation or deactivation, along with key operational parameters such as velocity and path trajectory for each robot. The output commands ensure a high level of coordination among various construction robots, aligning their operations with the overall construction plan.

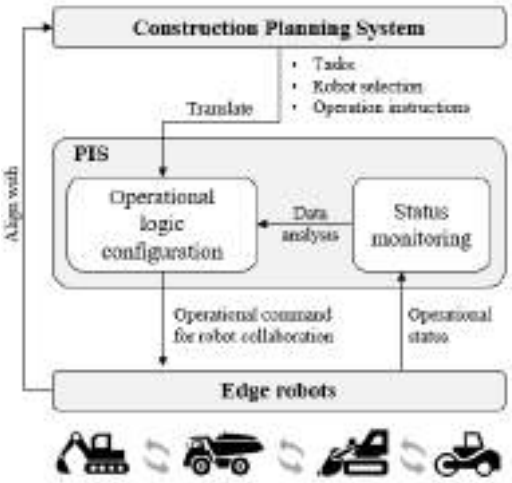


Figure 2. Overall system architecture

4 Prototype development

4.1 Platform Selection

Based on the system requirement in Section 3.1, Node-RED [13], an open-source, web-based visual programming tool, was selected as the development platform for the PIS to plan and control coordinated operation of construction robots. Its low-code, drag-and-drop interface enables civil engineers to design and configure operational logic without requiring extensive programming expertise. This allows civil engineers to focus on defining workflows that align with construction plans, efficiently translating project requirements into executable robotic behaviors.

Node-RED also excels in data visualization, providing a wide array of nodes to acquire, process, and display operational data. A node is an executable process that performs a specific computational task. By wiring up relevant nodes in Node-RED, users can quickly prototype dashboards featuring a variety of visual elements to monitor edge robots effectively and derive actionable insights into their operations. Additionally, the integration of Node-RED with ROS facilitated by the *rclnodejs* package [14] ensures seamless communication with ROS components such as topics, services, and actions. This capability allows users to visually develop and implement ROS-based control algorithms for multi-robot coordination. In this study, the proposed PIS leverages the Node-RED platform and an *rclnodejs*-based plugin software developed by EduArt Robotik [15] to directly interface with ROS.

4.2 Scenario Selection

The prototyping of the proposed PIS in this study focuses on automating the coordination of construction machines within earthworks projects, primarily due to the repetitive yet predictable nature of the workflow associated with earthwork construction. Earthwork operations, such as excavation, material transportation, and soil compaction, often require complex coordination among various construction machines, making them an ideal testing ground for collaborative automation. Additionally, the wide, open spaces with minimal external interference typical of earthwork sites further simplify the implementation of autonomous technologies, providing a controlled environment for testing and validation.

The workflow for earthworks construction can be broken down into a series of tasks, with the prototype in this study focusing on coordinating operations between an excavator and a dump truck using Node-RED and the relevant ROS client library. Figure 3 gives an example of the construction plan, specifying the detailed construction tasks involving the two machines, along with the timing as for each task to take place. To visualize and test their cooperative behavior in a controlled and risk-free environment, a Gazebo-based simulator was chosen for its simplicity and seamless integration with ROS. This ensures that the system's operational logic and coordination strategies are thoroughly validated prior to practical implementation in real-world construction scenarios.

	Scenario	Task	Trigger for the task
Task 1		**Dump truck**: Navigate to the excavation site **Excavator**: Stand-by	When the target position of dump truck is specified
Task 2		**Dump truck**: Stand-by **Excavator**: Digging, lifting, rotating, dumping (repeat)	When dump truck reaches its stand-by position at the excavation site
Task 3		**Dump truck**: Navigate to the dump site **Excavator**: Stand-by	When soil quantity reaches the weight limit of dump truck
Task 4		**Dump truck**: Soil dumping **Excavator**: Stand-by	When dump truck reaches the dump site

Figure 3. Example of a construction plan with detailed construction tasks and corresponding operation timings within an earthwork project

4.3 Prototype Development and Result Discussion

The prototype development thus involves translating the details of the construction plan above into equivalent flow-based operational logic that governs the work coordination of robots. Figure 4 shows the detailed communication flow between the Node-RED-based PIS and the edge robots. On the edge side, the robots execute high-level ROS control programs that govern their individual operations, such as navigation and excavation. Meanwhile, by incorporating the construction plan requirements into the operational logic, the PIS ensures the precise timing and conditions under which the edge control programs are executed, ensuring seamless collaboration among the machines. This operational logic is divided into three distinct parts: Part A, Part B, and Part C, each corresponding to a specific Node-RED flow, as illustrated in Figure 5. These parts collectively enable the PIS to orchestrate the coordination of tasks through intuitive flow-based programming. The following sections provide a detailed discussion of how the PIS facilitates task coordination.

4.3.1 Task 1 (Navigate to the excavation site)

Task 1 involves navigating the dump truck from its current position to reach a specified target position and orientation at an excavation site. This process implements path tracking control using the Pure Pursuit algorithm [16], which continuously calculates and adjusts the steering direction of the dump truck as it moves toward its target. This path tracking control is implemented using an action server–client communication framework in ROS. A ROS node on the edge robot initializes the action server that executes the Pure Pursuit algorithm and waits for goal requests from the Action Client a-1 in the PIS. The Action Client a-1 then sends the navigation goal to the running action server that triggers the navigation. The Node-RED configuration for this action client is detailed in Part A of Figure 5.

4.3.2 Task 2 (Excavation)

Task 2 automates the excavation process through a ROS node that integrates a subscriber and a publisher. The subscriber on the edge robot continuously monitors the **/flag_ex** topic to determine whether to proceed with excavation. Upon successful completion of Task 1 (the dump truck reaches its stand-by position at the excavation site), the action server sends the navigation result back to the PIS. This triggers the Publisher b-1 in the PIS to set the **/flag_ex** topic to "true", initiating the excavation process. The publisher as part of the excavation control program for Task 2 then sends joint position commands to control the excavator's movements. Simultaneously, the **/num_cycles** topic tracks the number of excavation cycles performed and thereby providing real-time feedback on the progress of the operation. The relevant Node-RED flow for coordinating Task 2 is shown in Part B of Figure 5.

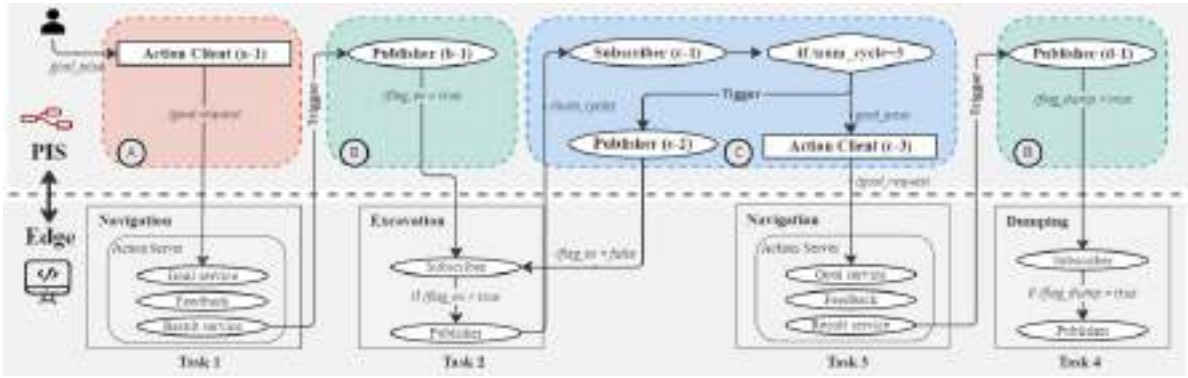


Figure 4. Overall communication flow between the PIS and edge side

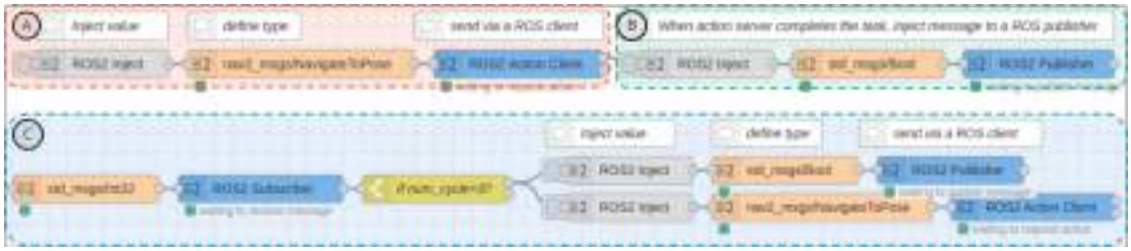


Figure 5. Node-RED flow configuration for work coordination

4.3.3 Task 3 (Navigate to the dump site)

Task 3, similar to Task 1, but this time involves navigating the dump truck to a designated dump site. In practice, excavation halts when the truck reaches its weight limit, whereas in the case of a simulated environment, this condition is approximated by counting the number of excavation cycles. The PIS sets up a Subscriber (c-1) to listen on the **/num_cycles** topic. After three cycles of excavation, the PIS triggers a Publisher (c-2) to set the **/flag_ex** topic to "false", effectively pausing the excavation work from Task 2. At the same time, the PIS sends a goal request via the Action Client c-3 to initiate the dump truck's navigation towards the dump site as per Task 3. This operational logic ensures that the dump truck starts moving only after the required amount of soil has been excavated, maintaining a smooth transition between the cessation of excavation and the initiation of material transportation. The detailed flow configuration for Task 3 is illustrated in Part C of Figure 5.

4.3.4 Task 4 (Soil dumping)

Task 4 oversees the soil dumping operation, managed by a ROS node to control the movement of the dump truck's vessel. On the edge robot side, the ROS node first starts a subscriber to monitor the **/flag_dump** topic to determine the timing for soil dumping, which then instructs the publisher to send joint control commands to the vessel actuator. Upon receiving the action results from Task 3, meaning when the dump truck has reached its designated position at the dump site, the PIS triggers the Publisher d-1 to set the **/flag_dump** topic to 'true' and hence initiates the dumping process. The setup for Publisher d-1, similar to Publisher b-1 and d-1, is detailed in Part B of Figure 5.

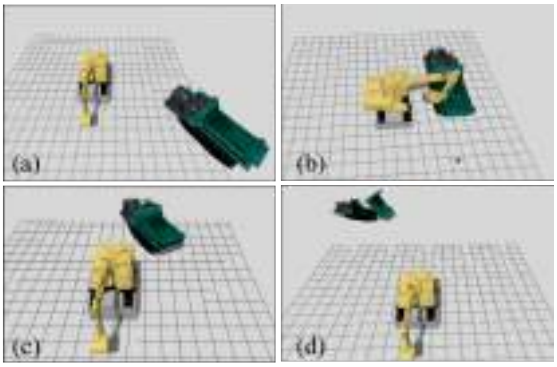


Figure 6. A snapshot of the cooperative behaviors of machines in Gazebo simulation: (a) Navigate to the excavation site; (b) Excavation and loading; (c) Navigate to the dump site; (d) Soil dumping

4.3.5 Validation of results in Gazebo simulation

Figure 6 provides a snapshot of the resulted machine behaviors in the simulation with respect to different tasks,

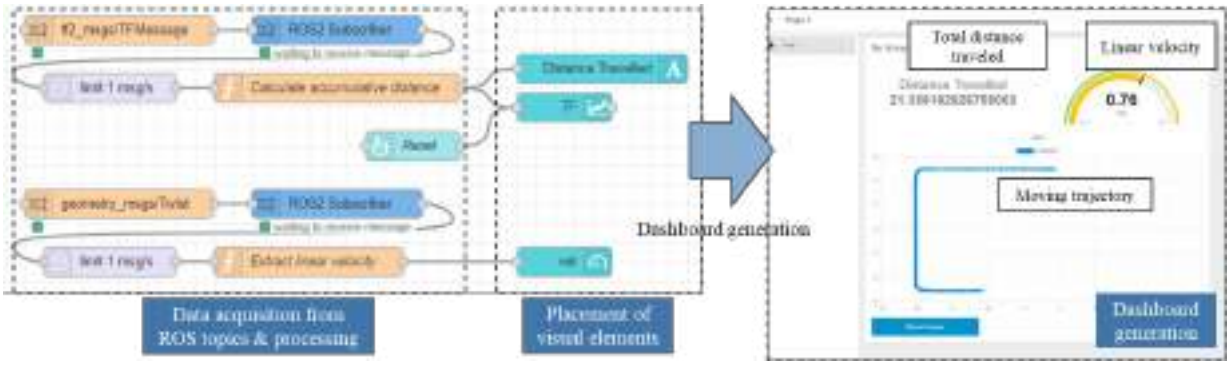


Figure 7. Sample Node-RED flow and its output dashboard for monitoring vehicle velocity, travel distance and trajectory

illustrating how the PIS orchestrates the excavator and dump truck to perform their respective tasks in a coordinated manner. The entire workflow from Task 1 (navigation to the excavation site) to Task 4 (soil dumping) is executed autonomously with no human intervention, with the PIS determining the optimal operation timings for each task based on predefined logic reflective of the construction plan. The simulation result demonstrates the potential of the PIS to design intuitive, flow-based operational logic for coordinating multiple construction tasks seamlessly.

4.3.6 Dashboard development

In order to facilitate the construction management process, a Node-RED dashboard was developed to provide real-time monitoring of the operational status of the construction machines throughout the task execution. By subscribing to relevant ROS topics, the system collects and processes data to generate insightful visualizations, offering a clear overview of machine performance. Figure 7 provides a simple example of the Node-RED flow and its corresponding dashboard, which tracks the linear velocity, travel distance and trajectory of the dump truck. By integrating with additional ROS topics, the dashboard is designed to expand its monitoring capabilities to include more construction management-oriented metrics, such as excavated soil quantity for tracking work progress and machine utilization rates for optimizing resource allocation.

5 Conclusion and future improvement

This study demonstrates the development of a low-code-based PIS leveraging Node-RED platform to integrate construction planning with the coordinated operation of multiple construction robots. The PIS simplifies configuration of operational logic for civil engineers through an intuitive drag-and-drop interface, reducing the need for extensive programming knowledge. By ensuring compatibility with ROS middleware through the *rclnodejs* package, the system enables seamless integration with existing ROS-based robotic systems. Additionally, the PIS provides robust data visualization capabilities, offering real-time insights into the operational status of construction activities and robot performance. The prototype demonstrated the coordination of excavator and dump truck operations in an earthwork scenario using ROS topics and actions. This showcases the PIS's ability to unify isolated automated tasks into a collaborative framework that aligns with the overarching construction plan with minimized human intervention.

Future work will expand the system to include a bulldozer for soil leveling and a vibration roller for compaction to allow for a more completed configuration of work coordination among all machines involved in earthwork operations. This will enable the PIS to handle more complex scenarios and edge cases that account for real-world construction scenarios. The dashboard will also be improved by integrating additional relevant ROS topics to monitor more construction management-oriented metrics for quality and quantity control. Additionally, a more advanced physics engine-based simulation environment will be adopted to accurately model soil properties, material interactions and machine dynamics, ensuring robust testing before field deployment. Finally, the PIS is expected to be deployed on actual construction machines to validate its effectiveness and practical applicability in real-world settings.

References

[1] T. Bock. The future of construction automation: Technological disruption and the upcoming ubiquity of robotics. *Automation in Construction*, 59:113–121, 2015. doi:https://doi.org/10.1016/j.autcon.2015.07.022.

[2] T. Elbashbishy and I. H. El-adaway. Skilled worker shortage across key labor-intensive construction trades in union versus nonunion environments. *Journal of Management in Engineering*, 40(1), 2024. doi:10.1061/JMENEA.MEENG-5649.

[3] Ministry of Land, Infrastructure, Transport, and Tourism. Status of the i-Construction initiatives (in Japanese). On-line: `https://www.mlit.go.jp/tec/i-construction/pdf/06.7_kikaku_siryou6.pdf`, Accessed: 12/17/2024.

[4] Ministry of Land, Infrastructure, Transport, and Tourism. i-Construction 2.0 - the automation of construction sites (in Japanese). On-line: `https://www.mlit.go.jp/tec/constplan/content/001738240.pdf`, Accessed: 12/17/2024.

[5] H. Suzuki, G. Yamauchi, D. Endo, and T. Hashimoto. The efforts of the public works research institute to promote the development of autonomous construction technology. *Journal of Control, Measurement, and System Integration*, 61(9):651–655, 2022. doi:10.11499/sicejl.61.651.

[6] T. Akegawa, K. Ohno, S. Kojima, K. Yamada, W. Go, T. Suzuki, S. Kiribayashi, T. Komatsu, N. Miyamoto, T. Suzuki, Y. Shibata, K. Asano, K. Nagatani, and S. Tadokoro. Path planning of the turning back of an autonomous large-scale six-wheeled dump truck for loading/leaving sediment based on backhoe work. In *2022 IEEE/SICE International Symposium on System Integration*, pages 511–518, United States, 2022. doi:10.1109/SII52469.2022.9708879.

[7] K. Kikuchi, K. Nagatani, T. Komatsu, S. Kiribayashi, K. Asano, Y. Shibata, K. Ohno, T. Suzuki, and Y. Hirata. Position estimation using environment-installed laser range finders and traveling control for autonomous surface compression work of vibration roller. *Journal of the Robotics Society of Japan*, 38 (9):872–881, 2020. doi:10.7210/jrsj.38.872.

[8] J. Hinze. *Construction Planning and Scheduling*. Pearson, 2011. ISBN 9780132473989.

[9] S. Lian and F. Matsushita. System design of an integrated controller for coordinated operation of autonomous construction machinery. *Intelligence, Informatics and Infrastructure*, 3(3):10–14, 2022. doi:10.11532/jsceiii.3.3_10.

[10] K. Sano, S. Sato, and R. Nakamura. Demonstration of a cooperative control system for a group of automatic construction machines. *Construction Machinery*, 75(7):46–51, 2023.

[11] F. Matsushita, R. Yajima, K. Ozawa, and K. Nagatani. Development of civil engineering / robotics integrated education program aimed at creating innovation in ict earthwork. *Journal of Japan Society of Civil Engineers (Education)*, 78(1):38–52, 2022. doi:10.2208/jscejeep.78.1_38.

[12] G. Yamauchi, E. Daisuke, H. Suzuki, and T. Hashimoto. Proposal of an open platform for autonomous construction machinery development. In *Proceedings of the 40th International Symposium on Automation and Robotics in Construction*, pages 186–191, India, 2023. doi:10.22260/ISARC2023/0027.

[13] Contributors OpenJS Foundation. Node-RED. On-line: `https://nodered.org`, Accessed: 12/04/2024.

[14] RobotWebTools. rclnodejs package. On-line: `https://github.com/RobotWebTools/rclnodejs`, Accessed: 12/10/2024.

[15] EduArt-Robotik. edu-nodered-ros2-plugin. On-line: `https://github.com/EduArt-Robotik/edu_nodered_ros2_plugin`, Accessed: 12/10/2024.

[16] C. Coulter. *Implementation of the Pure Pursuit Path Tracking Algorithm*. Camegie Mellon University, Pittsburgh, 1992.

From BIM to Autonomous Navigation: Using BIM Models to Enable Autonomous Navigation and Localization in Construction Environments

Jörg Husemann[1], Maximilian Kunz[1], Marcel Suiker[1] and Karsten Berns[1]

[1]University of Kaiserslautern-Landau, Germany

joerg.husemann@cs.rptu.de, maximilian.kunz@cs.rptu.de, suiker@rptu.de, karsten.berns@cs.rptu.de

Abstract -

The automation of construction vehicles and digitization using building information modeling (BIM) are currently two of the most popular research topics in the area of construction. This work combines the two areas by utilizing the information from a BIM model to enable a vehicle to localize and navigate through a building autonomously. A robot equipped with an industry arm should receive marking tasks, which specify a position to navigate to and positions to mark with the robotic arm. These tasks should mimic typical tasks on a construction site, like drilling or cutting holes. The localization within the building is based on LiDAR data, which are compared to the structure of the building, which was received from a BIM model. The path planning uses the BIM model to find a path to the target area. The concept is first tested in a simulation using the Unreal Engine 5, which imports the BIM model and uses simulated LiDAR scanners. Furthermore, the concept is evaluated using a real robot and a building consisting of three rooms and an area of roughly 100 square meters.

Keywords -

autonomous navigation, localization, building information modeling

1 Introduction

Digitalization using Building Information Modeling (BIM) and advancements in automation of construction machines are currently two popular research topics. On the one hand, the advancements with BIM aim to have fully digital documentation of construction sites to monitor the worksite's progress continuously. On the other hand, research in autonomous construction machinery aims to reduce risks for construction workers and mitigate the effects of the shortage of skilled labor. While there is research in both fields, only a few works aim to try to use digital models to improve the processes of autonomous robots. This paper presents an approach in which a robotic platform equipped with a robotic arm should drive through a building imitating a drilling task by marking spots on the inside walls. A BIM model of the environment is used as a knowledge base to localize within the building and to find a path to navigate to given target positions. The localization is done using a LiDAR scanner and is performed in two steps. First, the current room is detected by detecting the outline of the current room and comparing it to the outlines of the rooms in the BIM model. For precise localization, a match between the current scan and a scan of the earlier detected room is compared. By calculating the offset between the two scans and taking the robot's odometry into account, a more precise location within the room is determined. The navigation algorithm uses an elastic band fusion algorithm to navigate the building. As the knowledge base for this algorithm, a grid map is created out of the BIM model. Out of the map, some initial waypoints are calculated using a flood fill algorithm. These points are given to the elastic band fusion algorithm, which coordinates the robot's navigation. The concept is implemented in the Finroc framework[1], which is a framework for intelligent robot[1][2].



Figure 1. Image of Baubot MRS5-1920 from Baubot equipped with an MOTOMAN HC20DTP produced by YASKAWA.

The experiments are executed using the Baubot MRS5-1920[2], a robotic platform equipped with a

[1]https://www.finroc.org/
[2]https://www.baubot.com/

MOTOMAN HC20DTP robotic arm produced by YASKAWA[3](Figure 1). The experiments aim to drive to a given position within the model and then do a marking on a specified position at the wall. Experiments are performed in both simulations and real scenarios. Unreal Engine 5 is used for the simulation. Here, the BIM model is directly imported as a test environment, and a 3D model of the robot is imported. Both navigation and localization are tested using simulated LiDAR data [3]. For the experiment on the real robot, a setup with three rooms and a size of roughly 100 square meters is built. In this environment, the robot is moving through several rooms to a given target location. After reaching the location, the robotic arm marks a set of points on the wall. The position marked are adapted based on the position of the mobile platform. By doing so, some inaccuracies of the navigation can be compensated during the marking process.

The paper is structured in the following way. Section 2 introduces some previous works utilizing BIM models in the scope of robotic tasks and other related aspects as localization. Then in section 3 our developed system is introduced including the localization, path planning algorithm and the navigation. Section 4 shows the experiments in the simulation environment and with the real robot. Finally the work is summarized in the conclusion 5.

2 Related Work

While there are many theoretical concepts on how the information saved in BIM models could be used to support the use of autonomous robots in construction, only very few works use these concepts in real applications. The most common application utilizing a combination of BIM models and autonomous vehicles is the scanning and monitoring of construction progress. In this area, Pietro et al.[4] proposes a concept to monitor the progress of a construction site. They propose a robotic platform with a sensor setup suitable for the application. A general idea of how a map can be created based on a BIM model and which sensors are needed to localize within the environment is sketched. To validate the concept, simulated tests in a simplified environment are run. The same domain is tackled in the approach of Park et al.[5]. They are using a Spot robot to scan the progress in an indoor environment. They utilize BIM models to plan and optimize the scan positions that the robot has to approach. The target is to minimize the number of static scans needed to cover the whole environment, thus minimizing the needed time to scan the whole building. In future works, the scans should also be registered automatically to achieve a fully autonomous process. A similar approach with a different application is presented by Hu et al.[6]. They present an approach to monitor the air quality in an office environment, documenting it inside a BIM model. An autonomous robot drives through an office environment, measuring the air quality. The measurements are incorporated into the BIM model to help identify problems with air conditioning. Their experiments state a precision of roughly 0.10 m using a Velodyne-16 LiDAR scanner mounted on a four-wheeled mobile platform. The information from BIM models can also be utilized to improve various aspects of robotics, such as localization. Zhao et al.[7] use BIM models to solve the wake-up problem in indoor environments. By detecting features of the environment and matching them with the features saved in a BIM model, they manage to find the robot's position in an indoor environment. If there are not enough features at the robot's current position, the environment is explored until the position can be verified. Depending on the environment, the robot had to travel between 15 m and 50 m until finishing the initial positioning. Doing so, they achieve a translational precision of 0.15 m – 0.21 m and a rotational precision of1 1.2° – 2.4°. Another approach in which an autonomous robot is navigating based on a BIM model is presented by Iqbal et al.[8]. The approach uses a small miniature robot to draw floor plans on a sheet of paper. While this work integrated information from a BIM model to an autonomous robot, the application is far from real scenarios since it uses a small robot with a pen to draw two-dimensional floor plans on a sheet of paper. There are also general concepts on how to use BIM models without a specific application. For instance, Chen et al.[9] present a concept of using BIM models to do global path planning within an indoor environment. They use an Industry Foundation Class (IFC) file of the environment as a knowledge base and generate a map from it. Then, they use a modified A* algorithm to plan a path. The experiments are performed in a simplified building simulation and a model of an example robot. While there are already various concepts on how the models can be used in different scenarios, only very few works show how an autonomous robot can participate in the building progress while utilizing the information saved in BIM models.

3 Localization and Navigation Approach

In this work, two major problems of autonomous locomotion are tackled by utilizing the information saved in a BIM model. First, the robot should localize within the building, which includes identifying the current room and then detecting the precise position of the robot within this room. Second, the model is used to create a map of the building, which is the basis for planning a path to the target position and then navigating it while adapting to the current situation within the environment.

[3]https://www.yaskawa.com/

3.1 Localization

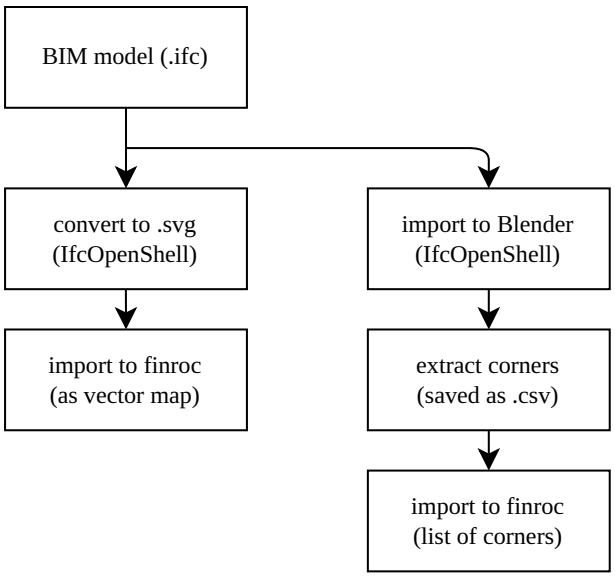


Figure 2. Processing pipeline to use the BIM model for the localization approach.

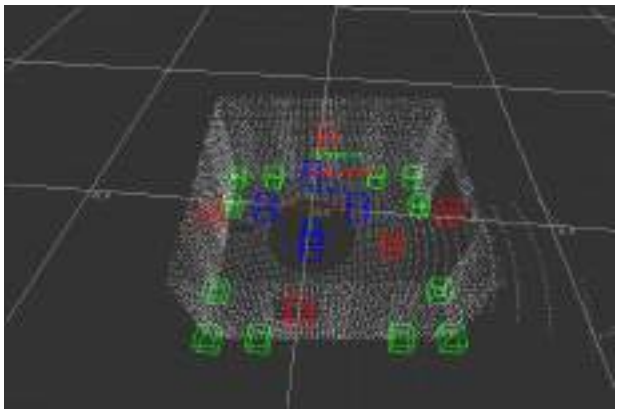


Figure 3. Visualization of feature extraction of a current LiDAR scan in the simulation. The green squares show the corners of the room together with the adjacent walls. The red squares represent the center of gravity of the wall depending on the number of points. The normals of every wall are visualized with the blue squares.

The foundation of our localization approaches is always the BIM model of the building we are currently in. We assume that the 3D model is up to date and contains all necessary components for localization and navigation. As explained in the introduction, we only utilize a LiDAR sensor to detect our environment, thus we are restricted to 3D geometries like walls and corners. We experimented with different pipelines for processing the BIM model stored in a .ifc file to generate a map we can use efficiently, resulting in the pipeline shown in Figure 2. In all of our approaches, we used the 3D modeling software Blender[4] which features an addon for visualizing, creating and working with .ifc files called BlenderBIM[5]. This allows us to use existing Blender features and export functionalities as well as writing our own python scripts inside Blender. Additionally, we made use of the open source IFC toolkit IfcOpenShell[6], e.g. for converting the .ifc file. For our first approach, we used the latter tool to export the BIM model as a .svg file. By doing so, we can interpret the basic geometrical shape of the rooms at a specific height and read the file into our Finroc framework. The downside of losing the Z dimension is neglectable since we extract the data at a specific height, only omitting the wall size. After having access to the building data, we can extract the corners with their corresponding wall normals. After we have created our map of the floor, we do the same for our current scan via the LiDAR sensor. A visualization can be seen in Figure 3. First, we filter the point cloud, then we extract the corners (green) and normal vectors of the wall (blue). As a result, we get a simplified point cloud consisting of corners and normals of the corresponding walls. We match this reduced room representation to our map of all corners using point cloud registration via RANSAC to get the estimated transform. This localization methods works well if we are in a relatively small room where we can see at least two corners. As a fallback, we use basic geometrical features in a similar matter. Similar to the approach by [10] we identify parallel and orthogonal walls since they are invariant to rotational changes and can be found both in the BIM model and the current LiDAR scan quite reliably. For the digital model, our preprocessing steps are similar to the ones we already explained. With the data in .svg format, we extract the walls, convert them into basic lines and extract orthogonal walls (corners) and parallel walls (corridors). In our current scan, we first filter the point cloud, project it to 2D and use the PCL library [11] to extract the walls. In the final step, we compare which features we currently see to what we have stored in the map and try to find the best match. The last influence of our localization is the position estimation based on odometry data we receive from the robot. For short distances, i.e. driving from one room to another, this data was accurate enough for our navigation. After reaching the middle of the room the other localization methods worked as a recalibration for the odometry error.

[4] https://www.blender.org/
[5] https://blender-addons.org/blenderbim-addon/
[6] https://docs.ifcopenshell.org/index.html

3.2 Navigation

The navigation approach consists of two parts. First, a map is generated from the BIM model. Then, the map is used to calculate a preliminary path, which points are used as initial guesses for an elastic band fusion algorithm.

3.2.1 Map Generation from BIM Model

For the navigation, the aim is to use the knowledge saved in the BIM model to create a map of the surroundings. In our scenario, we used the model to create a grid map with the basic structure of the building to be later enhanced with obstacles that can dynamically appear. To do so, some processing steps have to be performed. First, the BIM model available in the Industry Foundation Classes (IFC) format is converted to the Scalable Vector Graphics (SVG) format. The SVG format is usually used for vector graphics and is an efficient way to save geometries. The geometries of the SVG file are rendered to a two-dimensional matrix, which corresponds to the size of the grid map. The geometries' position relative to the BIM model's origin and the map's cell size are considered when mapping the geometries. Finally, the matrix is used to create a grid map in which each element intersecting with one of the geometries is marked as an obstacle in the map, and the other cells are marked as free. Figure 4 shows the BIM model of the test environment and the resulting gird-map. The small cell size for the grid map causes some issues during the calculation of the path planning algorithms. However, it is needed since the passages in the experiment environment are very narrow, considering the large size of the robot. Therefore, the small cell size is used to avoid further reduction of the available space by the map's resolution.

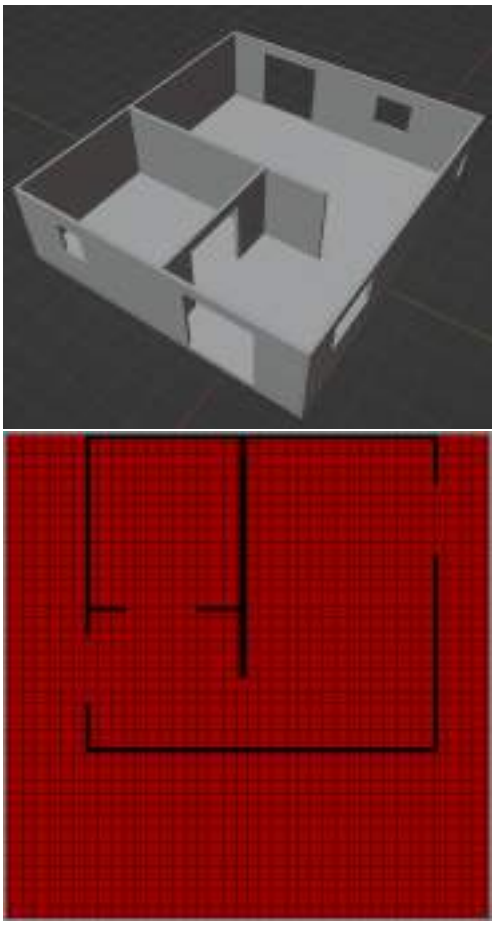


Figure 4. On top the BIM model of the test environment is shown. The image on the bottom shows the resulting grid map with a cell size of 0.06 m. Red indicates free areas and black indicates obstacles.

3.2.2 Elastic Band Fusion

For the actual driving, an elastic band fusion algorithm is implemented [12]. The elastic band method requires a path of target points, which finally leads to the target. Then, the robot always navigates to the next target point. The target points are adapted during navigation based on detected obstacles and the vehicle's position change. As soon as a point is reached, the next point is targeted. This procedure is done until the target is reached. In the case of the marking task, the robot turns to the correct orientation towards the wall after reaching the target point. To calculate the initial points for the algorithm, a flood fill algorithm [13] is used. The algorithm starts at the target pose and assigns to all neighboring cells the distance to the target, which is one in this case. Then, the algorithm iterates to all newly assigned cells and adds the previous cell's distance increased by one to the neighbors that do not have a distance assigned yet. After assigning a distance to the robot's current position, a path can be found by starting at the robot's position and always selecting a neighboring cell with a lower distance than the current one. Theoretically, each of the cells visited when selecting the path can be used as one initial point of the elastic bands algorithm. Since the cell size in our example map is extremely small, the number of points is too high for the algorithm. Therefore, only a subset of the cells is used as the initial target point. Therefore, using a simplified map for the flood fill algorithm, which has a larger cell size, is more meaningful.

The bubbles are created around the calculated target points and are later modified based on the robot's movement. Their size varies based on the critical obstacle, which is the closest obstacle to the bubble 5. The position of each intermediate bubble can be changed based on internal forces between neighboring bubbles, which represent the cohesion strength, and due to an external force, which pushes the point away from the critical obstacle.

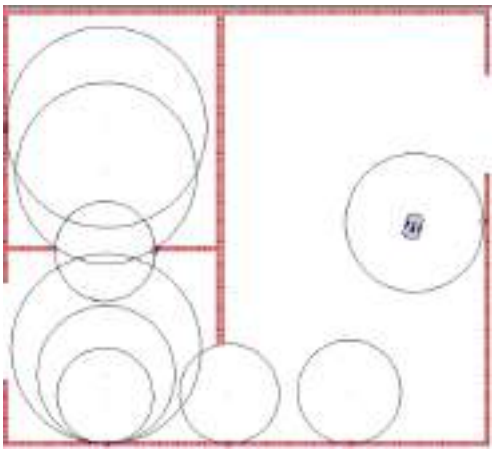


Figure 5. Initial points of the elastic bands fusion. The circles indicate the distance to the next obstacle. If the robots position changes or new obstacles are introduced to the map, the points adjust.

The internal force between two bubbles is defined in the following way

$$\vec{f}_{\text{int},ij} = \alpha_{\text{int}} \cdot \begin{cases} 0 & \text{if } ||\vec{c}_{bi} - \vec{c}_{bj}|| \leq \epsilon \\ \frac{\vec{c}_{bj} - \vec{c}_{bi}}{||\vec{c}_{bi} - \vec{c}_{bj}||} & \text{otherwise} \end{cases} \quad (1)$$

where c_{bi} and c_{bj} are the center points of the two adjacent bubbles. The external force is calculated based on the bubble's distance to the critical obstacle ($\vec{p}_i^*$).

$$\vec{f}_{\text{ext},i} = \alpha_{\text{ext}} \cdot \begin{cases} 0 & \text{if } r_{bi} \leq \epsilon \text{ or } r_{bi} \geq r_{\text{lim}} \\ \frac{r_{\text{lim}} - r_{bi}}{r_{bi}} \cdot (\vec{c}_{bi} - \vec{p}_i) & \text{otherwise} \end{cases} \quad (2)$$

The external force is only larger than zero if the critical obstacle is within the distance threshold r_{lim}. The only exceptions to this behavior are the first bubble, which is fixed to the robot, and the last one, which is fixed at the target position. The shift of a bubble c_{bi} in a discrete time step $t \rightarrow t+1$ is defined in the following way

$$\vec{c}_{bi,t+1} = \vec{c}_{bi,t} + \Delta\vec{c}_{bi} \quad (3)$$

where the delta is calculated based on the two internal forces between the bubble and its neighboring bubbles and the external force on the c_i. The delta is calculated by the sum of the two internal forces and the external force weighted with the value $\alpha_{tot,i}$.

$$\Delta c_{b,i} = \alpha_{\text{tot},i} \cdot \left(f_{\text{int},i,i-1} + f_{\text{int},i,i+1} + f_{\text{ext},i} \right) \quad (4)$$

Whenever the robot approximately reaches the next bubble, the reached bubble is removed, and the next one is chosen as the new target point. This procedure is continued until the bubble representing the target position is reached.

3.3 Marking

The marking task is performed by the robotic arm mounted on the platform. Due to the high precision achievable by the arm, it is responsible for compensating for inaccuracies in the platform's placement. During the marking task, the robot's base is not moving to avoid introducing additional errors. The general process of the marking task is simple. For the movement of the arm, the target positions are first transformed into the platform coordinate system and then into the local coordinate system of the robotics arm. Then, the TCP of the arm is moved to the desired positions. The movement of the arm is done by a function provided by the robotic arm, which is able to move the TCP to a desired local target position. Since only a global target position is given as input, the starting position of the marking progress is calculated to be 0.5 m in front of the marking position. Then, the arm is moved in small steps orthogonally towards the wall. In the target point calculation, the pose of the platform is considered to compensate for inaccuracies in its placement. Since there is no feedback from the arm indicating contact with the wall, a marking tool is used. This tool is equipped with a pen, which is pushed into the tool with a contact. A sensor indicates the contact of the pen. On contact, a new position is set on the arm, which is in front of the wall. As soon as this position is reached, the arm is either moved to the next marking point if another point is reachable from the current position, or the arm is moved back to the driving configuration to enable the platform to drive to the next position.

4 Experiments

To evaluate the concept, tests in simulation and a real test environment are performed. For both experiments, the same BIM model was used. For the simulated tests, the gaming engine Unreal Engine 5 is used. The real experiments are performed in an industry hall where an experimental building with three rooms is built. The setup of the two environments can be seen in Figure 6.

4.1 Hardware Setup

The mobile robot Baubot MRS5-19201 developed by the company Baubot[7] is used for the real experiments. The platform is equipped with the robotic arm MOTOMAN HC20DTP developed by Yaskawa. Additionally, the robot is equipped with an Ouster OS0 64 LiDAR scanner, which is used for perception, and an Intel NUC 13 pro as computing unit. The robot has a width of 1.26 m and a length of 1.88 m excluding the arm. Since the arm has a safety mechanism that is triggered whenever a specific torque

[7] https://www.baubot.com/



Figure 6. Test environment in the Unreal Engine simulation and the real test environment.

threshold is exceeded on the joint, the arm has to be placed in a specific driving position to avoid triggering this mechanism. Due to that, the length of the robot with the arm increases to roughly 3 m.

4.2 Simulated Experiments

A virtual environment based on the model is created for the simulated tests by importing the model into the simulation engine Unreal Engine 5 (Figure 6). Furthermore, a 3D model of the robot is used, which has the same properties as the real machine. Also, the sensors of the real machine are simulated, namely an Ouster OS0 64, the sensor of the marking tool, and the sensors measuring the inertial properties of the robot. In the simulated experiments, the parameters for the elastic band fusion algorithm are verified. Figure 7 shows the setup of the bubbles for a robot, starting in the small room and driving to the large one. The second image of Figure 7 shows the situation after the intermediate room is reached.

The resulting parameters are shown in table 1. The algorithm is also tested in larger environments proving the functionality.

Parameter	Value
r_{lim}	2.00
r_{robot}	0.90
α_{Int}	0.0001
α_{Ext}	0.0001
ϵ	0.050

Table 1. Parameter setup for the elastic band fusion algorithm

4.3 Experiments in Real Test Environment

In the experiments in the real building, the robot is placed in the small room, as also shown in the simulated experiments. Then, a position to drive to in the large room is selected, and a point to mark from this position. Then, the robot is enabled and starts driving to the target position. The initial localization is done using the LiDAR scanner. While driving mainly the odometry is used while resetting it from time to time with a new LiDAR position. The localization is done this way because the odometry is much more stable for short distances. In some places, especially while driving through a door, the LiDAR localization has large errors. Here, the odometry proved to be the best localization method. While driving through the doors, the navigation algorithm also tracks the robot's current room. By that, it is only needed to detect the current room in the beginning and track the current room afterward using the driven trajectory. In the real experiments, the elastic band failed in some cases. Those occurred due to the very narrow passages the robot has to drive. Because of that, fixed bubbles are generated within the middle of the doors and in front of the door in both connected rooms. By that, the robot was guaranteed to be approaching the door at an angle that enabled it to pass through. The most critical passage is shown in Figure 8. It can be seen that the margin for error while driving to the door is minimal, mainly because of the large length of the robot and the arm in the driving position. This issue only appeared in the shown passage. In the other cases, where the door is in the middle of the wall, the navigation is much easier since there is more space to maneuver the robot to pass through.

Figure 9 shows the two most relevant states of the marking process. The left image shows the inital position of the marker. Then the arm slowly moves perpendicularly towards the wall. Since there is very little room for the marker to move into the tool this movement was performed very slowly. As soon as the marker has contact to the wall as seen in the right image, the arm moves back to the driving position and the task is complete.

For obvious reasons, the main property of evaluating such a marking task is to check the accuracy of the marking position. During the process, there are multiple sources of

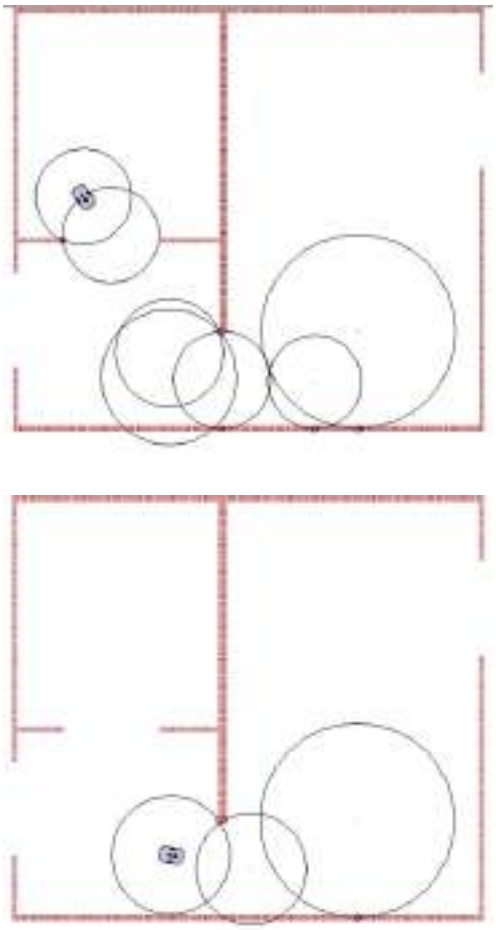


Figure 7. Bubbles of the elastic band fusion algorithm at the start and after driving half of the distance to the target position.



Figure 8. Robot driving through narrow passage in test building.

error, of which some can be optimized by the robotic platform, and some are introduced by the BIM model. During the test run, the precision of the marking varied between 0.2 m and 0.5 m. During the experiments, we found several error sources which have room for improvement. First, adjusting the arm movement to the possible imprecise positioning towards the wall introduced an error. This error grows with the difference between the desired perpendicular position and the actual position toward the wall. This can be either optimized by improving the precision of the driving of the platform or by improving the adaptation of the marking trajectory. The localization error introduces another issue. The quality of the LiDAR localization depends on the visual features as described in section 3. If the marking position is in an area with few features, the error increases. Another aspect that could improve the accuracies mentioned before would be to calibrate the sensors and the arm's position precisely. Currently, the positions are measured manually, which reduces overall precision. Those are the main error sources during our experiments. There are some smaller sources like the repetition accuracy of the arm, which can be up to 0.02 m, and the accuracy of the BIM model compared to the real building. While the accuracy is not yet suitable for real construction applications, improving the previously mentioned points could enable the robot to achieve much higher accuracy.

5 Conclusion

In this paper, an approach to using information from a BIM model to develop an autonomous robot for construction tasks is presented. The proposed concept uses the information to implement a localization approach based on point clouds and a planning and navigation approach. The aim is to navigate through a building while performing marking tasks with the robot's arm. The localization approach detects features of the current room to first identify the room the robot is placed in and then uses the features to precisely position itself within the room. For path planning, a grid map is created using the BIM model. Then, an elastic band fusion algorithm is implemented. A flood fill algorithm is used to find a suitable path to the target point. Then, this path is adapted based on the robot's current position and the LiDAR scanner's measured obstacles. Experiments in both a simulation and a real scenario evaluated the concept. The real tests are executed by a mobile



Figure 9. Robotic arm during the marking process. The left image shows the arm after moving into the initial position. The right image shows the arm after a contact before moving back to the initial position.

platform equipped with an arm. For these tests, a building with three rooms is constructed, and the robot navigates through this building to reach a target position. Then, it marks a given position with the arm. The experiments show that using the BIM model is sufficient for the localization and navigation tasks. Nevertheless, there is still room for improvement, especially regarding accuracy.

Acknowledgement

This work is part of the HumanTech project and has received funding from the European Union under Grant Agreement No. 101058236. HumanTech is a three-year project that started on the 1st June 2022 and terminates on the 31st May 2025.

References

[1] Max Reichardt, Tobias Föhst, and Karsten Berns. Introducing finroc: A convenient real-time framework for robotics based on a systematic design approach. *Robotics Research Lab, Department of Computer Science, University of Kaiserslautern, Kaiserslautern, Germany, Technical Report*, 2012.

[2] Max Reichardt, S Schütz, and Karsten Berns. One fits more-on highly modular quality-driven design of robotic frameworks and middleware. *Journal of Software Engineering for Robotics (JOSER)*, 8:141–153, 2017.

[3] Manuel Philipp Vogel, Maximilian Kunz, Eike Gassen, and Karsten Berns. Lidar-geil: Lidar gpu exploitation in lightsimulations. In *SIMULTECH*, pages 71–81, 2023.

[4] Samuel A Prieto, B Garcia de Soto, and Antonio Adan. A methodology to monitor construction progress using autonomous robots. In *Proceedings of the 37th International Symposium on Automation and Robotics in Construction (ISARC)*, pages 1515–1522. International Association for Automation and Robotics in Construction (IAARC), 2020.

[5] Sangyoon Park, Sanghyun Yoon, Sungha Ju, and Joon Heo. Bim-based scan planning for scanning with a quadruped walking robot. *Automation in Construction*, 152:104911, 2023.

[6] Xi Hu and Rayan H Assaad. A bim-enabled digital twin framework for real-time indoor environment monitoring and visualization by integrating autonomous robotics, lidar-based 3d mobile mapping, iot sensing, and indoor positioning technologies. *Journal of Building Engineering*, 86:108901, 2024.

[7] Xinge Zhao and Chien Chern Cheah. Bim-based indoor mobile robot initialization for construction automation using object detection. *Automation in Construction*, 146:104647, 2023.

[8] Fahad Iqbal, Shiraz Ahmed, Fayiz Amin, Siddra Qayyum, and Fahim Ullah. Integrating bim–iot and autonomous mobile robots for construction site layout printing. *Buildings*, 13(9):2212, 2023.

[9] Zhengyi Chen, Keyu Chen, Changhao Song, Xiao Zhang, Jack CP Cheng, and Dezhi Li. Global path planning based on bim and physics engine for ugvs in indoor environments. *Automation in Construction*, 139:104263, 2022.

[10] Wolfgang D Rencken, Wendlin Feiten, and Raoul Zöllner. Relocalisation by partial map matching. In *Sensor Based Intelligent Robots: International Workshop Dagstuhl Castle, Germany, September 28-October 2, 1998. Selected Papers*, pages 21–35. Springer, 1999.

[11] Radu Bogdan Rusu and Steve Cousins. 3D is here: Point Cloud Library (PCL). In *IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, China, May 9-13 2011.

[12] Sean Quinlan and Oussama Khatib. Elastic bands: Connecting path planning and control. In *[1993] Proceedings IEEE International Conference on Robotics and Automation*, pages 802–807. IEEE, 1993.

[13] SV Burtsev and Ye P Kuzmin. An efficient flood-filling algorithm. *Computers & graphics*, 17(5):549–561, 1993.

Case study: Robotic Underwater Inspection of Concrete Dam Face Prior to Refurbishment

Frédéric Nadeau[1], Marion Nourry[1], Guillaume Boivin[1], Eric Boudreault[1], and Carmin Drolet[2]

[1]Hydro-Quebec, 1800 boul. Lionel-Boulet, Varennes, QC, Canada
[2]Hydro-Quebec, 855 Rue Sainte-Catherine E, Montréal, QC, Canada
nadeau.frederic2@hydroquebec.com, nourry.marion@hydroquebec.com, boivin.guillaume@hydroquebec.com, boudreault.eric@hydroquebec.com, drolet.carmin@hydroquebec.com

Abstract –

This case study examines the use of underwater robotic systems to assess cofferdam support faces prior to refurbishment. The inspection process employs a suite of specialized equipment, including a surface preparation robot, geometry measurement tools, and non-destructive testing (NDT) technologies for concrete delamination detection. The integration of these technologies enhances the accuracy and efficiency of structural assessments, ensuring comprehensive data collection and analysis while eliminating health and safety constraints for workers. The findings demonstrate the potential of robotic inspection systems to significantly improve the safety and effectiveness of dam infrastructure maintenance.

Keywords –

Concrete Delamination; Geometry Measurement; NDT; Robot; Underwater; Divers; Dams; Refurbishment

1 Introduction

Hydro-Québec has many hydroelectric generating stations and spillways, some of which were built more than 100 years ago. Our case study focuses on the La Gabelle dam (Figure 1) that celebrated its 100th anniversary in 2024. It has reached a point in its life cycle where some maintenances must be planned. To target the right intervention to restore the life of mechanical equipment and concrete structures, it is important to have quality data that is sufficiently representative of their condition.

In the case of the La Gabelle dam, spillway gates must be refurbished, thus requiring cofferdams to give dry access to workers. Before the cofferdams can be installed, piers' support faces must be inspected to ensure that the cofferdams will be adequately supported and offer appropriate watertightness.

Historically, divers conducted these inspections manually, leading to numerous operational and qualitative challenges. Among these, one can consider the health and safety risks for the divers associated with the dam's conditions (cracks, water infiltrations, etc.) and water currents. Additionally, significant doubt and uncertainty arise regarding the reliability of the measurements due to the following reasons:

- For measuring surface roughness, inspection is done visually using a ruler in turbid and turbulent water.
- For identifying delamination of concrete, inspection is done by manually hitting the concrete surface with a hammer and listening to the resulting sound wave through diving helmet.

To overcome these challenges, Hydro-Québec's Research Institute has developed robotic inspection solutions dedicated to the inspection of hydraulic dams and their features. This paper presents how these robotic solutions were used at La Gabelle to provide civil engineers with reliable measurements while avoiding the health and safety constraints associated with deploying divers in the water.



Figure 1. La Gabelle dam, upstream view of the concrete piers and inspection areas.

2 Surface Preparation

Before conducting laser measurement of geometry and concrete hammering, surface preparation must be conducted. The main purpose of this operation is to remove algae, mussels, and concrete parts weakly clinging on the surface. As shown in Figure 2, this is achieved using two cavitation cleaning domes mounted on a trolley, maneuvered by a multifunctional platform. The platform is shown in Figure 3. Rotating nozzles inside the domes create a strong suction force, ensuring the system remains in contact with the surface. The system can clean rough deposits like barnacles and mussels in up to 17 seconds per square meter, and for algae, it can clean a strip 0.89 meters wide at a speed of 0.5 meters per second.

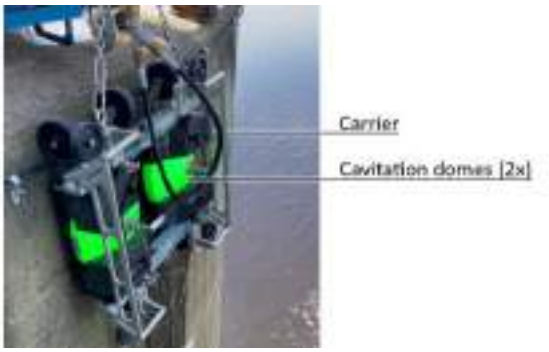


Figure 2. Cavitation cleaning system

The cavitation cleaning process involves water passing through specially shaped nozzles, generating millions of steam bubbles. These bubbles collapse upon hitting the surface, creating tiny pressure waves that remove scale and biofouling such as barnacles, algae, and mussels. This method is highly efficient, easy to integrate onto a carrier, and does not release any residue from the material itself.

Compared to other underwater cleaning technologies, cavitation cleaning is highly effective [1]. Ultrasonic and UV methods have limited capacity and slow efficiency in removing biofouling. Cavitation, bush rotary brushes and water jets achieves over 90% bio-fouling removal. Cavitation cleaning stands out with an efficiency range of 100-1500 m²/h, making it one of the most efficient methods available.

3 Geometry Measurements

Geometry measurements of the contact surfaces are required to evaluate the proper support and sealing of the cofferdams. Flatness and parallelism of the contacts will insure its proper operation. Geometry deviations may result from concrete swelling (e.g. alkali-aggregate reaction) or surface degradation. Erosion and spalling are often found within the tidal zone, where ice thrusts apply.

To assess the geometry, a survey team measured the contact surfaces on the dam, above and below the waterline. Surface digitization technologies are preferred. The resulting 3D point cloud data can be processed to generate vertical cross-sections and heat maps, effectively highlighting any deviation. Offering precision non-contact measurements, a Leica ATS-600 direct scanning laser tracker was used for measurement above the waterline.

Detecting damage on submerged structural elements is often challenging, especially when precision is required. To collect engineering-level three-dimensional measurements underwater, Hydro-Quebec developed the WireScan [2] [3], laser scanner for vertical surfaces that can achieve a ± 1mm accuracy in still water. The system is portable and deployed from the deck, eliminating the need for divers (Figure 3).



Figure 3. WireScan measurements on a spillway. Underwater scanning (top right). Modular multifunctional platform (left).

WireScan uses two plumb lines as vertical references, maintaining high precision over the range (10 meters at La Gabelle). Supported by its tether, the laser camera slides vertically along wires stretched by weights. Each scan covers approximately 400 mm wide and is positioned in the 3D space with reference spheres measured with the ATS-600.

As depicted in Figure 4**Erreur ! Source du renvoi introuvable.**, results show a large coverage of the top of both piers (with the ATS-600). Below the waterline, the scans focus on the narrow sealing surfaces. Surface damage is visible on the tidal zone just below the waterline.

From the 3D data, it is possible to measure:

- Verticality of the lifting plane
- Flatness defects on sealing surfaces
- Pier-to-pier width at the contact points

The results are shown in Figure 5. The left pier is deviated by +/-15mm around the vertical, while the right pier is deviated up to 60mm to the right. The

discontinuity on the left pier corresponds to the horizontal joint shown on the 3D scan.

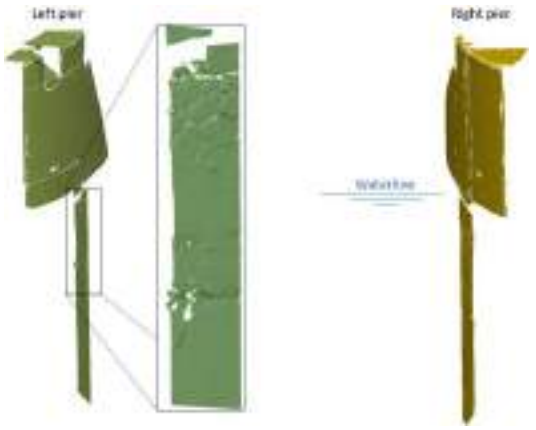


Figure 4. 3D scans of the cofferdam contact surfaces, La Gabelle.

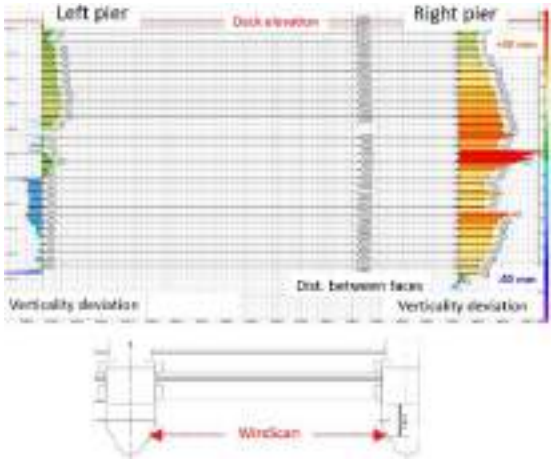


Figure 5. Measurement of the sealing surfaces, elevation view, La Gabelle.

4 Concrete Delamination Detection

To conduct delamination detection, we used an in house developed trolley, see Figure 6, equipped with a motorized hammer to perform impacts and with two hydrophones to acquire the resulting acoustic waves. Striking force consistency is improved compared to traditional methods by mounting the hammer on a shaft axially coupled to a servo motor. As the hammer rotates around the axis, it provides two impact points, 30 cm apart, enhancing inspection coverage. This setup allows precise control of the striking force and additionally offers insights such as the impact angle and the restitution factor. The impact angle can be used to detect surface imperfections, while the restitution force can help improve delamination detection and surface condition. Additional equipment has been mounted on the trolley: a camera and a line laser, allowing for detailed visual inspections and providing valuable information about the concrete's texture, and a vortex suction cup that keeps the trolley firmly attached to the concrete surface it is inspecting. The multifunctional platform mounted on top of the dam, shown in Figure 3, is equipped with a winch, and allows precise positioning of the inspection trolley over the concrete surface.



Figure 6. Hammering Trolley.

Hydrophones provide sound pressure resulting (in Pascals) of each impact. From there, we can calculate the sound exposure (see Equation (1)) where *p(t)* is the sound pressure reading from the hydrophone in respect to time t in seconds. Using Equation (2) & (3), we obtain the sound exposure level in a dB scale, sometimes referred as SEL_{SS} [4].

$$E = \int_{t_0}^{t_1} p(t)^2 \, dt \tag{1}$$

$$SEL_{SS} = 10 \log_{10} \frac{E}{E_0} \tag{2}$$

$$E_0 = 1 \cdot 10^{-6} \tag{3}$$

Sound exposure levels for each single strike, SEL_{SS}, are computed for each strike's respective duration, where $\Delta t = t_1 - t_0$ is approximately 80ms. In-air measurement would use a different reference level. E_0 and E are given in $Pa^2 \cdot s$.

System has been validated on a concrete block manufactured with known defects. We were able to determine with great accuracy the location of each defect embedded within the block. A software has been developed to map all the readings from an inspection into a heat map, see Figure 7. Our software enables the playback of each impact individually allowing reviewer to determine if a specific section if subject to

delamination. Further improvements could be envisioned such as use of Impact-Echo method [5]. However, surface roughness could pose a challenge in producing high-frequency compression wave.

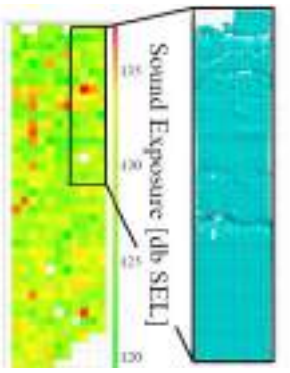


Figure 7. Hammer strike (left) vs WireScan (right).

When compared to measurements obtained from the WireScan, both inspection tools were able to find areas where wears and tears are present. The variation in sound exposure across the inspected surface (+/- 5 %) is not significant enough to correlate with delamination.

5 Conclusion

The adoption of robots in construction projects is anticipated to rise in the coming years, driven by their numerous benefits, including enhanced worker safety, consistency, and precision, among others. In this case study, we demonstrated how robotized inspection solutions developed at Hydro-Québec's Research Institute were successfully used to obtain surface roughness measurements and to assert presence of concrete delamination on concrete piers at the La Gabelle dam prior to the installation of cofferdams. The benefits expected from the use of robots were confirmed accordingly.

As a starter, all the operational constraints related to health and safety concerns over divers conducting inspections were eliminated. This should be sufficient to justify the robotic inspection project, as the health and safety constraints for divers are so restrictive that, at times, inspection phases of piers are completely abandoned, leading to cofferdam installation work being planned without data on the actual condition of the concrete, which has historically resulted in significant cost overruns.

Additionally, the inspection data that was delivered to the civil engineering team at Hydro-Québec provided unpreceded analysis capabilities. As compared with traditional inspection methods, which only provide maps with approximate areas that require attention, we provided millimeter precision surface roughness measurements, quantitative impact maps that reflect the state of concrete delamination, audio files for playback of all the impacts, and inspection videos. This valuable data will not only enable civil engineers to take more accurate decisions, reducing the risk of cost overruns, but also allow them to compare inspection data of different dams and at different moments of their lifespan, which will inevitably lead to a better understanding of concrete degradation mechanisms.

While the robotic inspection systems demonstrated significant advantages in terms of safety, precision, and data quality, there are still some limitations to consider. One major limitation is the requirement for easy and large access to the top of the structure directly above the inspection area. This can be challenging in certain dam configurations or when overhangs are present, which may obstruct the deployment of the inspection equipment. Additionally, the current system's effectiveness can be hindered by extreme underwater conditions, such as high turbidity or strong currents, which may affect the accuracy of the measurements and the stability of the robotic platforms. Future developments should focus on enhancing the adaptability of the robotic systems to various structural configurations and improving their performance in diverse underwater environments.

6 References

[1] Houssam, A., Dandan, K., Ananiez, A. and Kalaykov, I. Underwater robotics: Surface Cleaning Technics, Adhesion and Locomotion Systems, International. *Journal of Advanced Robotic Systems*, 2015.

[2] Miralles, F., Boivin, G., *A New Compact Solution for the Rapid Measurement of Underwater Gate Guides*, Proceedings of the International Conference on Applied Robotics for the Power Industry, 2014.

[3] International Hydropower Association., *Extending the life of hydropower plants with 3D laser scanning.* On-line: https://www.hydropower.org/case-study/underwater-3d-scanning-for-hydropower, Accessed: 16/12/2024.

[4] BOEM. *Introduction to Sound and Acoustic Assessment.* On-line: https://www.boem.gov/sites/default/files/documents/renewable-energy/NY%20Bight_DraftPEIS_AppJ_IntroSoundAcousticAssessment_508.pdf, Accessed: 16/12/2024.

[5] M. Sansalone and N.J. Carino, *Impact-echo: A method for flaw detection in concrete using transient stress waves.* National Bureau of Standards, 1986, NBSIR 86-3452.

Robotized tower crane: Simulation and high-level action planning system

Mohammad Reza Kolani[1], Stavros Nousias[1] and André Borrmann[1]

[1]Chair of Computing in Civil and Building Engineering, TUM Georg Nemetschek Institute
Technical University of Munich, Germany

mohammadreza.kolani@tum.de, stavros.nousias@tum.de, andre.borrmann@tum.de

Abstract -

Tower cranes are essential in the construction industry and play a critical role in lifting and transporting materials. Despite significant advancements in the construction industry, a comprehensive framework for simulating tower crane operations and an Actions Management Node (AMN) to coordinate these operations effectively are currently lacking. This research presents a comprehensive framework that integrates a full-scale tower crane model with the Robot Operating System (ROS) and simulates its functionalities using a physics-based simulation. The framework not only facilitates operational-level control of a real-scale tower crane, but also incorporates the action management node that abstracts the complexity of crane functionalities. This abstraction enables users to provide a high-level control plan that the action manager executes with precise and efficient sequences of actions to lift and transport materials. The proposed framework, along with the action management node, is evaluated through a physics-based simulation in two different scenarios where construction components are transported from initial positions to designated target locations. The results demonstrate the framework's capability to effectively handle all crane operations, demonstrating its scalability and strong potential for real-world applications in tower crane automation.

Keywords -

Robotic tower crane; Tower crane simulation; Automated transporting system; Action planning; Construction robotics

1 Introduction

Rapid industrialization and urbanization drive a growing demand for housing, factories, and infrastructure as productivity in the construction industry remains stagnant [1, 2]. Tower cranes (TCs) are the most important equipment on construction sites and play a critical role in lifting and transporting structural materials such as steel beams and columns and concrete segments. In addition, TCs are used for the placement of prefabricated elements and heavy equipment, which is important for the assembly of structural frameworks and the installation of machinery. As a result, the overall productivity of construction projects is significantly influenced by the effective planning of crane operations. In addition, in industrial settings such as power plants, TCs are utilized to assemble heavy machinery and shipyards to transport containers and large equipment.

In practical scenarios, operations managers and planners often rely on their experience and subjective judgment to design lift plans, utilizing only 2D drawings or static 3D models [3]. Typically, planners evaluate crane accessibility to material pick-up and placement points and provide approximate estimations of lift task durations. However, these evaluations fail to consider the real movements and dynamics of the crane under operational settings on site [3]. These static planning methods can lead to inefficient and impractical plans for crane operations, compromising the effectiveness of the entire lifting process [4].

Therefore, the efficient operation of TCs on construction and logistics sites plays a critical role in reducing the time required for each load while maintaining safety standards for workers and operators. In this context, the robotization of TC and automation of its operations not only significantly enhance operational efficiency but also improve overall productivity, since the overall productivity of construction projects is heavily affected by the proper planning and operations of a TC [5]. Furthermore, this automation reduces human errors, operational risks, and enables precise load handling. Therefore, the capabilities of TCs make them crucial across various sectors, particularly in the construction industry, to improve productivity and efficiency, and robotization of TCs is an important step toward the envisioned robotized construction site [6].

Numerous studies have explored various aspects of tower cranes in planning, operation, control strategies, and automation techniques. In particular, several research studies investigate various aspects of tower crane controllers, highlighting advancements in control strategies [7–10]. A sliding mode control scheme was developed to enhance the control performance of 5-DOF TCs by precisely driving the trolley, jib, and hook to their target positions while suppressing payload swings caused by unknown payload masses, friction, and wind disturbances,

without relying on linearization techniques [10]. Another approach employed data-driven methods to reduce path tracking and positioning errors in automated TCs without permanently integrating additional sensors. This method used a regression model to predict path errors, enabling the computation of a compensatory hook path that ensures the measured trajectory closely aligns with the desired one [1]. Additionally, an offline trajectory planning algorithm was proposed to guide payloads along multiple straight connection lines formed by waypoints, with smooth transitions calculated at their intersections, further enhancing path planning precision [11].

In particular, significant research has been achieved in the field of robotic TCs to automate and improve crane operations [5, 12–16]. A motion compensation algorithm was developed for multi-step speed control to achieve the exact displacement based on dynamically optimizing the time duration at each planned velocity [13]. A re-planning module developed constitutes of a Decision Support System (DSS) and a Path Re-planner (PRP) based on a decision-making algorithm to enable near real-time collision-free path optimization [16]. To address operational challenges, an automated lifting path tracking system was developed to eliminate the 'blind spot' issue in tower crane operations [17]. This system incorporated laser devices to assess its feasibility across varying environmental conditions and dynamic payload swinging scenarios. Furthermore, [15] proposed an automated lift path planning method for tower cranes based on point clouds and utilizes an octree-based sampling strategy to generate a roadmap. However, existing planning approaches often do not consider lifting time considerations and face frequent infeasibility in planned paths. To overcome these limitations, [5] introduced a reinforcement learning (RL) algorithm that replicates human decision-making processes by determining actions based on visual inputs. This system employs a lidar sensor to detect surrounding obstacles, thereby preventing collisions and enhancing operational safety, like a human operator. Additionally, [12] developed time-varying linear quadratic regulators (LQR) for trolley and jib control employing a proportional-integral-derivative (PID) method. The proposed controllers were evaluated through the ROS framework and Gazebo simulator. In addition, a robotic truck crane system (RTCS) was developed in a simulation environment and was validated by a small-scale prototype of a truck crane [14].

To enhance productivity in the construction industry, a crucial step is the development of robotized tower cranes capable of transporting objects without the need for expert operators. However, achieving this level of automation requires prior simulation due to the safety risks and costs associated with real-world testing. Therefore, a framework that simulates TCs's operations in realistic scenarios in construction and logistic sites is the first step towards the automation of TCs. This framework should then provide various construction site scenarios and transportation tasks. These can provide better management of TCs and provide precise operation time per load, which can be used for real implementation. The simulation and automation of TCs have the potential to improve efficiency, ensure safer operations, and maximize productivity, especially in construction environments where multiple TCs are in use.

This study presents a comprehensive framework that emphasizes action handling and simplifies the complexities of operational-level commands. It achieves this through the development of an action management node, which translates user-provided high-level plans into precise instructions for the tower crane's controller, enabling the seamless execution of desired motions. The framework, built on the ROS [18], leverages its capabilities for robotic automation and machinery control. ROS provides the best middleware that enables seamless integration of various nodes, allowing for real-time communication between modules and efficient message parsing. To validate the proposed action management node and framework, we integrated them with the Gazebo simulator [19] and conducted simulations using various scenarios involving different objects. The results demonstrate the effectiveness of the framework in simulating tower crane operations with precision and reliability.

The rest of the paper is organized as follows: A detailed description of the tower crane model used in this research is presented in section 2. Following that, Sections 3 and 4 outline the design and implementation of the Action Management Node and the simulation framework, respectively. Section 5 presents the results, demonstrating the effectiveness of the proposed framework. Lastly, Sections 6 and 7 deliver a discussion of the results and conclude the study.

2 Tower Crane Modelling

A tower crane comprises several key components, including the ground base segment, body segment, climbing segment, cabin and motors, apex, counterweight segment, jib base, main and last segments, trolley, cables, and hook, as shown in Figure 1. These components collectively determine the operational workspace of the crane. Adjustments to the working space can be achieved by adding or removing body and jib segments, enabling customization to meet specific site requirements.

Tower cranes, particularly hammerhead tower cranes, are often modeled as five-degree-of-freedom (5-DOF) nonlinear dynamic systems [10]. Three DOFs are actuated—jib rotation, trolley movement, and hoisting cable operation—while two are unactuated, representing radial and tangential swinging. The generalized coordinates for the actuated DOFs are γ (jib rotation), x_t (trolley posi-

tion), and l_h (hoisting cable length). For the unactuated DOFs, the generalized coordinates are ϕ (tangential swing) and θ (radial swing). Figure 1 provides a schematic representation of a robotized tower crane, highlighting its structural components, jib segments, and trolley-hook assembly, which are responsible for its core functionalities.

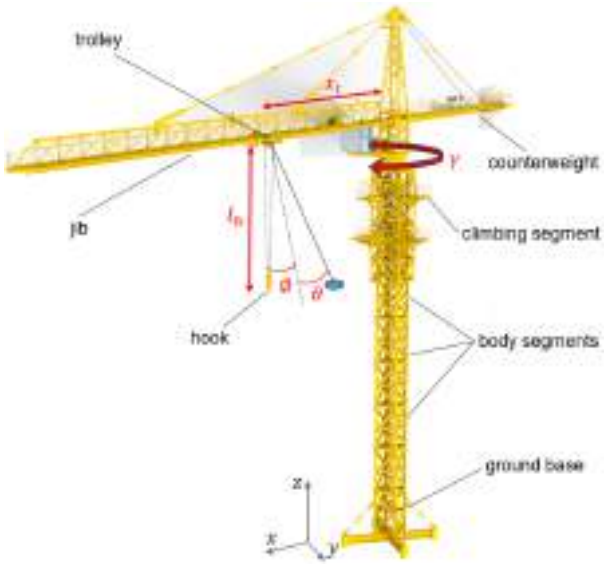


Figure 1. Schematic representation of the tower crane investigated in this study. The red color shows the DOFs of the tower crane.

This study focuses exclusively on the three actuated DOFs, treating the tower crane as a 3-DOF robotic arm. The unactuated DOFs, representing pendulum-like swings, are beyond the scope of this work. The jib, similar to a robotic arm, is mounted on the structural component (known as a mast) and driven by an electric motor coupled with a gear mechanism, allowing rotational motion. The trolley, which moves radially along the jib, is controlled by a cable-driven mechanism powered by a second electric motor. Finally, a pulley system operated by a third electric motor facilitates the vertical hoisting of loads.

The complete action functionalities of the three actuated DOFs are implemented and simulated in this study as part of a robotized tower crane model. This approach underscores the crane's operational capabilities while abstracting complexities related to the unactuated DOFs, ensuring a focused investigation of its robotic aspects.

The initial models of TCs are typically designed in formats such as .prt (Siemens NX) or .SLDPRT (SolidWorks). To integrate these models into robotics frameworks, they must be converted into formats compatible with simulation environments. This process involves modifying CAD files, converting their formats, and generating simulation-supported description files. These description files serve not only to define the crane's geometry but also to provide physical properties, including mass, inertia, joint limits, connections, and kinematics. In addition, modifying CAD files allows for minimizing the computational complexity of the simulation while preserving the operational and behavioral accuracy of the tower crane.

These steps are crucial, as they allow the ROS framework to effectively interpret the connectivity and relative motion of the crane's components, ensuring accurate simulation and control of the robotized tower crane.

3 Action Management Node

The main tower crane operations are gripping the load, lifting, transporting from one location to another, lowering, and releasing [12]. The Action Management Node (AMN) is responsible for handling all these operations by receiving high-level plans from users and converting them to low-level commands that can be used by the TC's planner and controller. The AMN incorporates all the functionalities of a TC within completely separate actions, enabling the transporting of objects from their initial positions to desired target locations. The AMN processes input data such as the object's name and the required sequence of operations, which are then transmitted to the planning and control nodes for the execution.

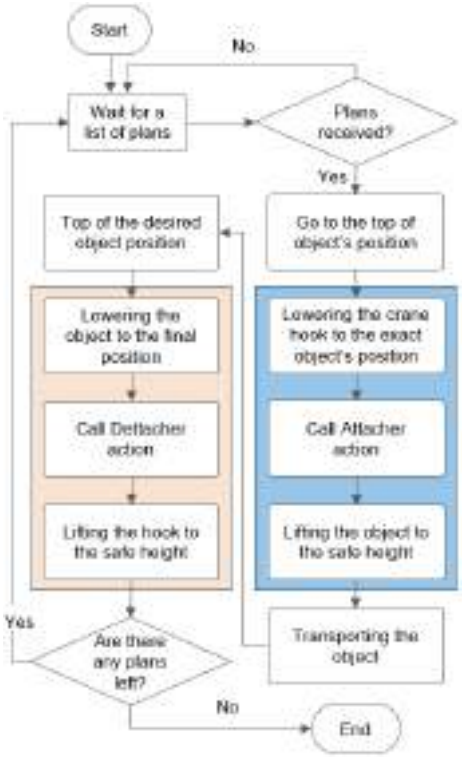


Figure 2. High-level flowchart illustrating the entire process of Action Management Node

The developed node provides continuous feedback to the user, enabling real-time tracking of the crane's status during execution. After completing the operation, the AMN sends a final status result to the user. Notably, the user can interrupt and cancel an operation at any stage of execution, ensuring flexibility and safety.

The workflow for a complete operation, from transporting an object from its initial position to its designated target location, is depicted in Figure 2. This process involves a series of sequential actions organized into three main phases: transportation, lifting (blue box), and placement (orange box). Each phase includes specific actions to ensure the precise and safe handling of materials during crane operations. The lowering action is for operational accuracy, while the lifting is critical for maintaining safety. The AMN incorporates a dedicated action to raise the hook to a predefined safe height, thereby minimizing the risk of collisions during transportation.

Given the assumption of a robotized tower crane, two additional actions, Attacher and Detacher, are introduced to perform tasks traditionally handled by human operators, securing the object to the hook and releasing it at the target location. These actions enable the system to autonomously handle objects, ensuring the same operation. When the tower crane hook reaches the object, the AMN executes the attach plan, securely attaching the object to the hook. At the target location, upon receiving the detach plan, the AMN releases the object. This automated workflow ensures efficient and reliable execution of tasks in an autonomous robotic tower crane system.

4 Simulation Framework

To simulate and control the tower crane, several core modules are essential, including the TC modeling, planning, control, and simulation, each composed of multiple sub-modules, as shown in Figure 3. These modules manage the entire process, from crane modeling to executing motion plans within the simulation environment. ROS [18], provides the best middleware that enables seamless integration of various nodes, allowing for real-time communication between modules and efficient message parsing. Furthermore, ROS's modular architecture supports scalability, enabling the addition of new functionalities or modifications to existing modules without disrupting the overall system's integrity. This enables continuous enhancement of the framework, ensuring its adaptability to various operational scenarios and evolving requirements in the construction industry.

Once the crane model is loaded into the system, the state publisher node continuously broadcasts the joint states and transformations of the crane, ensuring that its positions and orientations are consistently updated. This allows other modules, such as motion planning and control, to access

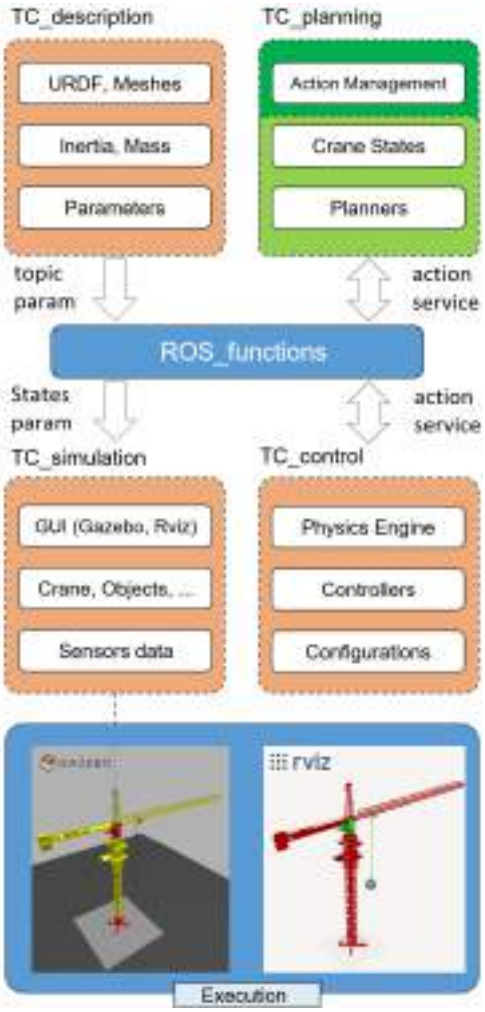


Figure 3. Flow diagram of the simulation framework

up-to-date kinematic data of the crane's current state.

A motion planning node was developed for transportation operations using MoveIt! [20], considering the crane's configurations and properties to ensure accurate and reliable motion planning. This node generates a trajectory consisting of a sequence of joint positions, defining the crane's motion from its initial state to the desired target position. Each joint position in the trajectory is associated with a specific timestamp, defining a time-parameterized trajectory. Once this trajectory is generated, a controller is required to drive the system's joint states to follow the planned positions at the corresponding timestamps.

A controller node was developed by adapting and enhancing existing control libraries originally designed for robotic manipulators. The control node ensures smooth and precise crane movements by implementing acceleration and deceleration profiles, optimizing motion dynamics for efficiency and accuracy. It interfaces with Gazebo

[19], a physics-based simulation environment, to simulate and execute planned trajectories. The controller operates as a closed-loop system, generating control commands for the crane model while continuously receiving real-time feedback to ensure accurate trajectory tracking and execution.

The action management node integrates this iterative feedback mechanism, enabling it to monitor the execution process and send the results to the user. This feedback loop continues until the controller either successfully completes the planned actions or sends an error message for the AMN. By incorporating this functionality, the AMN equips the tower crane with advanced action planning capabilities tailored to various construction scenarios.

5 Results

In this section, the simulation results of the tower crane operations using a physics-based simulation environment are presented. To evaluate the performance and effectiveness of the proposed framework, we utilized Gazebo, a widely-used simulation tool that allows for testing and validation in a virtual environment prior to real deployment. Gazebo simulates the dynamics of the crane, including the effects of gravity, friction, and other physical forces, such as wind and magnetic fields, which are crucial for accurate real-world operation. To validate the proposed framework, two transportation scenarios with different object types and different initial and target positions were evaluated within simplified construction sites. The simulation results presented here demonstrate the framework's ability to accurately replicate tower crane operations, validate motion planning algorithms, and assess the crane's performance in various dynamic conditions.

5.1 Conexes plan

Figure 4 shows the sequential steps involved in transporting conexes from their initial positions to their target locations. The process is systematically divided into six distinct stages for each conex. The detailed breakdown of these stages is as follows:

(a) Action Processing: The Action Management Node (AMN) receives and processes the user-defined plan and initiates the sequence of actions required to execute the operation.

(b) Initial Positioning: The conex is positioned on the ground, prepared for lifting. The AMN starts the plan, beginning with the "Picking" action. This step involves determining the exact initial position of the conex within the workspace.

(c) Object Attachment: The object is attached using the attachment action. To avoid potential collisions with surrounding components in the construction site, this step incorporates a controlled movement in the z-direction.

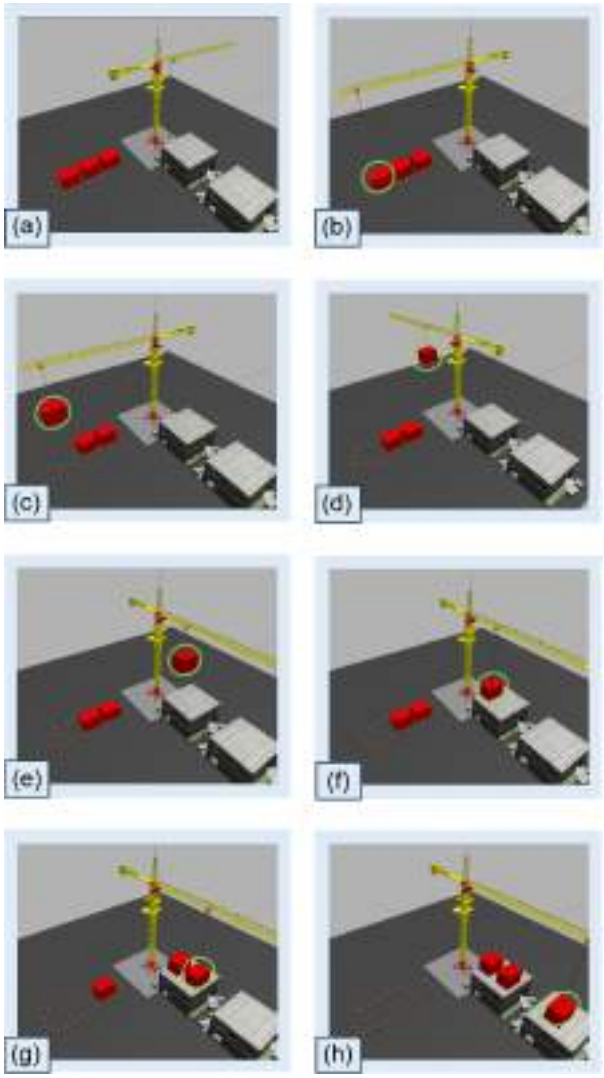


Figure 4. Objects transporting sequences: (a) home position, (b) picking position, (c) attaching the first object, (d) transporting the object, (e) placing position, (f) detaching the first object (g) placing the second object, (h) placing the third object

(d) Object Transportation: The system computes the optimal trajectory for transporting the conex from its initial position to a point directly above the target location, considering TC's working space and controller parameters. This step ensures collision-free movement.

(e) Object Placement: The conex is precisely placed at the target location. To avoid collisions, the AMN

employs accurate control for object placement in this step.

(f) Object Detachment: Finally, the conex is detached from the hook mechanism. Depending on the operational requirements, the crane either returns to its home state or proceeds with plans for the next objects.

(g) ,(h) Placement of the second and the third objects

This approach highlights the AMN's capability to efficiently manage complex transport scenarios while ensuring precision and safety throughout the operation.

5.2 Beams plan

Figure 5 shows another scenario for transporting construction material—beams, in this case—from source place to the destination. This transportation process includes several crane motions for each beam from precise lifting, transporting, and placing. Figure 5 shows: (a) the initial positioning of beams in the construction site, (b) the transportation phase where the crane is actively maneuvering the elements along trajectories, and (c) the final placement at the designated target positions.

Each stage is illustrated with detailed top-down views. The material source area, highlighted in blue, is located at the rear of the TC where beams are initially placed. These beams must be transported to the target area, highlighted in green, positioned at the top of the building within the construction site. The Action Management Node (AMN) executes the transportation plan sequentially for each beam, continuing the process until the entire plan is successfully completed or an error is encountered.

The evaluation scenarios involving conexes and beams validate the proposed AMN and the framework. These tests demonstrate the framework's ability to achieve precise positioning of materials and enable efficient crane operations. The results show the framework's potential to meet the operational requirements of real-world construction environments.

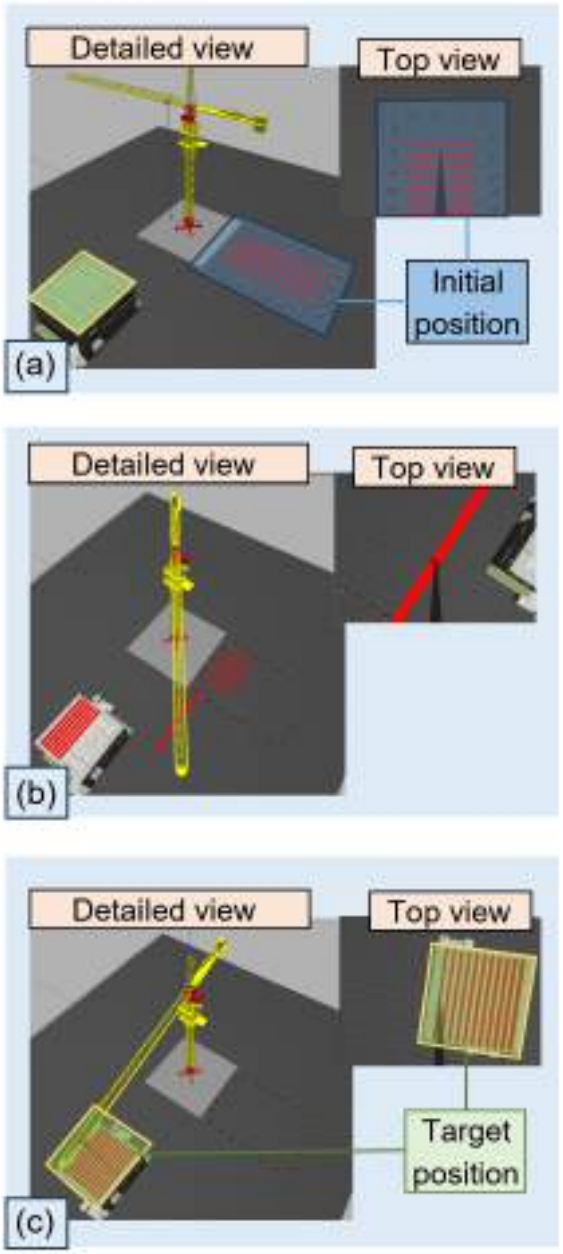


Figure 5. Beams transporting sequences: (a) initial state, (b) in process, (c) final state

6 Discussion

This study introduces and evaluates a comprehensive framework for automating tower crane operations using robotic frameworks. The proposed system integrates TC functionalities within a physics-based simulation environment, supporting both operational-level control and high-level plans. By abstracting the complexities of crane operation, the framework allows users to control the TC without the need for an expert operator. In addition, the action management node demonstrated its effectiveness in two different scenarios, confirming its feasibility and robustness toward automation planning.

This study focused on controlling the actuated joints of the tower crane within a realistic simulation environment, excluding radial and tangential swings from its scope. The assumption is ideal, but the hoisting could experience swing and rotation due to wind or inertia of mass during object transportation. These dynamics are external to TC functionalities and are typically compensated through control strategies and algorithms. It is also important to note that robotic simulators, such as Gazebo, rely on rigid

body dynamics and do not natively support flexible or non-rigid structures. However, the flexibility of TC components, such as ropes, can be approximated by modeling them as chains of rigid links. While this approach closely replicates real-world dynamics, it significantly increases computational load on physics simulation.

Future research will address cable swings and tower crane structural deformation by integrating advanced mathematical models or sensor-based feedback control methods to enhance the accuracy and realism of the framework. Furthermore, we plan to deploy our framework to a scaled-down version of a tower crane in the future.

The modular and scalable nature of the framework allows for its extension to other crane types by providing their unique configurations, characteristics, and operational requirements. Additionally, the framework is built on the Robot Operating System (ROS), which facilitates seamless integration of distributed nodes through robust communication protocols and acts as a layer of abstraction on top of robotic hardware (actors and sensors) or simulation engines, such as Gazebo. Although the current implementation is optimized for Gazebo, the framework can be adapted to interface with other physics engines. However, this would necessitate additional development efforts to address differences in communication protocols and simulation interfaces.

Finally, this framework holds the potential to interface with real-world tower cranes in future applications, provided the cranes are equipped with suitable drivers to enable external connections. However, limitations and specific requirements should be considered when transitioning the framework to real-world operations to ensure reliability and accuracy in practical scenarios. These limitations include safety and compliance issues, as implementing a fully or semi-autonomous robotic crane in construction environments requires compliance with strict safety regulations and industry standards, which may impose additional design and operational constraints. Furthermore, sensor and localization challenges must be addressed, as precise localization of the crane and its load depends on sensor accuracy. External disturbances, and calibration errors could degrade performance, particularly in outdoor construction sites. Additionally, due to the increased communication between the framework and sensors, high-performance computing resources are required to handle real-time data processing and motion planning efficiently.

7 Conclusion

This research presents a simulation framework and action management node for planning and control of the jib, trolley, and hoisting using high-level information from the user and translating them to sequences of actions for the tower crane operations. The developed framework is validated on the physics-based simulator with a real-scaled tower crane model. The developed framework is scalable to a real TC as far as the desired data types are provided to the ROS framework.

Acknowledgements

The presented research was funded by the TUM Innovation Network “CoConstruct”.

References

[1] Mark Burkhardt, Andreas Gienger, Lena Joachim, Norbert Haala, Uwe Sörgel, and Oliver Sawodny. Data-based error compensation for georeferenced payload path tracking of automated tower cranes. *Mechatronics*, 94:103028, 2023. doi:10.1016/j.mechatronics.2023.103028.

[2] United nations, department of economic and social affairs, population division, world urbanization prospects: The 2014 revision, highlights, 2014. On-line: `https://www.compassion.com/multimedia/world-urbanization-prospects.pdf`.

[3] Zhiqian Zhang and Wei Pan. Lift planning and optimization in construction: A thirty-year review. *Automation in Construction*, 118:103271, 2020. doi:10.1016/j.autcon.2020.103271.

[4] Z Zhang and W Pan. Virtual reality (vr) supported lift planning for modular integrated construction (mic) of high-rise buildings. *HKIE Transactions*, 2019. doi:10.33430/V26N3THIE-2019-0015.

[5] SungHwan Cho and SangUk Han. Reinforcement learning-based simulation and automation for tower crane 3d lift planning. *Automation in Construction*, 144:104620, 2022. doi:10.1016/j.autcon.2022.104620.

[6] A. Borrmann, T. Bruckmann, K. Dörfler, T. Hartmann, and K. Smarsly. Towards realizing the information backbone of robotized construction – computational methods and cyber-physical architectures for collaborative robotic fleets. In *Proc. of the CIB W78 Conference 2021*, Oct 2021.

[7] Qi Zhou, Shiyi Zhao, Hongyi Li, Renquan Lu, and Chengwei Wu. Adaptive neural network tracking control for robotic manipulators with dead zone. *IEEE Transactions on Neural Networks and Learning Systems*, 30(12):3611–3620, 2018. doi:10.1109/TSMC.2020.3048722.

[8] Gerasimos Rigatos, Masoud Abbaszadeh, and Jorge Pomares. Nonlinear optimal control for the 4-dof underactuated robotic tower crane. *Autonomous Intelligent Systems*, 2(1):21, 2022. doi:10.1007/s43684-022-00040-4.

[9] Lobna T Aboserre and Ayman A El-Badawy. Robust integral sliding mode control of tower cranes. *Journal of Vibration and Control*, 27(9-10):1171–1183, 2021. doi:10.1177/1077546320938183.

[10] Chuanjing Hou, Can Liu, Zhi Li, and Duchen Zhong. An improved non-singular fast terminal sliding mode control scheme for 5-dof tower cranes with the unknown payload masses, frictions and wind disturbances. *ISA transactions*, 149:81–93, 2024. doi:10.1016/j.isatra.2024.04.001.

[11] Mark Burkhardt, Andreas Gienger, and Oliver Sawodny. Optimization-based multipoint trajectory planning along straight lines for tower cranes. *IEEE Transactions on Control Systems Technology*, 2023. doi:10.1109/TCST.2023.3308762.

[12] Muhammad Muddassir, Mohamed AA Abdelkareem, Tarek Zayed, and Zoubeir Lafhaj. Automating the tower crane: Integrating the development and simulation of path planning and trajectory tracking of tower crane in ros framework. In *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, volume 41, pages 1065–1072. IAARC Publications, 2024.

[13] Teerapat Kian Xiong Ku, Bingran Zuo, and Wei Tech Ang. Robotic tower cranes with hardware-in-the-loop: Enhancing construction safety and efficiency. *Automation in Construction*, 168:105765, 2024. doi:10.1016/j.autcon.2024.105765.

[14] Xiao Lin, Songchun Chen, Hongling Guo, and Ziyang Guo. Developing a robotic system for construction truck crane. In *International Symposium on Advancement of Construction Management and Real Estate*, pages 11–23. Springer, 2022. URL https://link.springer.com/book/10.1007/978-981-99-3626-7.

[15] Xiao Lin, Yu Han, Hongling Guo, Zhubang Luo, and Ziyang Guo. Lift path planning for tower cranes based on environmental point clouds. *Automation in Construction*, 155:105046, 2023. doi:10.1016/j.autcon.2023.105046.

[16] Souravik Dutta, Yiyu Cai, Lihui Huang, and Jianmin Zheng. Automatic re-planning of lifting paths for robotized tower cranes in dynamic bim environments. *Automation in Construction*, 110:102998, 2020. doi:10.1016/j.autcon.2019.102998.

[17] Ghang Lee, Hong-Hyun Kim, Chi-Joo Lee, Sung-Il Ham, Seok-Heon Yun, Hunhee Cho, Bong Keun Kim, Gu Taek Kim, and Kyunghwan Kim. A laser-technology-based lifting-path tracking system for a robotic tower crane. *Automation in Construction*, 18(7):865–874, 2009. doi:10.1016/j.autcon.2009.03.011.

[18] Steven Macenski, Tully Foote, Brian Gerkey, Chris Lalancette, and William Woodall. Robot operating system 2: Design, architecture, and uses in the wild. *Science Robotics*, 7(66): eabm6074, 2022. doi:10.1126/scirobotics.abm6074. URL https://www.science.org/doi/abs/10.1126/scirobotics.abm6074.

[19] Nathan Koenig and Andrew Howard. Design and use paradigms for gazebo, an open-source multi-robot simulator. In *2004 IEEE/RSJ international conference on intelligent robots and systems (IROS)(IEEE Cat. No. 04CH37566)*, volume 3, pages 2149–2154. Ieee, 2004. doi:10.1109/IROS.2004.1389727.

[20] David Coleman, Ioan Sucan, Sachin Chitta, and Nikolaus Correll. Reducing the barrier to entry of complex robotic software: a moveit! case study. *arXiv preprint arXiv:1404.3785*, 2014. doi:10.6092/JOSER_2014_05_01_p3.

Senser Fusion and Hierarchical Inspection Strategy for Aerial-Ground Multi-robot System

Xiayu Zhao[1], Tianyu Ren[1], Yizhi Liu[2], Houtan Jebelli[1]

[1]Department of Civil and Environmental Engineering, University of Illinois Urbana-Campaign, Champaign, IL, USA, xiayuz3@illinois.edu, tianyur2@illinois.edu, hjebelli@illinois.edu
[2]Department of Civil and Environmental Engineering, Syracuse University, Syracuse, NY 13244, USA, yliu580@syr.edu

Abstract –

Infrastructure inspection requires balancing coverage and detail for effective assessment. This paper presents a multi-robot inspection system combining aerial and ground platforms through a hierarchical approach. A hexacopter drone performs rapid site mapping, while specialized ground robots (hexapod and tracked variants) conduct detailed inspections. The system implements bidirectional learning where aerial mapping guides ground robot deployment, while ground inspection data refines aerial strategies. Multi-modal sensor data integration uses an Extended Kalman Filter framework to create unified structural health representations. Experimental results demonstrate 96.3% feature extraction accuracy, sub-millimeter registration between sensors, 3.2-second anomaly response time, and 99.7% collision-free operation. The system shows significant potential for inspecting large-scale infrastructure including bridges, industrial facilities, and power plants.

Keywords –

Multi-robot Inspection; Infrastructure Assessment; Sensor Fusion; Hierarchical Strategy; Bidirectional Learning.

1 Introduction

Infrastructure forms the backbone of modern society, enabling transportation, energy distribution, and industrial operations. Maintaining this critical infrastructure is essential for ensuring public safety, economic stability, and operational reliability. However, traditional manual inspection methods face significant challenges, including time constraints, labor intensity, and accessibility limitations, particularly in large-scale or hazardous environments [1,2].

Recent advances in robotic inspection systems have demonstrated promise in addressing these challenges through automation [3–5]. Single-platform solutions, such as autonomous drones or ground robots, offer improved efficiency and accessibility compared to manual methods [6–8]. However, these systems often struggle to balance comprehensive coverage with detailed examination. Aerial platforms excel at rapid site mapping but typically lack the resolution for detailed structural analysis, while ground robots provide high-fidelity inspection data but are limited by slower operational speeds and restricted mobility. Furthermore, existing robotic inspection systems typically operate with limited adaptability, following predetermined paths without the capability to dynamically adjust their strategies based on real-time environmental feedback [9]. This rigid approach can result in inefficient resource utilization and potentially missed critical areas, particularly in complex or evolving inspection environments.

This paper presents a novel multi-robot inspection system that addresses these limitations through a hierarchical strategy combining aerial and ground platforms. The system employs a specially equipped hexacopter drone featuring 30m-range LiDAR, 4K RGB imaging, and high-resolution thermal sensors for preliminary site mapping, working in concert with specialized ground robots - a hexapod capable of navigating 30° inclines and a tracked variant offering 0.1mm crack detection resolution.

A key innovation of our approach lies in its bidirectional learning mechanism, where inspection strategies are continuously refined through real-time data sharing between platforms. The aerial module efficiently generates preliminary site maps using RANSAC-based segmentation for feature extraction, while ground robots provide high-fidelity inspection data through contact-based sensors and close-range imaging. This collaborative approach is enhanced by a sophisticated sensor fusion framework that integrates visual, thermal, and LiDAR data through an Extended Kalman Filter, achieving sub-millimeter registration accuracy. The system implements a market-based task allocation strategy with token-based access control, enabling efficient coordination between platforms while preventing inspection conflicts. This coordination is

further optimized through a Q-learning based adaptation framework that continuously refines inspection strategies based on accumulated experience and environmental feedback. Experimental results demonstrate 96.3% accuracy in feature extraction and 3.2-second anomaly response time. The system maintains 99.7% collision-free operation while reducing overall inspection time, establishing new benchmarks for automated infrastructure assessment. These achievements suggest the potential for widespread adoption in various infrastructure maintenance applications, from bridge inspection to industrial facility assessment.

2 Related Works

Infrastructure inspection has evolved significantly from traditional manual methods to increasingly automated approaches. Manual inspection methods, while detailed, face substantial challenges including time constraints, labor intensity, and accessibility limitations [10]. These limitations become particularly acute in large-scale infrastructure inspection, where human inspectors often struggle to access critical areas safely and efficiently [11]. Recent advances in robotics have introduced various single-platform solutions for infrastructure inspection. Aerial platforms, particularly UAVs, have demonstrated considerable success in rapid large-scale inspections. Tu and Liang [6] developed a UAV-based bridge inspection system achieving comprehensive coverage through automated flight paths. However, these aerial systems typically sacrifice detail for coverage area.

Multi-robot systems have emerged as a promising solution to address the limitations of single-platform approaches. Cacace et al. [12] demonstrated the advantages of combining aerial and ground robots for industrial facility inspection, but their approach did not implement bidirectional learning between platforms. Yang et al. [13] explored a hierarchical inspection strategy for bridge monitoring, albeit without a comprehensive sensor fusion framework. More recent studies have advanced multi‐robot sensor fusion by emphasizing robustness and adaptability. For instance, Amorim et al. [14] introduced a modified particle filter that robustly fuses visual data from multiple robots to track targets even under severe occlusion, effectively mitigating sensor limitations. Zhou et al. [15] leveraged graph neural networks to combine multi‐view visual information, significantly enhancing individual robot perception accuracy and resilience to sensor failures. Sensor fusion techniques have become increasingly crucial in infrastructure inspection. Huang et al. [16] developed a framework for integrating visual and thermal data in structural assessments, while Nam et al. [17] implemented LiDAR-visual fusion for enhanced defect detection. However, these approaches typically focus on single-platform applications.

Adaptive path planning and learning mechanisms represent a growing focus in robotic inspection systems. Zhang et al. [18] implemented reinforcement learning for optimizing UAV inspection paths, while Zhao et al. [8] developed adaptive navigation strategies for ground robots in complex environments. The integration of market-based task allocation in multi-robot systems, as demonstrated by [19], has shown promise in optimizing resource utilization, though primarily in controlled laboratory settings.

3 Methodology

The multi-robot inspection system combines complementary aerial and ground platforms. The aerial drone uses 3D point cloud technology for large-scale mapping and identifying regions of interest. Ground robots (hexapod and tracked variants) perform detailed inspections of these areas with high-resolution sensors, enabling in-depth structural analysis.

3.1 System Architecture

As shown in Fig. 1, the system architecture is designed to facilitate seamless communication and collaboration between the aerial and ground platforms. A centralized control system coordinates the activities of both modules, ensuring an efficient and adaptive inspection workflow. Each platform is equipped with wireless communication modules to enable real-time data sharing and bidirectional learning.

The aerial module consists of a custom-designed hexacopter drone equipped with an integrated sensor suite. At its core is a primary LiDAR sensor offering 30m range with 0.5° angular resolution, complemented by a 4K RGB camera providing a 120° field of view. Thermal imaging capabilities are delivered through a FLIR (Forward Looking Infrared) camera with 640×512 resolution. An onboard computing unit handles real-time data processing, while an RTK (Real-Time Kinematic) GPS system ensures precise positioning with ±2cm

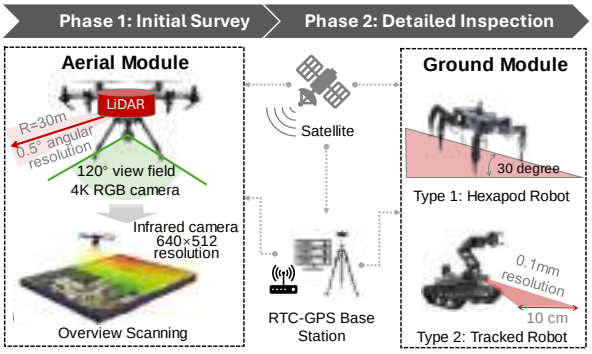


Figure 1. System with inspection phases overview

accuracy.

Two specialized ground robots complement the aerial system. The first is a hexapod robot featuring a 6-legged configuration with 18 degrees of freedom. This platform can carry a 2.5kg sensor payload and navigate terrains with up to 30° incline. Its sensor package includes high-resolution contact sensors for surface analysis and close-range LiDAR for detailed geometric mapping.

The second ground platform is a tracked robot designed for stability and precision. Operating at a maximum speed of 0.5 m/s, it incorporates an integrated stabilization system for precise measurements. The robot carries a specialized surface crack detection camera capable of 0.1mm resolution at 10cm distance for comprehensive structural analysis.

The proposed multi-robot inspection system integrates aerial and ground platforms. The system state at time t can be represented as:

$$S(t) = \{P_{a(t)}, P_{g(t)}, M(t), R(t)\} \quad (1)$$

where $P_{a(t)}$represents the aerial platform state, $P_g(t)$ the ground robot states, $M(t)$ the environmental map, and $R(t)$ the set of identified regions of interest.

The aerial platform state is defined as:

$$P_a(t) = [x_a(t), v_a(t), \omega_a(t), E_a(t)]^T \quad (2)$$

where $x_a(t) \in \mathbb{R}^3$ is position, $v_a(t) \in \mathbb{R}^3$ is velocity, $\omega_a(t) \in \mathbb{R}^3$ is angular velocity, and $E_a(t)$ is remaining energy.

For ground robots, each platform state is:

$$P_g^i(t) = \left[x_g^i(t), v_g^i(t), \theta_g^i(t), E_g^i(t)\right]^T \quad (3)$$

where i denotes the specific ground robot (hexapod or tracked).

The system employs a mesh network topology operating wireless communication. Communication reliability is ensured through redundant channels, while edge computing nodes enable distributed processing. A central server coordinates data fusion and mission control operations across all platforms.

3.2 Hierarchical Inspection Strategy

The system implements bidirectional data flow between platforms. The aerial platform provides global site mapping updates, new Region of Interest (ROI) identifications, mission priority adjustments, and environmental hazard alerts to ground units. Conversely, ground robots transmit high-resolution inspection data, updated anomaly classifications, local environment maps, and status information to the aerial platform.

Phase 1: Initial Survey

The aerial platform initiates inspection using cellular decomposition for coverage planning, operating at altitudes between 30-50m for overview scanning. The system maintains a 60% overlap ratio to ensure accurate 3D reconstruction, generating real-time point clouds with density exceeding 100 points/m².

ROI identification employs RANSAC-based segmentation for feature extraction, coupled with geometric analysis for anomaly detection. The system conducts thermal mapping to identify subsurface defects. Priority scoring incorporates geometric deviation thresholds exceeding 2cm, thermal gradient anomalies greater than 5°C, surface texture irregularities, and structural criticality indices.

The hierarchical inspection strategy prioritizes regions based on a combined score:

$$Score(ROI_i) = \beta_1 C(ROI_i) + \beta_2 A(ROI_i) + \beta_3 D(ROI_i) \quad (4)$$

where $C(ROI_i)$ is criticality, $A(ROI_i)$ is accessibility, and $D(ROI_i)$ is detection confidence for region i.

Phase 2: Detailed Inspection

The detailed inspection task allocation follows a dynamic assignment protocol based on robot capabilities. The system employs a comprehensive cost function incorporating distance to target, robot specialization, battery status, and inspection priority. Ground robot deployment utilizes the enhanced RRT* (Rapidly Exploring Random Tree) algorithm for path planning, complemented by a dynamic window approach for obstacle avoidance. The sliding window optimization approach periodically refines local trajectories based on updated environmental information and inspection requirements. This optimization considers multiple objectives including minimizing path length, maximizing sensor coverage, and maintaining stable sensor positioning for high-quality data collection. The refinement process incorporates feedback from both successful and failed inspection attempts, building a knowledge base that informs future planning decisions.

3.3 Sensor Fusion Framework

The system fuses multi-modal data streams from complementary sensors (Fig 2-a). The aerial platform uses LiDAR for structural mapping, thermal imaging for temperature analysis, and high-resolution cameras for surface inspection. Ground robots carry enhanced versions of these sensors with superior resolution for detailed close-range examination.

3.3.1 Data Collection and Synchronization

The data collection process maintains precise temporal alignment through hardware-level synchronization, achieving millisecond timing accuracy across all sensors. Each measurement receives accurate timestamps and position tags, enabling coherent data integration. The system implements rigorous calibration procedures to optimize intrinsic sensor parameters, establish inter-sensor transformation matrices, and compensate for environmental factors that could affect

measurement accuracy.

Data preprocessing begins with adaptive threshold filtering to remove noise while preserving critical structural information. The system applies statistical outlier removal techniques and motion compensation algorithms to ensure data quality. A crucial step in the pipeline involves spatial alignment of LiDAR, thermal, and visual data streams using advanced registration algorithms. This process creates a coherent multi-modal representation of the inspection site, where each data point combines structural, thermal, and visual information.

The LiDAR data provides precise 3D structural information, capturing geometric features with millimeter-level accuracy. Thermal imaging data reveals temperature variations that may indicate subsurface defects, material fatigue, or thermal stress patterns not visible to other sensors. High-resolution visual data enables detailed surface analysis, particularly for identifying cracks, corrosion, and material degradation. The fusion of these complementary data streams enables the detection of complex structural anomalies that might be missed by single-sensor approaches.

The multi-modal sensor fusion process combines data from different sensors using a weighted fusion approach:

$$F(X) = \sum_{i=1}^{n} w_i S_i(X) \tag{5}$$

where $F(X)$ is the fused data point, $S_i(X)$ represents individual sensor measurements, and w_i are confidence weights satisfying $\sum_{i=1}^{n} w_i = 1$.

3.3.2 Feature Extraction and Unified Data Integration

The feature extraction phase employs multiple specialized algorithms working in parallel. Visual feature extraction utilizes SIFT/SURF algorithms to identify surface anomalies, while geometric processing identifies structural deviations from expected patterns. Thermal pattern recognition algorithms detect temperature anomalies and track thermal gradients across structural elements.

Feature confidence scoring is calculated as:

$$C_f = \alpha_v Q_v + \alpha_t Q_t + \alpha_l Q_l \tag{6}$$

where Q_v, Q_t, and Q_l represent quality metrics for visual, thermal, and LiDAR data respectively, with $\alpha_v + \alpha_t + \alpha_l = 1$.

The final stage of sensor fusion employs an Extended Kalman Filter for state estimation, combining the extracted features into a unified structural health assessment. Point cloud registration using Iterative Closest Point (ICP) algorithms ensures precise spatial alignment of all data sources. The system implements confidence-weighted integration of features from different modalities, prioritizing the most reliable data sources for each type of structural assessment.

3.3.3 Sensor Modalities Alignment and Visualization

The alignment of different sensor modalities is achieved through a transformation matrix:

$$T = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \tag{7}$$

where $R \in SO(3)$ is the rotation matrix and $t \in R^3$ is the translation vector.

The fusion framework produces comprehensive inspection outputs including detailed 3D textured meshes overlaid with thermal data. These visualizations combine geometric, thermal, and visual information into an intuitive representation of infrastructure health. The system generates detailed defect mapping and classification reports, along with confidence metrics for each detected anomaly. Quality assessment metrics provide quantitative measures of inspection coverage and data reliability, enabling informed decision-making for infrastructure maintenance planning.

By integrating diverse sensor data streams and employing advanced processing algorithms, the framework achieves superior inspection accuracy while minimizing false positives. The resulting unified representation of infrastructure health provides unprecedented insight into structural conditions, enabling more effective maintenance planning and risk assessment. The implementation of our sensor fusion framework leverages several established open-source libraries: Point Cloud Library (PCL v1.12.1) for point cloud processing and registration [20], OpenCV (v4.7.0) for image processing and feature extraction [21], and Robot Operating System (ROS Noetic) for inter-platform communication and data synchronization [22]. The RANSAC-based segmentation utilizes optimized implementations from the MRPT (Mobile Robot Programming Toolkit) [23].

3.4 Path Planning and Adaptation

The path planning and adaptation framework (Fig. 2-b) implements a hierarchical strategy that combines global coverage optimization with reactive local navigation. This multi-layered approach enables efficient site inspection while maintaining robustness to environmental changes and emerging inspection requirements.

3.4.1 Global Planning and Multi-Robot Coordination

The global path planning optimization problem is formulated as:

$$\min P \sum_{k=1}^{K} \left(\lambda_1 E(p_k) + \lambda_2 T(p_k) + \lambda_3 R(p_k)\right) \tag{8}$$

subject to:

$$|v(t)| \leq v_{max}, |\omega(t)| \leq \omega_{max}, E(t) \geq E_{min} \quad (9)$$

where $E(p_k)$ is energy cost, $T(p_k)$ is time cost, and $R(p_k)$ is risk cost for path segment p_k.

The system uses modified Boustrophedon decomposition [24] to partition inspection space efficiently, considering structural complexity, accessibility, and priorities while minimizing energy consumption. A global planner creates a task dependency graph for mission scheduling, employing dynamic programming to optimize resources based on battery constraints, sensor capabilities, and real-time requirements. A distributed coordination framework with token-based access control prevents inspection conflicts, using a market-based approach where robots bid for tasks based on position, battery life, and capabilities. Real-time communication enables continuous strategy refinement between aerial and ground platforms.

3.4.2 Local Navigation and Reactive Control

The system implements a multi-layered navigation framework where each robot platform employs specialized algorithms optimized for their mobility characteristics. At its core, ground robots utilize an enhanced RRT* algorithm for local path planning, which integrates real-time obstacle detection and dynamic avoidance capabilities. This planning framework continuously maintains optimal safety distances while ensuring comprehensive inspection coverage through trajectory adjustments based on real-time sensor feedback.

Task allocation within the system follows a sophisticated dynamic assignment protocol that evaluates robot capabilities through a comprehensive cost function. This function synthesizes multiple parameters including Euclidean distance to target, platform-specific specialization factors, current battery levels, and inspection priority weights. The cost function for task assignment can be expressed as:

$$\mathrm{C_{task}(r,t)} = \alpha_\mathrm{d} \mathrm{D(r,t)} + \alpha_\mathrm{s} \mathrm{S(r,t)} + \alpha_\mathrm{b} \mathrm{B(r)} + \alpha_\mathrm{p} \mathrm{P(t)} \quad (10)$$

where $D(r,t)$ represents the distance cost between robot r and task t, $S(r,t)$ is the specialization matching score, $B(r)$ indicates battery level, and $P(t)$ denotes task priority. The weights α_i satisfy $\sum \alpha_i = 1$.

The local navigation system incorporates a dynamic window approach for reactive obstacle avoidance, which optimizes robot velocity v and angular velocity ω within an admissible velocity space: $V_a(x) = (v,\omega)\,|v \in [0, v_{max}], \omega \in [-\omega_{max}, \omega_{max}],\ v \leq \sqrt{2 \cdot dist(x,obs) \cdot a_{max}}, |\omega| \leq \sqrt{2 \cdot \theta(x,obs) \cdot \alpha_{max}}$ where $dist(x,obs)$ is the distance to the nearest obstacle and $\theta(x,obs)$ is the angular deviation.

Trajectory optimization occurs through a sliding window approach that periodically refines local paths based on updated environmental data and inspection requirements. The optimization objective function combines multiple criteria:

$$J(\tau) = w_1 \int_t^{t+T} L(\tau(t))dt + w_2 \int_t^{t+T} C(\tau(t))dt + w_3 \int_t^{t+T} Q(\tau(t))dt \quad (11)$$

where $L(\tau)$ represents path length, $C(\tau)$ denotes coverage quality, and $Q(\tau)$ indicates sensor positioning stability. The weights w_i are dynamically adjusted based on inspection priorities.

3.4.3 Adaptive Learning and Environmental Response

The adaptation framework incorporates reinforcement learning techniques to continuously improve navigation and inspection strategies. The learning system maintains a state representation that includes current position, sensor coverage metrics, and environmental features. Action selection is guided by a policy that balances exploration of uncertain areas with exploitation of known inspection requirements. The reward function considers multiple objectives including coverage completeness, data quality, and energy efficiency.

For adaptive learning, the Q-learning update rule is:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_t + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)] \quad (12)$$

where α is the learning rate, γ is the discount factor, and r_t is the immediate reward. This optimization considers multiple objectives including minimizing path length, maximizing sensor coverage, and maintaining stable sensor positioning for high-quality data collection. The refinement process incorporates feedback from both successful and failed inspection attempts, building a knowledge base that informs future planning decisions.

This adaptive capability, combined with the continuous learning and optimization processes, ensures consistent high-quality inspection results across diverse infrastructure types and operating conditions.

4 Simulation Experiment

4.1 Experimental Setup

The evaluation of the proposed multi-robot inspection system was conducted through comprehensive simulation experiments designed to test the hierarchical inspection strategy, sensor fusion framework, and adaptive path planning capabilities. The simulation environment was developed using ROS Gazebo, incorporating physics-based modeling to ensure realistic robot-environment interactions and sensor behavior.

The test environment (Fig. 2-c) was designed as a construction site spanning 20 × 20 m with corridor widths ranging from 1.5 m to 3 m. The environment features

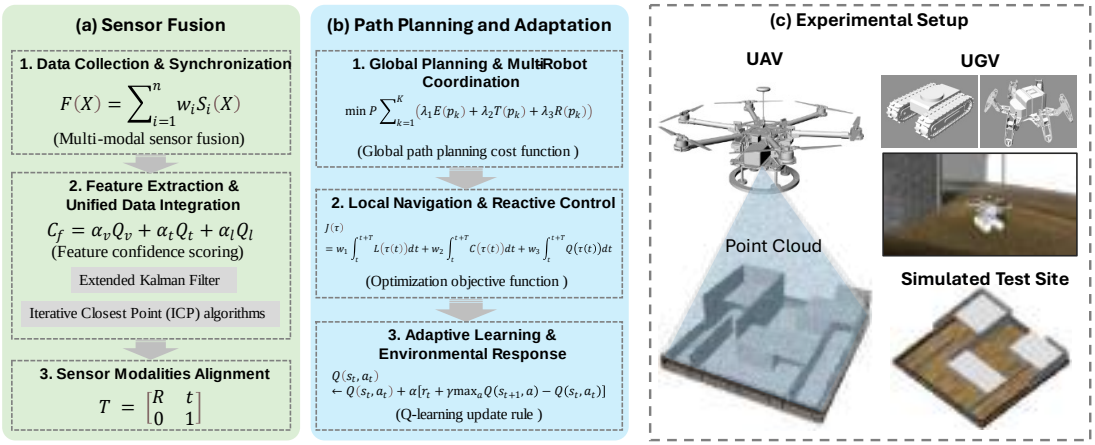


Figure 2. Methodology: (a) Mathematical framework for Sensor Fusion (b) Mathematical framework for Path Planning and Adaptation (c) Experimental Setup with UAV and UGV models

vertical structures representing construction scaffolding, with heights varying between 2 m and 6 m. The site contains both uncovered regions and areas beneath uncompleted floors, creating zones that are unflyable for UAV operation. Predefined structural anomalies were strategically placed throughout the structure, including surface cracks (0.5-2mm width), thermal variations (±5°C from baseline), and geometric deformations (2-5cm deviation from design specifications).

The simulated aerial platform matched the specifications detailed in the methodology, featuring a hexacopter configuration with 30m-range LiDAR (0.5° angular resolution), 4K RGB camera (120° FOV), and FLIR thermal camera (640×512 resolution). The RTK-GPS simulation provided ±2cm positioning accuracy. The ground robots included both the 18-DOF hexapod with 5kg payload capacity and 30° incline capability, and the tracked variant with 0.5 m/s maximum speed and 0.1mm crack detection resolution at 10cm distance.

4.2 Test Scenarios

The experimental protocol was structured to validate the three core components of the system: hierarchical inspection strategy, sensor fusion framework, and adaptive path planning. The simulation implemented the mesh network topology described in the methodology, with wireless communication between platforms and edge computing nodes for distributed processing.

The aerial platform's initial survey followed the cellular decomposition strategy outlined in the methodology, operating at 30-50m altitude with 60% overlap ratio for accurate 3D reconstruction. The point cloud density exceeded 100 points/m², enabling RANSAC-based segmentation for feature extraction. ROI identification utilized the priority scoring system defined in Equation 4, incorporating geometric deviations exceeding 2cm and thermal gradients greater than 5°C.

Ground robot deployment implemented the enhanced RRT* algorithm for path planning, with the sliding window optimization approach for trajectory refinement. The simulation tested the dynamic assignment protocol using the comprehensive cost function from Equation 9, evaluating distance, specialization, battery status, and inspection priority for task allocation.

4.3 Experimental Workflow

The simulation workflow followed the sensor fusion framework detailed in the methodology, implementing millisecond timing accuracy for data synchronization across all platforms. The data preprocessing pipeline included adaptive threshold filtering and motion compensation, followed by spatial alignment of LiDAR, thermal, and visual data streams using the transformation matrix defined in Equation 7.

Feature extraction employed SIFT/SURF algorithms for visual anomaly detection, while geometric processing identified structural deviations. The Extended Kalman Filter state estimation integrated extracted features according to the confidence scoring system defined in Equation 6. The ICP algorithm ensured precise spatial alignment of multi-modal data, with confidence-weighted integration prioritizing the most reliable data sources for each type of structural assessment.

The adaptive learning framework utilized the Q-learning update rule specified in Equation 11, with the reward function considering coverage completeness, data quality, and energy efficiency. The market-based coordination approach managed simultaneous operation of aerial and ground platforms, with the token-based access control mechanism preventing inspection conflicts.

5 Results and Discussion

The simulation experiments validated the effectiveness of our hierarchical multi-robot inspection

system through comprehensive performance evaluation of its core components. The aerial platform's global path planning optimization achieved exceptional efficiency, generating high-density point clouds (127 points/m²) that exceeded design specifications while maintaining precise positioning (±2cm accuracy). This preliminary mapping capability directly enhanced the RANSAC-based segmentation process, achieving 96.3% accuracy in feature extraction and enabling optimal task allocation for ground robots. The system generates three primary output data types from our simulation environment: structural point clouds, multi-spectral inspection maps, and classification metadata. Data is processed both onboard for real-time decisions and offloaded for comprehensive offline analysis. Outputs are formatted for BIM compatibility, enabling integration with existing infrastructure management systems. Offline processing requires approximately 45 minutes per 1000m² on standard engineering workstations.

Building upon this foundation, the sensor fusion framework demonstrated superior integration of multi-modal data streams. The transformation matrix achieved 0.7mm alignment accuracy between sensor modalities, while the weighted fusion approach dynamically adjusted confidence weights to optimize anomaly detection.. This integrated approach resulted in 95.4% accuracy for surface anomaly detection (calculated as TP/(TP+FP+FN) against 200 ground-truth annotations) and 98.2% precision in thermal pattern recognition (TP/(TP+FP)). Sensor integration errors influenced inspection performance through registration misalignment (2.8% of false detections), temporal synchronization drift (1.3% of measurements), and sensor-specific noise (±1.7% threshold variation). The system mitigated these effects through dynamic confidence weighting based on estimated reliability factors from signal-to-noise ratios and registration quality metrics.

The adaptive path planning system further enhanced these capabilities through efficient resource utilization. The enhanced RRT* algorithm, combined with the dynamic window approach, maintained path deviations within 2.8cm. This optimization extended to the Q-learning based adaptation framework, which achieved 3.2-second response times to new anomalies while maintaining 99.7% collision-free operation through the token-based access control mechanism.

The experimental results demonstrate the significant advantages of our integrated approach to infrastructure inspection, particularly in addressing the traditional trade-off between coverage and detail. The hierarchical strategy's success stems from the effective combination of aerial mapping capabilities with specialized ground robot functions, enabling comprehensive inspection while maintaining high data fidelity. This integration proves especially valuable in complex structural environments, where the system's adaptive capabilities ensure thorough coverage while optimizing resource utilization. The bidirectional learning mechanism's performance highlights the importance of real-time adaptation in infrastructure inspection. This improvement directly translates to more reliable structural health assessments, with the multi-modal approach effectively identifying both surface and subsurface defects that might be missed by conventional inspection methods.

Table 1. Performance of each system component

System Component	Metric	Performance
Mapping & Detection		
Aerial Mapping	Point Cloud Density	127 points/m²
	Feature Extraction Accuracy	96.30%
Anomaly Detection	Visual Detection Accuracy	95.40%
	Thermal Pattern Recognition	98.20%
Sensor Integration		
Multi-Modal Fusion	Registration Error	0.7mm
	Data Synchronization	<1ms
Navigation & Control		
Ground Robot Navigation	Path Deviation	2.8cm
System Adaptation	Anomaly Response Time	3.2s
Platform Coordination	Collision-Free Operation	99.70%

6 Conclusion

This paper presents a novel multi-robot inspection system that effectively addresses the fundamental challenges in infrastructure assessment through a hierarchical strategy combining aerial and ground platforms. The system's core innovation lies in its bidirectional learning mechanism and advanced sensor fusion framework, which enable dynamic adaptation to complex inspection environments while maintaining high data fidelity. Through comprehensive simulation experiments, we demonstrated that this integrated approach achieves significant improvements in both efficiency and accuracy compared to traditional inspection methods.

The experimental results validate our hierarchical inspection strategy's effectiveness, with the aerial mapping achieving 96.3% feature extraction accuracy while the ground robots maintained precise navigation within 2.8cm path deviation. The sensor fusion framework demonstrated exceptional performance through sub-millimeter registration accuracy and 3.2-second anomaly response time.

With robust performance across key metrics, the system shows significant promise for broad adoption in infrastructure maintenance. Its 99.7% collision-free operation sets a new standard for collaborative robotic inspection, confirming the feasibility of integrating multiple sensing modalities. Future work should explore deep reinforcement learning and advanced noise reduction for sensor fusion. Implementing wireless map

communication between aerial and ground platforms warrants further investigation to address bandwidth, latency, and signal degradation. Field experiments will measure communication reliability, assess data compression for point cloud transmission, and introduce fault-tolerance protocols to ensure stable operation in varied environments.

This research advances automated infrastructure inspection by presenting a scalable, efficient solution that balances comprehensive coverage with detailed examination. The demonstrated improvements in inspection efficiency and accuracy could significantly enhance infrastructure maintenance, ultimately boosting public safety and structural reliability. Despite promising results, limitations include the simulation environment's imperfect mimicry of real-world conditions, high computational demands for real-time sensor fusion, and potential performance drops in adverse settings. Future efforts will focus on field validation, algorithmic optimization for edge computing, fault detection, and integrating additional sensor modalities such as ground-penetrating radar and acoustic emissions monitoring.

References

[1] S. Dorafshan, M. Maguire, Bridge inspection: human performance, unmanned aerial systems and automation, J Civ Struct Health Monit 8 (2018) 443–476. https://doi.org/10.1007/S13349-018-0285-4/TABLES/3.

[2] A. Ojha, Y. Liu, S. Shayesteh, H. Jebelli, W.E. Sitzabee, Affordable Multiagent Robotic System for Same-Level Fall Hazard Detection in Indoor Construction Environments, Journal of Computing in Civil Engineering 37 (2023). https://doi.org/10.1061/(ASCE)CP.1943-5487.0001052.

[3] X. Zhao, H. Jebelli, A Computational Method for Real-time Roof Defect Segmentation in Robotic Inspection, Computer-Aided Civil and Infrastructure Engineering (2025). https://doi.org/10.1111/mice.13471.

[4] J. Sanchez-Cubillo, J. Del Ser, J.L. Martin, Toward Fully Automated Inspection of Critical Assets Supported by Autonomous Mobile Robots, Vision Sensors, and Artificial Intelligence, Sensors 2024, Vol. 24, Page 3721 24 (2024) 3721. https://doi.org/10.3390/S24123721.

[5] D. Lattanzi, G. Miller, Review of Robotic Infrastructure Inspection Systems, Journal of Infrastructure Systems 23 (2017) 04017004. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000353.

[6] Y. Tu, L. Liang, Research and application of UAV-based intelligent bridge inspection methods, Https://Doi.Org/10.1117/12.3053013 13445 (2024) 422–429. https://doi.org/10.1117/12.3053013.

[7] T. Ren, H. Jebelli, Efficient 3D robotic mapping and navigation method in complex construction environments, Computer-Aided Civil and Infrastructure Engineering (2024). https://doi.org/10.1111/MICE.13353.

[8] X. Zhao, T. Ren, H. Jebelli, Adaptive Intelligence for Robot Navigation Efficiency with a Deep-Reinforcement-Learning-Based Cyber-Physical System, 2024 International Conference on Computing in Civil Engineering (I3CE), 2024.

[9] Y. Liu, H. Jebelli, Intention-aware robot motion planning for safe worker–robot collaboration, Computer-Aided Civil and Infrastructure Engineering 39 (2024) 2242–2269. https://doi.org/10.1111/MICE.13129.

[10] B.-W. Jo, Y.S. Lee, J.-H. Kim, K.-W. Yoon, A review of advanced bridge inspection technologies based on robotic systems and image processing, International Journal of Contents 14 (2018). https://doi.org/10.5392/IJOC.2018.14.3.017.

[11] P.J. Sanchez-Cuevas, G. Heredia, A. Ollero, Multirotor UAS for bridge inspection by contact using the ceiling effect, 2017 International Conference on Unmanned Aircraft Systems, ICUAS 2017 (2017) 767–774. https://doi.org/10.1109/ICUAS.2017.7991412.

[12] J. Cacace, G.A. Fontanelli, V. Lippiello, A Novel Hybrid Aerial-Ground Manipulator for Pipeline Inspection tasks, AIRPHARO 2021 - 1st AIRPHARO Workshop on Aerial Robotic Systems Physically Interacting with the Environment (2021). https://doi.org/10.1109/AIRPHARO52252.2021.9571034.

[13] X. Yang, Z. Ji, J. Wu, Y.K. Lai, C. Wei, G. Liu, R. Setchi, Hierarchical Reinforcement Learning With Universal Policies for Multistep Robotic Manipulation, IEEE Trans Neural Netw Learn Syst 33 (2022) 4727–4741. https://doi.org/10.1109/TNNLS.2021.3059912.

[14] T.G.S. Amorim, L.A. Souto, T. P. Do Nascimento, M. Saska, Multi-Robot Sensor Fusion Target Tracking with Observation Constraints, IEEE Access 9 (2021) 52557–52568. https://doi.org/10.1109/ACCESS.2021.3070180.

[15] Y. Zhou, J. Xiao, Y. Zhou, G. Loianno, Multi-Robot Collaborative Perception With Graph Neural Networks, IEEE Robot Autom Lett 7 (2022) 2289–2296. https://doi.org/10.1109/LRA.2022.3141661.

[16] H. Huang, Y. Cai, C. Zhang, Y. Lu, A. Hammad, L. Fan, Crack detection of masonry structure based on thermal and visible image fusion and semantic segmentation, Autom Constr 158 (2024) 105213. https://doi.org/10.1016/J.AUTCON.2023.105213.

[17] D. Van Nam, P.T. Danh, C.H. Park, G.W. Kim, Fusion consistency for industrial robot navigation: An integrated SLAM framework with multiple 2D LiDAR-visual-inertial sensors, Computers and Electrical Engineering 120 (2024) 109607. https://doi.org/10.1016/J.COMPELECENG.2024.109607.

[18] B. Zhang, Z. Mao, W. Liu, J. Liu, Geometric Reinforcement Learning for Path Planning of UAVs, Journal of Intelligent and Robotic Systems: Theory and Applications 77 (2015) 391–409. https://doi.org/10.1007/S10846-013-9901-Z/METRICS.

[19] A. Pustowka, E.F. Caicedo, Market-based task allocation in a multi-robot surveillance system, Proceedings - 2012 Brazilian Robotics Symposium and Latin American Robotics Symposium, SBR-LARS 2012 (2012) 185–189. https://doi.org/10.1109/SBR-LARS.2012.37.

[20] R.B. Rusu, S. Cousins, 3D is here: Point Cloud Library (PCL), Proc IEEE Int Conf Robot Autom (2011). https://doi.org/10.1109/ICRA.2011.5980567.

[21] opencv/opencv: Open Source Computer Vision Library, (n.d.). https://github.com/opencv/opencv (accessed March 11, 2025).

[22] M. Quigley, B. Gerkey, K. Conley, J. Faust, T. Foote, J. Leibs, E. Berger, R. Wheeler, A. Ng, ROS: an open-source Robot Operating System, n.d. http://stair.stanford.edu.

[23] MRPT — MRPT 2.14.8 documentation, (n.d.). https://docs.mrpt.org/reference/latest/ (accessed March 11, 2025).

[24] H. Choset, P. Pignon, Coverage Path Planning: The Boustrophedon Cellular Decomposition, Field and Service Robotics (1998) 203–209. https://doi.org/10.1007/978-1-4471-1273-0_32.

Vision-Guided Camera Pose Estimation for Robotic Rebar Tying

Shaopeng Xu[1], Huiguang Wang[1], Xiaoyi Lv[1], Lu Deng[1,2], Bo Jin[1,3], and Jingjing Guo[1,2,*]

[1]College of Civil Engineering, Hunan University, China
[2]Hunan Provincial Key Laboratory for Damage Diagnosis for Engineering Structures, Hunan University, China
[3]Hunan Huda Construction Supervision Co., Ltd, Changsha, China
[*]Corresponding Author
xsp@hnu.edu.cn, whg0917@hnu.edu.cn, lyuxiaoyi@hnu.edu.cn, denglu@hnu.edu.cn, jinbo@hnu.edu.cn, [*]guojingjing@hnu.edu.cn

Abstract –

Rebar tying is a critical yet time-consuming process in construction, often criticized by high labor intensity, repetitive motions, and low efficiency. These challenges have led to the development of rebar tying robots, which offer a promising solution to automate and enhance the process. However, existing rebar tying robots are unable to deal with varying orientations of working surface in practical environments, making it difficult for cameras to consistently maintain an ideal pose. To address these challenges, this paper proposes a vision-guided camera pose estimation method. This method includes three steps: (1) image preprosessing, (2) feature point detection and matching, and (3) transformation matrix calculation. Through these steps, the optimal camera pose can be estimated from images captured at random initial poses, allowing the camera to autonomously adjust to its ideal pose. Furthermore, an image evaluation method is introduced, forming a feedback loop with the pose estimation process to ensure high-quality image capture. The proposed method achieves a 99.5% success rate for pose estimation within three attempts, with an average computation time of 1.05 seconds. This approach helps improve the efficiency and accuracy of rebar tying operations, facilitating the automation of the rebar tying process for planar rebar cages.

Keywords –

Rebar tying, Camera pose estimation, Computer vision, Feature point matching, Rebar cage

1 Introduction

Rebar tying is a labor-intensive and repetitive process in rebar product manufacturing, often leading to physical strain [1] and reduced efficiency. To mitigate these challenges, rebar tying robots have been developed, such as Tybot [2], TOMOROBO [3], and RBBD-Bot2.0 [4], reflecting a trend toward automation in rebar tying.

Most rebar tying robots rely on vision guidance for their operations [5,6,7], mainly dealing with horizontal rebar meshes, where the camera is fixed to the robot body and does not require pose adjustments. However, in practical applications, rebar cages may be placed in various orientations, such as horizontally for composite slabs and vertically or inclined for T-beams and box girders in bridges, requiring adaptive adjustments to the camera pose. A 6-Degree-of-Freedom (DoF) robotic arm allows flexible positioning, ensuring that the camera maintains a perpendicular view of the rebar mesh, as shown in Figure 1. Additionally, efficient rebar tying requires that (1) the captured rebar mesh area remains rectangular (as shown in Figure 1); (2) rebar intersection points do not align with image edges, minimizing overlap and optimizing tying efficiency. Manual camera adjustments are time-consuming and impractical in fast-paced environments, necessitating an autonomous pose estimation method.

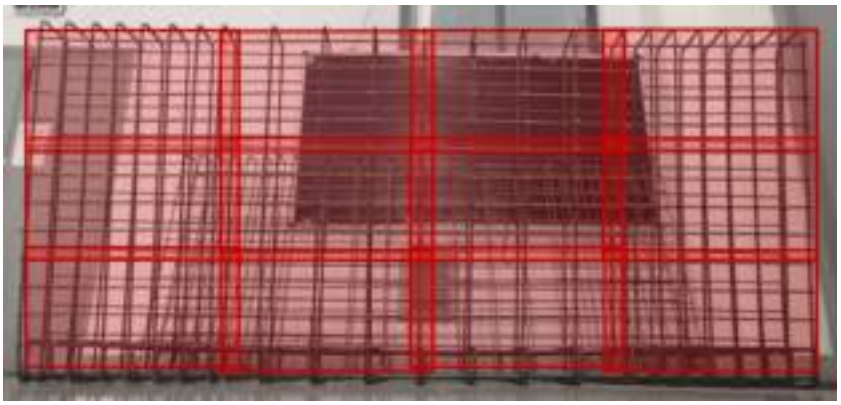


Figure 1. A rebar mesh divided into multiple rectangular work areas

Camera pose estimation involves determining the camera's position and orientation in 3D space, often using feature point matching techniques such as SIFT [8], SURF [9], and ORB [10]. However, these traditional methods struggle with rebar cages due to their uniform structure, leading to mismatches and inaccurate results.

Recent deep learning advancements, such as SuperPoint [11] and SuperGlue [12], have improved feature point detection and matching. SuperPoint provides high-quality keypoints and descriptors, while SuperGlue, based on Graph Neural Networks (GNN) [13], enhances matching accuracy. This study integrates SuperPoint, SuperGlue, and the Perspective-n-Point (PnP) algorithm to achieve accurate and reliable camera pose estimation.

To address the challenges posed by varying rebar cage orientations, this paper proposes a computer vision-based camera pose estimation method with two key innovations: (1) integrating a feature detector, a feature matcher, and PnP algorithm to determine the optimal camera pose from randomly initialized images, and (2) incorporating an image evaluation mechanism to establish a feedback loop for autonomous pose adjustment. This enables rebar tying robots to achieve optimal camera pose adjustment, improving operational efficiency and accuracy.

2 Method

The proposed method comprises three modules as illustrated in Figure 2: (1) image preprocessing; (2) the 6-DOF pose estimation of the camera; (3) image evaluation.

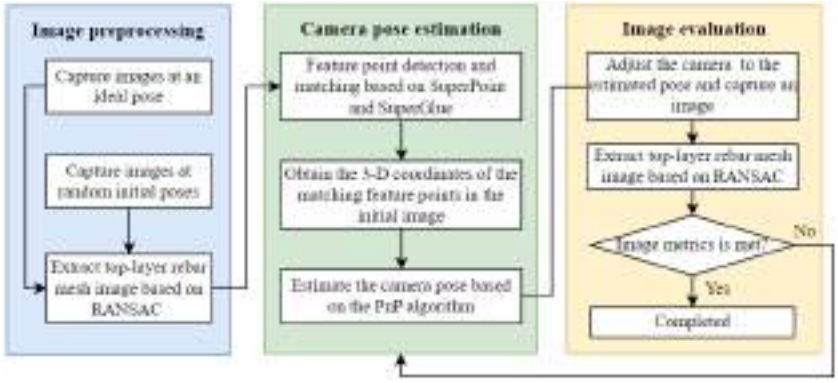


Figure 2. The workflow of the proposed method

2.1 Image Preprocessing

The proposed method requires two inputs: an RGB image and depth map, captured in an ideal pose and random initial pose, respectively. The RGB image and depth map captured in random initial poses are intended to consider the effects from the unanimous ground vehicle (UGV), which will be deviated due to its low positioning accuracy and complicated construction sites. This deviation causes the Z axis of the structured light camera, mounted at the end of the robotic arms, hard to maintain perpendicular to the top-layer rebar mesh of

the rebar cage.

The raw RGB image captured by the structured light camera contains excessive background information, making image matching and rebar crosspoints recognition difficult. To address this, it is essential to filter the background information of the raw RGB image. The process follows these steps: (1) align the captured RGB image and depth map to obtain the 3D point cloud of the rebar cage; (2) the point cloud of the top-layer rebar mesh is contained in the plane closest to the camera's optical center, which is the origin of the camera coordinate system; (3) a method based on RANdom SAmple Consensus [14] (RANSAC) is adopted to fit planes; (4) the plane closest to the camera's optical center is extracted.

After extracting the plane closest to the optical center of the camera, the point cloud belonging to this plane can be obtained. The pixel coordinates corresponding to the point cloud date can be calculated by depth values and the camera's intrinsic matrix, and an RGB image with background information removed can be obtained.

2.2 Camera Pose Estimation

This section describes a method for estimating the optimal camera pose based on RGB images and depth maps captured from a random initial pose. Initially, we perform feature point detection and matching on two filtered RGB images, which have excluded background information. Then, based on the coordinates of these matched feature points, the PnP algorithm is used to obtain the estimated camera pose.

2.2.1 Detect and Match the Feature Points

Recently, a significant number of algorithms have been substantially applied to detect the feature points between two images captured in different poses, such as FAST corner detector [15], SIFT, ORB, etc. However, these algorithms cannot cope with RGB images in complex scenarios, which exhibit poor robust during the process of detecting. Therefore, we adopt an approach based on neural network named SuperPoint, which can detect the feature points without manual annotations. After that, utilize a matching algorithm named SuperGlue to match feature points.

SuperPoint is a self-supervised network for training feature point and descriptors. In this framework, the authors trained a full convolutional network (FCN) based on synthetic dataset, which includes simple geometric shapes with no ambiguity in the feature point locations named MagicPoint. However, this pre-trained network performs poor performance in real dataset. As a result, they introduced Homographic Adaptation for boosting performance from synthetic dataset to real dataset. Finally, this framework can efficiently detect the location of feature points and provide descriptors for feature points. In this paper, we created a dataset consisting of 375 preprocessed rebar cage images as the training dataset for SuperPoint. These images were captured by the camera from different shooting distances and angles relative to the rebar cage. The image collection area is shown in Figure 3.



Figure 3. Image capture area of the SuperPoint training dataset

After detecting the positions of feature points and obtaining their descriptors, they are input into a GNN called SuperGlue. This network can match two sets of feature points using an attention-based flexible context aggregation mechanism.

2.2.2 Camera Pose Estimation

After extracting and matching feature points using the SuperPoint and SuperGlue algorithms, the pixel coordinates of the matched feature points in the initial image (filtered image obtained at a random initial pose) and the reference image (filtered image obtained at the ideal pose) can be obtained. By combining the depth map captured by the camera at the random initial pose, the 3D coordinates of the matched feature points in the initial image can be obtained. The input to the PnP algorithm includes the 3D coordinates of the matched feature points in the initial image and the pixel coordinates of the matched feature points in the reference image.

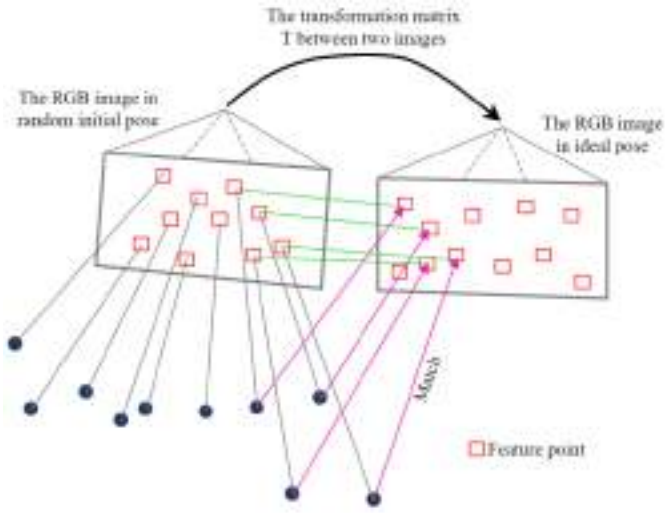


Figure 4. The core target of PnP algorithm

The core target of PnP algorithm is to get the transformation matrix from random initial pose to the ideal pose, which is illustrated in Figure 4. When getting the input of PnP algorithm, the transformation matrix T can be calculated based on Levenberg-Marquardt method [16]. As a result, the estimated pose can be determined by the equation below:

$$Pose_{estimated} = Pose_{initial} \cdot T \quad (1)$$

2.3 Image Evaluation

After obtaining the estimated pose of the camera through the PnP algorithm, move the camera to the estimated pose and collect images, and preprocess the original images to obtain the top-layer rebar mesh image. In order to ensure that the image meets the requirements, the image needs to be evaluated. The specific process is as follows: (1) the rebar crosspoints in the top-layer rebar mesh image are identified and located based on the YOLOv8-pose [17] keypoint detection algorithm; (2) find the maximum and minimum pixel coordinates along the u-axis and v-axis among all the rebar crosspoints, i.e., u_{max}, u_{min}, v_{max}, and v_{min}, as shown in Figure 5; (3) three image evaluation metrics can be calculated to assess the quality of filtered RGB images, where 'a' denotes the pixel width of the image and "b" represents the pixel height of the image.

$$delta_u = abs(1 - \frac{u_{max}}{a} - \frac{u_{min}}{a}) \quad (2)$$

$$delta_v = abs(1 - \frac{v_{max}}{b} - \frac{v_{min}}{b}) \quad (3)$$

$$sum_u = 1 - \frac{u_{max}}{a} + \frac{u_{min}}{a} \quad (4)$$

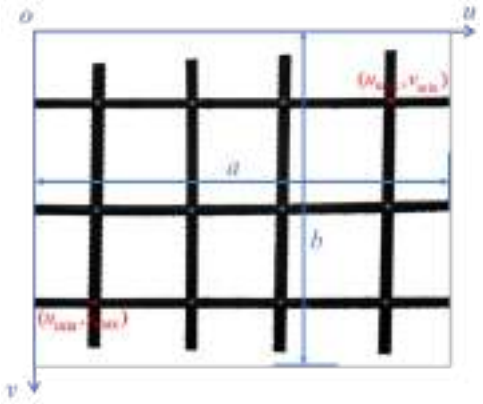


Figure 5. An ideal captured filtered RGB image

It is noticeable that all crosspoints are located on the central area of the filtered RGB image when $delta_u < 0.08, delta_v < 0.08$. Additionally, while all crosspoints are located on the center area of the RGB image, there are still a lot of places around the image that do not contain crosspoints. To make crosspoints are full of the RGB image, sum_u should be in a range from 0.1 to 0.4.

If the filtered image captured by the camera at the estimated pose meets the above metric requirements, the pose estimation is considered successful. Otherwise, if the captured filtered image does not meet the above requirements, it indicates that the camera has not yet reached its ideal pose. The filtered image captured by the camera at the current pose is used as input to the pose estimation module, and the camera pose estimation is repeated until the captured filtered image meets the above metric requirements.

3 Experiments and Results

3.1 Experiment Device

The proposed method was deployed on the rebar tying robot independently developed by our research group, as shown in Figure 6. This robot consists primarily of a mobile chassis, an industrial computer, and a 6-DOF robotic arm equipped with a structured light camera and a rebar tying actuator at the end. The structured light camera can simultaneously capture RGB images and depth maps, providing three-dimensional information of the captured objects. It is mounted at the end of the robotic arm in an "eye-in-hand" configuration, and the transformation matrix from the camera coordinate system to the robotic arm base coordinate system has been obtained through hand-eye calibration.

The industrial computer used in this study is configured as follows: Operating System: Ubuntu 20.04; CPU: 8-core ARM Cortex-A78; GPU: NVIDIA Ampere with 1,792 CUDA cores; RAM: 32GB. The development environment is primarily based on Python.



Figure 6. Rebar tying robot developed by our research group

3.2 Experiment Settings

The filtered rebar cage image shown in Figure 7 was obtained by gradually adjusting the camera pose, capturing images, and preprocessing the images. The evaluation metrics for this image is $delta_u = 0.003, delta_v = 0.026, sum_u = 0.291$, which meets the requirements of the image evaluation module. Thus, it is used as the reference image for the pose estimation module. Filtered images captured and preprocessed at all initial poses are matched with this reference image in the pose estimation module. When capturing this image, the optical center of the camera was 48 cm away from the plane of the top-layer rebar mesh. The z-axis of the camera coordinate system was perpendicular to this plane, the x-axis was parallel to the transverse rebar, and the y-axis was parallel to the longitudinal rebar.

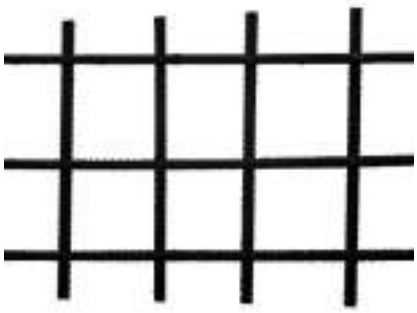


Figure 7. Reference image

To evaluate the performance of the pose estimation method when the camera is in different initial poses, the camera captured images from various shooting distances and angles as the input for the proposed pose estimation method. Additionally, to verify the robustness of the pose estimation method, the images used to test the method's performance were collected from different regions of the rebar cage, differing from the dataset used to train the SuperPoint network, as shown in Figure 3. And the design transverse rebar spacing and the design longitudinal rebar spacing of the rebar cage used in the experiment are both 10 cm.

After obtaining the initial image, the process for testing the performance of the pose estimation method is as follows: the initial image is input into the pose estimation module to calculate the estimated camera pose. The camera is adjusted to the estimated pose to capture an image, and the filtered image is input into the image evaluation module. If the image meets the required metrics, the camera pose estimation is complete, and the computation time for the pose estimation method is recorded. If the image does not meet the metrics, it is used as the initial image and input into the pose estimation module for re-estimation. This process is repeated until the captured image meets the required metrics, with the total computation time and the number of pose estimations recorded.

Considering that the recommended working distance of the structured-light camera is 30 cm to 60 cm, the camera's optical center is set at distances of 40 cm, 45 cm, and 50 cm from the plane of the top rebar mesh to ensure the depth values of rebar pixels in the filtered image fall within this range. This distance is hereafter referred to as the "shooting distance". Taking a shooting distance of 40 cm as an example, the camera pose is first adjusted so that the z-axis of the camera coordinate system is perpendicular to the plane of the top rebar mesh, the x-axis is parallel to the transverse rebar, and the y-axis is parallel to the longitudinal rebar. This pose is referred to as its "standard pose", as shown in Figure

8. Then, the camera can be rotated around its x-axis, y-axis, and z-axis respectively by the following angles: $-10°, -5°, 0°, 5°, 10°$ (counterclockwise is positive, clockwise is negative). Therefore, the camera has a total of 125 initial poses at this shooting distance, and a total of 375 initial poses at the three shooting distances.



Figure 8. Standard pose of the camera

3.3 Experiment Results

For each initial pose scenario, the following parameters are recorded: whether it can be adjusted to the ideal pose, the number of pose estimation attempts required to reach the ideal pose, and the computation time of the pose estimation method. The results are shown in Table 1. It can be seen from Table 1 that for the total of 375 initial pose scenarios at three shooting distances, the success rate of the pose estimation method within three attempts is 99.5%, and the success rate with just one attempt is 91.5%.

Table 1 Number of pose estimation attempts, corresponding scenario counts, and average computation time

Number of pose estimation attempts and the corresponding scenario counts							
Shooting distance (cm)	1	Success rate (%)	2	Success rate (%)	3	Success rate (%)	Failure
40	114	91.2	8	97.6	1	98.4	2
45	115	92.0	9	99.2	1	100.0	0
50	114	91.2	9	98.4	2	100.0	0
Total	343	91.5	26	98.4	4	99.5	2
Average computation time of pose estimation method (run once)						1.05s	

The failure of a single pose estimation attempt can be attributed to the large number of incorrect matching points in the initial image matching results, which leads to inaccurate pose estimation by the PnP algorithm. Consequently, the first estimated image fails to meet the evaluation metrics, necessitating additional pose estimation attempts. If the estimated pose renders the robotic arm unreachable or causes collisions, the pose estimation is considered a failure. The computation time in this study refers to the duration from receiving an image captured at an initial camera pose to generating the corresponding optimal camera pose that meets the predefined requirements. The proposed method achieves an average computation time of 1.05 seconds per pose estimation, meaning that the system can determine the desired camera pose within 1.05 seconds after receiving the initial image. This efficiency ensures minimal impact on production scheduling in precast plants and construction sites, making the method suitable for real-world deployment.

Figure 9 shows the test results of three initial pose scenarios, each of which includes: an image captured and preprocessed by the camera at the initial pose (referred to as "initial image"), the matching result with the reference image, and an image captured and preprocessed at the estimated pose (referred to as "estimated image"). In Figure 9(a), the initial pose of the camera for the initial image is: the shooting distance is 40 cm, and the camera rotates 10°, -10°, and -10° around its own x-axis, y-axis, and z-axis based on the standard pose. The match results show the matching between the initial image and the reference image. The image pose estimation module only requires a single run with a computation time of 1.062 seconds. The evaluation metric value for the estimated image is: $delta_u = 0.005, delta_v = 0.002, sum_u = 0.270$, which meets the requirements. It can be concluded that even when the camera's initial pose significantly deviates from its ideal pose, the proposed pose estimation method can still calculate an ideal pose, enabling the camera to capture rebar cage images that meet the evaluation metrics.

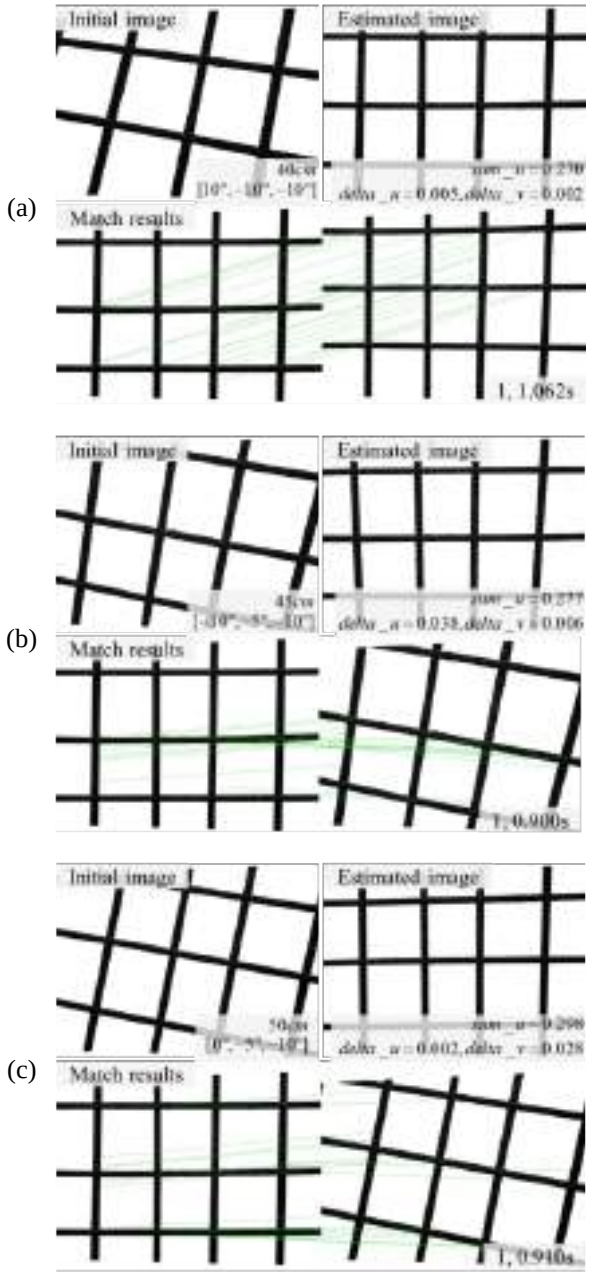


Figure 9. The initial image, matching result, and estimated image corresponding to the camera in different initial pose

4 Conclusion

This paper proposes a vision-guided camera pose estimation method for rebar tying robots, designed to calculate the ideal camera pose. This ideal pose facilitates efficient and high-quality rebar tying for rectangular planar rebar cages. The key innovations of the proposed method are as follows:

(1) A feature point-based camera pose estimation method is proposed. For the 375 initial pose scenarios tested in this study, the method achieves a 99.5% success rate for pose estimation within three attempts, with a 91.5% success rate for a single attempt. The average computation time for each estimation is 1.05 seconds, which is generally sufficient for practical applications.

(2) A method for evaluating the image quality captured by the rebar tying robot's camera is proposed. This method incorporates three evaluation metrics and forms a feedback loop with the camera pose estimation process, further ensuring the quality of the captured images. As a result, it contributes to the efficient and high-quality completion of rebar tying tasks for planar rebar cages.

However, the proposed pose estimation method requires multiple adjustments of the camera under certain initial pose scenarios and is only applicable to planar rebar cages. It is necessary to develop more efficient and robust pose estimation methods applicable to curved rebar cages and corresponding image evaluation metrics in the future.

Acknowledgment

The authors would like to acknowledge the financial support provided by the National Key Research and Development Program of China (Grant No. 2023YFC3806800), the National Natural Science Foundation of China (Grant No. 52308312), the Hunan Province Funding for Leading Scientific and Technological Innovation Talents (Grant No. 2021RC4025), and the Central Guidance for Local Science and Technology Development Fund Project (Grant No. 246Z5404G).

References

[1] M. F. Antwi-Afari, S. Anwer, W. Umer, H. Y. Mi, Y. Yu, S. Moon, and M. U. Hossain. Machine learning-based identification and classification of physical fatigue levels: A novel method based on a wearable insole device. International Journal of Industrial Ergonomics, 93:103404, 2023.

[2] Tybot. TyBot: REBAR TYING ROBOTS. On-line: https://www.constructionrobots.com/tybot, Accessed: 25/12/2024.

[3] TOMOROBO. REBAR TYING ROBOT TOMOROBO. On-line: https://en.kenrobo-

tech.com/tomorobo/, Accessed: 25/12/2024.

[4] RBBD-Bot2.0. The self-propelled intelligent rebar tying robot (RBBD-Bot 2.0) of China Construction Eighth Bureau made its debut. On-line: https://www.thepaper.cn/newsDetail_forward_20940885, Accessed: 25/12/2024.

[5] J. Jin, W. Zhang, F. Li, M. Li, Y. Shi, Z. Guo, and Q. Huang. Robotic binding of rebar based on active perception and planning. Automation in Construction, 132:103939, 2021.

[6] B. Cheng and L. Deng. Vision detection and path planning of mobile robots for rebar binding. Journal of Field Robotics, 41(6):1864–1886, 2024.

[7] R. Feng, Y. Jia, T. Wang, and H. Gan. Research on the System Design and Target Recognition Method of the Rebar-Tying Robot. Buildings, 14(3):838, 2024.

[8] D. G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the seventh IEEE international conference on computer vision*, 2:1150-1157, 1999.

[9] H. Bay, T. Tuytelaars, and L. V. Gool. Surf: Speeded up robust features. In *Computer Vision–ECCV 2006: 9th European Conference on Computer Vision*, Part I 9: 404-417, 2006.

[10] V. E. Rabaud. ORB: An efficient alternative to SIFT or SURF. In *2011 International conference on computer vision*, 2564-2571, 2011.

[11] D. Detone, T. Malisiewicz, and A. Rabinovich. SuperPoint: Self-supervised interest point detection and description. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 224-236, 2018.

[12] P.-E. Sarlin, D. Detone, T. Malisiewicz, A. Rabinovich, and E. Zurich. SuperGlue: Learning feature matching with graph neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 4938-4947, 2020.

[13] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini. The graph neural network model. IEEE Transactions on Neural Networks, 20(1): 61-80, 2009.

[14] M. A. Fischler and R. C. Bolles. Random sample consensus. Communications of the ACM, 24(6): 381-395, 1981.

[15] E. Rosten and T. Drummond. Machine learning for high-speed corner detection. In *Computer Vision–ECCV 2006: 9th European Conference on Computer Vision*, Part I 9: 430-443, 2006.

[16] H. P. Gavin. The Levenberg-Marquardt algorithm for nonlinear least squares curve-fitting problems. Department of Civil and Environmental Engineering, Duke University, 2019.

[17] S. Wang, L. Deng, J. Guo, M. Liu, and R. Cao. Automatic quality inspection of rebar spacing using vision-based deep learning with RGBD camera. In *Proceedings of the International Symposium on Automation and Robotics in Construction*, 41:57-64, 2024.

Reinforcement Learning with Vision-based State Estimation for Generalized Robotic Assembly

Mark Trovinger[1], Leonel Giacobbe[1], Tanner Grantham[1], Chuangchuang Sun[2], Jingdao Chen[1]

[1]Department of Computer Science and Engineering, Mississippi State University, USA
[2]Department of Aerospace Engineering, Mississippi State University, USA

mrt364@msstate.edu, lg1221@msstate.edu, tdg295@msstate.edu, csun@ae.msstate.edu, chenjingdao@cse.msstate.edu

Abstract -

Robotic manipulation is a promising technology that has the potential to further automate construction tasks such as bricklaying, concrete printing, or crane assembly of prefabricated components. With the increasing complexity of modern construction projects, more sophisticated planning algorithms are needed to optimize the efficiency of robotic manipulator operations. Reinforcement learning (RL) has emerged in recent research as a way to solve planning problems in high-dimensional state spaces or where the world model is unknown or noisy. Most RL algorithms for robotic manipulation target low-level control and is not scalable to multi-object or complex assembly tasks. This research proposes a reinforcement learning framework with integration of vision-based state estimation and inverse kinematics for generalized robotic assembly tasks. A contour detection and linear regression algorithm is used to perform initial state estimation and grasp point regression, whereas the reinforcement learning algorithm involves a discrete state space where transitions between states are handled using inverse kinematics. The proposed framework is validated through a physics simulation as well as a lab-based experimental setup with robotic assembly of wooden blocks.

Keywords -

Reinforcement learning; computer vision; robot assembly

1 Introduction

Automation of construction work is of great need in the building industry to allow construction tasks to be carried out more accurately and efficiently, especially in view of the increasing shortage of skilled workers. In the construction industry, roboticized work can come in the form of robotic bricklaying [1], concrete printing [2], or crane assembly of prefabricated components [3]. Still, using robotics technology for construction work remains challenging due to the unstructured and heterogeneous nature of construction environments [4].

Reinforcement learning (RL) has emerged as a promising research area for robotic control and robotic assembly of complex objects [5, 6]. Contrary to prior methods which rely on having a known physical model of an environment, reinforcement learning allows a robotic agent to learn the characteristics of an unknown environment directly through experience and taking actions in that environment.

Existing work in reinforcement learning for robotic assembly tasks [7, 6] usually define the states as end effector pose and actions as end effector displacements, both in continuous Cartesian space. However, the limitation of this approach is that the learning process involves long convergence times and low success rates due to the large state space and action space, making the task extremely difficult to learn. Additionally, robotic movement does not lend itself to discrete timesteps; robots require time to start and stop movement and so additional work is needed to ensure that an agent does not attempt to take actions when the robot is incapable of doing so. These obstacles can largely be overcome in simulation, but additional issues are faced when moving from a simulation of a robot to the real world.

To address these issues, this research proposes to leverage underlying inverse kinematics mechanisms for the robotic manipulator, allowing the reinforcement learning algorithm to act at a high level with a smaller state space, thus simplifying the learning problem. In addition, a vision module is used for state estimation, allowing the robot to adapt to equipment and materials being transported about in a dynamic construction environment.

2 Literature Review

2.1 Robotic assembly tasks

Robots are uniquely suited to handle many assembly tasks that are dull, dirty, or dangerous. As a result, robots are deployed to accomplish basic tasks in the construction industry [8], such as laying bricks [1], concrete 3D printing [2], etc. Other tasks include assembling interior structures, such as flooring installation [9], HVAC and plumbing installation [10], and assembling furniture such

as an IKEA chair using multiple collaborative robotic manipulators [11].

2.2 Planning methods in construction robotics

Planning methods find many applications in construction robots, such as for crane lift planning [12]. Work in [13] considers building information modeling in construction robotic task planning. [14] addresses the issue of long-horizon multi-robot planning in construction assembly. Sampling-based planners, such as RRT [15], have been used for crane planning and beyond. However, state-space planners and sampling-based planners mostly assume a known environment with static obstacles. To work in dynamic construction environments, multi-modal perception is key for real-time feedback and situational awareness in planning [16]. In the domain of crane lift planning, relevant work include dynamic 3D modeling in environments with low visibility and dense obstacles [17], and real-time monitoring of crane load [18].

2.3 Learning methods in construction robotics

Recently, learning methods have also been explored in construction robotics [19] to tackle problems of higher complexity and improve operation in uncertain environments. Robots empowered by reinforcement learning, for example, Deep Deterministic Policy Gradient (DDPG), have been used for the robotic assembly of timber joints [20]. Human-robot collaboration in construction [21] can also be achieved via reinforcement learning with safety and efficiency. Another learning paradigm, imitation learning, or Learning from Demonstration (LfD) [22] is also deployed in construction robots, such as performing quasi-repetitive construction tasks through human demonstration [23]. Moreover, [24] developed an imitation learning approach for robot teleoperation in construction. Nevertheless, research gaps remain in integrating these black-box learning-based methods with more established planning or kinematics frameworks, and also in integrating the manipulation system with vision-based feedback. There is also a research need for scalable reinforcement learning algorithms [25] that can work for complex, multi-agent [26] tasks in construction environments.

3 Methodology

3.1 Problem setup

In this research, tasks such as bricklaying, crane assembly etc. are represented as a generalized robotic assembly problem. The generalized robotic assembly problem involves assembling n components from initial states $\{s_1, s_2, s_3, ..., s_n\}$ to a complete structure with target states $\{t_1, t_2, t_3, ..., t_n\}$. The assembly task may involve order constraints such that some components need to be installed before other components are installed, which is specific to the assembly task. The total manipulation cost is the time taken to move each component from its initial state to the final state $\sum_n T(s_n, t_n)$. The key assumptions are that (i) the initial states are unknown, and (ii) the manipulation costs are unknown. Therefore, the optimization process involves finding the least time required for a robotic manipulator to complete the assembly task.

3.2 Vision-based initial state estimation and grasp point regression

The proposed method follows the framework described in Figure 1.The vision module starts by obtaining an image frame from a camera mounted near the end effector of the robotic arm. The robot's position remains consistent during the image acquisition process, ensuring a stable stream of extractable data. The image undergoes several preprocessing steps, including grayscale and denoising filters. Non-local means denoising is chosen due to its high retention of detail. Then, a contour detection algorithm is utilized to obtain the outline of prominent shapes within the frame. In this work, the specific target objects are rectangular wooden blocks, so our implementation uses OpenCV's contour detection algorithm to find the four corner points of any rectangular shape in the image. The preprocessing steps, including the grayscale conversion and denoising, were implemented to overcome the noise limitations induced by the camera included with the Kinova Robotic Arm. The area threshold for contour detection was set to 100 pixels, as that value effectively filtered out noise-induced false positives while maintaining accurate detection for rectangular blocks.

The position of the points in the frame are represented as $(xPixel, yPixel)$ tuples, with the top-left corner of the image serving as the origin $(0, 0)$. At this point, the tuples of the four vertices are known, but the specific object corners they represent (top/ bottom, left/ right) are not. To determine the correct assignment, we treat the 2 points with the highest $xPixel$ values as our right points, and the remaining points are treated as left points. Subsequently, the $yPixel$ values between the two right points and left points are compared. Among the right and left side points, the ones with smaller $yPixel$ values are designated as top points, and the remaining ones are assigned as bottom points. With these four points identified, we can solve for the correct rotation angle about the block's z-axis as well as the coordinates of the center of the block. These coordinates of all blocks detected are stored as the initial state of the assembly task.

Then, using the center of the block and its Z-axis rotation, we pass this data to a regression model to calculate the movement required by the robotic arm to

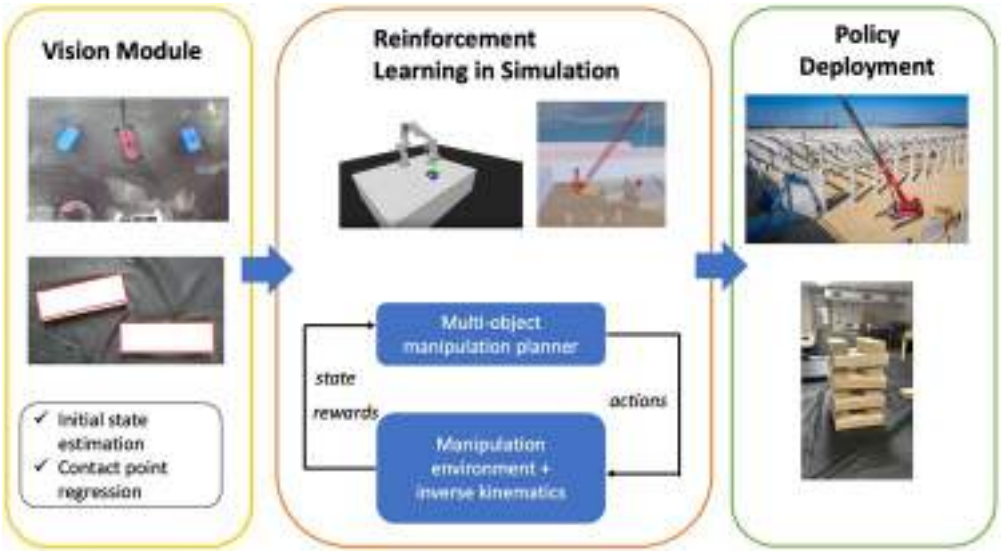


Figure 1. Proposed framework for reinforcement learning with vision-based state estimation for generalized robotic assembly (crane examples are adapted from [27])

successfully grasp the block. Our regression model is a linear model consists of 50 data points of the form $(xPixel, yPixel) \rightarrow (xMovement, yMovement)$, where $xPixel$ and $yPixel$ represent the coordinates of the center of the wooden block, and $xMovement$ and $yMovement$ represent the translational displacement required to position the robotic arm directly above the block to pick up. The training data set was manually annotated through human demonstrations by calculating the block's midpoint and recording the movements needed for the arm to transition from the initial camera position to a pick-up position above the block. The Z movement is not taken into account, as all blocks are assumed to be at the same height when placed on the surface where the arm lies. Figures 2 and 3 show visualizations of the regression model trained to predict the best robot manipulator pose given the visual position of the target object. In our study, the aforementioned linear regression model demonstrated excellent predictive performance, achieving R^2 scores of 0.9865 for the X-coordinates and 0.9985 for the Y-coordinates.

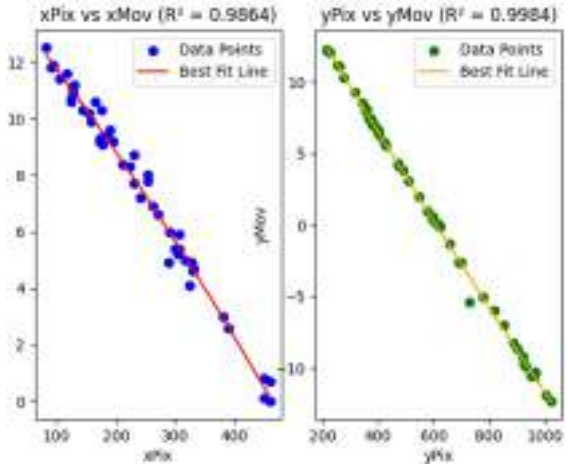


Figure 2. Plot of training data points and line of best fit for regression of desired manipulator pose

3.3 Simulation setup

The simulation consists of a modified Panda Gym environment from [7]. The Panda Gym environment was selected because it allows for simulating robotic manipulation tasks in an RL interface compatible with Stable-Baselines3 [28]. In order to utilize RL algorithms for planning, actions that normally consist of low-level movements in the arm, such as changing joint position and gripper width, are replaced by pre-planned actions. These actions are higher-level, and correspond to the underlying goal within the task; such as moving the arm to a specified location and opening and closing the gripper. In our initial work, we used the simulation of the Franka Emika Panda robot that is present in the Panda Gym environment. This robot was already present in the simulation software, and is similar to the Kinova robot that we were able to utilize it while implementing the Kinova robot. The simulation runs at 25 Hz[7], and each task is given 100 timesteps in order to complete each movement. Using the relationship that a single timestep equals 2ms, we are able to determine how long each subtask takes.

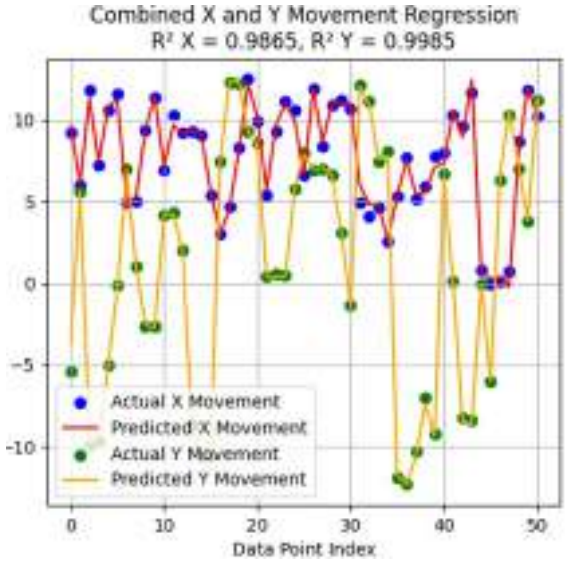


Figure 3. Graph of regression line between the desired robot manipulator pose and the visual position of the target object to be manipulated

3.4 Reinforcement learning

In typical robotic manipulator simulations, the state of the simulation is a set of continuous variables, that define the position of the components of the robot in simulated space. For example, in environments where the agent controls the end-effector's position, the coordinates of the end effector are present in the state or observation. These observations are used by the agent to select actions from a policy π, in order to maximize the returns (or sum of rewards) over the course of a training episode. These rewards are typically "sparse" in manipulator environments, meaning that the agent only receives a positive reward at the completion of the task.

Whereas, in our proposed RL framework, a discrete state space is used where transitions between discrete states are handled using inverse kinematics. The learning process thus assumes that there is an underlying inverse kinematics function which is an unknown model. The cost (i.e. manipulation time) of applying this unknown inverse kinematics model has to be learned through interactions with the environment. In this research, we also assume that this cost of manipulation can be approximated through a physics simulation. This way, the process of moving a single object can be viewed as "deterministic" since it is controlled by inverse kinematics and not varied by the RL agent. The RL agent then handles the process of optimizing the movement of multiple objects to complete an assembly task.

3.5 Implementation Details

All code and algorithms are deployed using a software stack consisting of the Kinova Gen3 API, OpenCV, Stable-Baselines3, PyBullet, and Panda-Gym's learning environment. OpenCV is used to estimate the real-world coordinates of wooden blocks positioned on a lab bench. Whereas, the Kinova API is used for manipulation planning, where our algorithm computes the desired goal state of the robot's end effector, and the underlying API will calculate the required joint angles to reach that position using inverse kinematics for the Kinova robot. The simulation environment is built upon Panda Gym, which allows for modular integration between robot simulations and reinforcement learning algorithms. The simulation environment uses PyBullet, a physics engine that models collisions, soft body dynamics, and rigid body dynamics to replicate real world environmental constraints as closely as possible. The Kinova Gen3 7DOF robot arm is loaded into simulation using the Unified Robot Description Format (URDF). The URDF contains information regarding mass, moments of inertia and force limits for each of the arm's 7 joints. Finally, the reinforcement learning algorithms used are obtained from the Stable Baselines3 library.

4 Results

4.1 Evaluation metrics in simulation

To evaluate the proposed framework, we design two assembly tasks which are the 3-layer tower structure (6 components) and the 3-layer wall structure (8 components). The wall structure is designed to simulate the brick laying task in construction. The tasks created in the simulator were designed to match, as closely as possible, the tasks performed in the lab setting. The metrics used for evaluating the performance of the agents are the assembly time and success rate, where a successful run is defined as one in which the blocks are detected, picked up, and deposited at the correct target positions accurately.

In order to evaluate our proposed "deterministic" approach compared to other RL algorithms, we evaluated three current state of the art algorithms used in robotic manipulation, Deep Deterministic Policy Gradients (DDPG)[29], Twin Delayed Deep Deterministic policy gradient (TD3)[30], and Soft Actor Critic (SAC)[31]. In addition, these algorithms make use of Hindsight Experience Replay (HER), which allows for a more sample efficient agent training process. HER was used due to the sparse nature of the rewards given in each task; with the agent only receiving a reward at end of the training loop, and even then, only when the agent has successfully completed the task. Hyperparameters used for the baseline testing match those found in HER[32].

For the RL agents, the training regime is 50M timesteps, and the hyperparameters used match those used in Panda Gym. The higher number of timesteps for training was possible due to training the agents in parallel. The reward each agent receives is -1 for each timestep taken for each subtask (moving the arm to the object, opening and closing the gripper, etc.), with a maximum number of timesteps allowed per objective task of 350 for the 3-layer tower, and 500 for the 3-layer wall and the total return and is represented as:

$$G = -\sum_{0}^{T}\sum_{0}^{\infty} t_{subtask} \quad (1)$$

Where G represents the total episodic return, T represents the total number of allowed timesteps and $t_{subtask}$ is the number of timesteps per deterministic subtask. Hyperparameters used in the deterministic approach are referenced from work by Da Costa[33].

Quantitative results for assembling the 3-layer tower and 3-layer wall are shown in Tables 1 and 2 respectively. Results show that the proposed algorithm achieves higher rewards (and therefore shorter number of steps to complete the assembly task) compared to other algorithms. The proposed algorithm also achieves a much higher success rate: since the algorithm solves the state transitions in a "deterministic" way using inverse kinematics, it is almost guaranteed to finish the assembly task. Whereas, other RL algorithms struggle to solve the task because they have to search for a policy in a very large state space and action space. Qualitative results of the completed assembly tasks in simulation are shown in Figure 4.

Table 1. Performance comparison for 3-layer tower assembly in simulation

Method	Reward	Success Rate
DDPG	-350	0.01
SAC	-349	0.02
TD3	-350	0.01
Proposed deterministic	-315*	1.00

Table 2. Performance comparison for 3-layer wall assembly in simulation

Method	Reward	Success Rate
DDPG	-498	0.015
SAC	-500	0.010
TD3	-500	0.010
Proposed deterministic	-494*	1.000

4.2 Qualitative results for lab testbed

Figure 5 shows proof-of-concept results for physically assembling a 3-layer hollow tower, and a 3-layer wall structure out of wooden blocks using a Kinova Gen3 L53 7 degree-of-freedom (DOF) robotic arm. The implementation makes use of the inverse kinematics module in the Kinova API, and the geometry of the structure to be assembled is pre-programmed in a Python script and deployed on the robot.

Table 3 shows a comparison of the total assembly time of the 3-layer tower as well as the 3-layer wall from the lab testbed. The time taken to place each component is highly dependent on the component's target position and orientation, as well as the specific implementation of the underlying inverse kinematics algorithms (the same end effector pose can be reached through many different possible joint angles). Reinforcement learning provides an opportunity to learn these interactions and arrive at an optimized policy to minimize assembly time.

Table 3. Comparison of total assembly time from the lab testbed

Structure	#Components	Assembly Time (s)
3-layer tower	6	195
3-layer wall	8	180

5 Limitations and Discussion

Currently, the vertical (Z-axis) distance between the arm's end effector and the target object to be manipulated is assumed to be fixed. Future work will look to overcome this limitation by implementing triangle similarity or using depth images, allowing for more generalized behavior that more closely resembles real-life construction environments. Moreover, the current model only accounts for rotation about the Z axis and not about the X and Y axes. Moving forward, the regression model can be adjusted to consider these additional rotation angles as parameters, and allow for more robust grasping and manipulation of objects. The current implementation only considers vision input in the initial stage for state estimation and grasp point regression. Ideally, continuous vision feedback will be important to correct state estimation errors or motion noise. Theoretically, our approach can scale to larger construction environments, as long as we can represent the scene accurately in simulation and train the agent in simulation. Our approach relies on having access to inverse kinematics and this is a key assumption. For complex construction machinery where an inverse kinematics solution is not available, probabilistic motion planning algorithms may be a viable alternative [34].

6 Conclusion

This research proposes a reinforcement learning framework with integration of vision-based state estimation and



Figure 4. Results for 3-layer tower assembly and 3-layer wall assembly in the Panda Gym simulation



Figure 5. Qualitative results of a 7-DOF Kinova robot arm assembling a hollow tower (left 3 images) and a wall structure (right 3 images) using wooden blocks

inverse kinematics for generalized robotic assembly tasks. The advantage of the proposed framework is a generalized formulation of robotic assembly tasks for reinforcement learning algorithms that can also be adapted to arbitrary initial states. Future work in this area involves exploring multi-agent reinforcement learning algorithms that can allow the assembly tasks to be completed by two or more robotic manipulators that can learn to coordinate with one another. Human-robot collaboration [35] will also be explored to investigate how robotic agents can collaborate with human workers to complete assembly tasks.

7 Acknowledgements

The work reported herein was supported by the National Science Foundation (NSF) (Award #IIS-2153101). Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

References

[1] Zakaria Dakhli and Zoubeir Lafhaj. Robotic mechanical design for brick-laying automation. *Cogent Engineering*, 4(1):1361600, 2017.

[2] Daniel Kajzr, Tomáš Myslivec, and Josef Černohorskỳ. Modelling, analysis and comparison of robot energy consumption for three-dimensional concrete printing technology. *Robotics*, 13(5):78, 2024.

[3] Aiyu Zhu, Tianhong Dai, Gangyan Xu, Pieter Pauwels, Bauke De Vries, and Meng Fang. Deep reinforcement learning for real-time assembly planning in robot-based prefabricated construction. *IEEE Transactions on Automation Science and Engineering*, 20(3):1515–1526, 2023. doi:10.1109/TASE.2023.3236805.

[4] Iro Armeni, Jingdao Chen, Chen Feng, Inbae Jeong, Ci-Jyun Liang, Kristian Morin, Nikos Tsagarakis, Xi Wang, and Liangjun Zhang. Construction robotics and automation [tc spot light]. *IEEE Robotics and Automation Magazine*, 31(4):186–192, 2024. doi:10.1109/MRA.2024.3481769.

[5] Lei Huang and Zhengbo Zou. Deep reinforcement learning-based construction robots collaboration for sequential tasks. In *Proceedings of the 1st Future of Construction Workshop at the International Conference on Robotics and Automation (ICRA 2022)*, pages 48–51, Philadelphia, PA, USA, May 2022. International Association for Automation and Robotics in Construction (IAARC). doi:10.22260/ICRA2022/0015.



Figure 6. The robot camera's top view of two blocks and an assembled hollow tower. Red lines indicate detection results from the vision-based state estimation algorithm.

[6] Weijia Cai and Zhengbo Zou. A reinforcement learning based approach for conducting multiple tasks using robots in virtual construction environments. In *Proceedings of the 1st Future of Construction Workshop at the International Conference on Robotics and Automation (ICRA 2022)*, pages 44–47, Philadelphia, PA, USA, May 2022. International Association for Automation and Robotics in Construction (IAARC). doi:10.22260/ICRA2022/0014.

[7] Quentin Gallouédec, Nicolas Cazin, Emmanuel Dellandréa, and Liming Chen. panda-gym: Open-Source Goal-Conditioned Environments for Robotic Learning. *4th Robot Learning Workshop: Self-Supervised and Lifelong Learning at NeurIPS*, 2021.

[8] Thomas Bock. Construction robotics. *Autonomous Robots*, 22:201–209, 2007.

[9] Kepa Iturralde, Malte Feucht, Daniel Illner, Rongbo Hu, Wen Pan, Thomas Linner, Thomas Bock, Ibon Eskudero, Mariola Rodriguez, Jose Gorrotxategi, et al. Cable-driven parallel robot for curtain wall module installation. *Automation in Construction*, 138:104235, 2022.

[10] Kangkang Duan, Christine Wun Ki Suen, and Zhengbo Zou. Robot morphology evolution for automated hvac system inspections using graph heuristic search and reinforcement learning. *Automation in Construction*, 153:104956, 2023.

[11] Francisco Suárez-Ruiz, Xian Zhou, and Quang-Cuong Pham. Can robots assemble an ikea chair? *Science Robotics*, 3(17): eaat6385, 2018. doi:10.1126/scirobotics.aat6385. URL `https://www.science.org/doi/abs/10.1126/scirobotics.aat6385`.

[12] Songbo Hu, Yihai Fang, and Yu Bai. Automation and optimization in crane lift planning: A critical review. *Advanced Engineering Informatics*, 49: 101346, 2021.

[13] Sungjin Kim, Matthew Peavy, Pei-Chi Huang, and Kyungki Kim. Development of bim-integrated construction robot task planning and simulation system. *Automation in Construction*, 127:103720, 2021.

[14] Valentin N Hartmann, Andreas Orthey, Danny Driess, Ozgur S Oguz, and Marc Toussaint. Long-horizon multi-robot rearrangement planning for construction assembly. *IEEE Transactions on Robotics*, 39(1):239–252, 2022.

[15] Ying Zhou, Endong Zhang, Hongling Guo, Yihai Fang, and Heng Li. Lifting path planning of mobile cranes based on an improved rrt algorithm. *Advanced Engineering Informatics*, 50:101376, 2021.

[16] Jingdao Chen, Pileun Kim, Dong-Ik Sun, Chang-Soo Han, Yong-Han Ahn, Jun Ueda, and Yong K. Cho. Workspace modeling: Visualization and pose estimation of teleoperated construction equipment from point clouds. In *37th International Symposium on Automation and Robotics in Construction (ISARC)*, pages 781–788, October 2020.

[17] Jingdao Chen, Yihai Fang, and Yong K. Cho. Real-time 3d crane workspace update using a hybrid visualization approach. *Journal of Computing in Civil Engineering*, 31(5):04017049, 2017. doi:10.1061/(ASCE)CP.1943-5487.0000698.

[18] Leon C. Price, Jingdao Chen, and Yong K. Cho. Dynamic crane workspace update for collision avoidance during blind lift operations. In Eduardo Toledo Santos and Sergio Scheer, editors, *Proceedings of the 18th International Conference on Computing in Civil and Building Engineering*, pages 959–970, Cham, 2020. Springer International Publishing. ISBN 978-3-030-51295-8.

[19] Juan Manuel Davila Delgado and Lukumon Oyedele. Robotics in construction: A critical review of the reinforcement learning and imitation learning

paradigms. *Advanced Engineering Informatics*, 54: 101787, 2022.

[20] Aleksandra Anna Apolinarska, Matteo Pacher, Hui Li, Nicholas Cote, Rafael Pastrana, Fabio Gramazio, and Matthias Kohler. Robotic assembly of timber joints using reinforcement learning. *Automation in Construction*, 125:103569, 2021.

[21] Jiannan Cai, Ao Du, Xiaoyun Liang, and Shuai Li. Prediction-based path planning for safe and efficient human–robot collaboration in construction via deep reinforcement learning. *Journal of Computing in Civil Engineering*, 37(1):04022046, 2023.

[22] Zuyuan Zhu and Huosheng Hu. Robot learning from demonstration in robotic assembly: A survey. *Robotics*, 7(2), 2018. ISSN 2218-6581. doi:10.3390/robotics7020017. URL https://www.mdpi.com/2218-6581/7/2/17.

[23] Ci-Jyun Liang, Vineet R Kamat, and Carol C Menassa. Teaching robots to perform quasi-repetitive construction tasks through human demonstration. *Automation in Construction*, 120:103370, 2020.

[24] Yan Li, Songyang Liu, Mengjun Wang, Shuai Li, and Jindong Tan. Teleoperation-driven and keyframe-based generalizable imitation learning for construction robots. *Journal of Computing in Civil Engineering*, 38(6):04024031, 2024.

[25] Chuangchuang Sun, Macheng Shen, and Jonathan P How. Scaling up multiagent reinforcement learning for robotic systems: Learn an adaptive sparse communication graph. In *2020 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 11755–11762. IEEE, 2020.

[26] Dong Ki Kim, Miao Liu, Matthew D Riemer, Chuangchuang Sun, Marwa Abdulhai, Golnaz Habibi, Sebastian Lopez-Cot, Gerald Tesauro, and Jonathan How. A policy gradient algorithm for learning to learn in multiagent reinforcement learning. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 5541–5550. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/kim21g.html.

[27] Leon C. Price, Jingdao Chen, Jisoo Park, and Yong K. Cho. Multisensor-driven real-time crane monitoring system for blind lift operations: Lessons learned from a case study. *Automation in Construction*, 124:103552, 2021. ISSN 0926-5805. doi:https://doi.org/10.1016/j.autcon.2021.103552. URL https://www.sciencedirect.com/science/article/pii/S0926580521000030.

[28] Antonin Raffin, Ashley Hill, Adam Gleave, Anssi Kanervisto, Maximilian Ernestus, and Noah Dormann. Stable-baselines3: Reliable reinforcement learning implementations. *Journal of Machine Learning Research*, 22(268):1–8, 2021. URL http://jmlr.org/papers/v22/20-1364.html.

[29] Timothy P. Lillicrap, Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning, 2019. URL https://arxiv.org/abs/1509.02971.

[30] Scott Fujimoto, Herke van Hoof, and David Meger. Addressing function approximation error in actor-critic methods, 2018. URL https://arxiv.org/abs/1802.09477.

[31] Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor, 2018. URL https://arxiv.org/abs/1801.01290.

[32] Marcin Andrychowicz, Filip Wolski, Alex Ray, Jonas Schneider, Rachel Fong, Peter Welinder, Bob McGrew, Josh Tobin, Pieter Abbeel, and Wojciech Zaremba. Hindsight experience replay, 2018. URL https://arxiv.org/abs/1707.01495.

[33] Théo Alves Da Costa. Solving the Traveling Salesman Problem with Reinforcement Learning — Eki.Lab — ekimetrics.github.io. https://ekimetrics.github.io/blog/2021/11/03/tsp/, 2021. [Accessed 14-03-2025].

[34] Seied Mohammad Langari, Faridaddin Vahdatikhaki, and Amin Hammad. Improving the performance of rrt path planning of excavators by embedding heuristic rules. *Advanced Engineering Informatics*, 62:102724, 2024.

[35] Ali Garshasbi, Jun Wang, Jingdao Chen, and Junfeng Ma. Human-robot collaboration in the construction industry: A mini-review. In *Proceedings of the 2nd Future of Construction Workshop at the International Conference on Robotics and Automation (ICRA 2023)*, pages 1–4, London, UK, July 2023. International Association for Automation and Robotics in Construction (IAARC). ISBN -. doi:10.22260/ICRA2023/0003.

Enhancing Robotic Vision through Deep Learning Techniques: From Detection to Construction

Yusuf Aykin[1], **Hans Sachs**[1], **Nikolai Gerzen**[1]

[1]Technische Hochschule Ostwestfalen-Lippe, Germany

yusuf.aykin@th-owl.de, hans.sachs@th-owl.de, nikolai.gerzen@th-owl.de

Abstract -

This paper presents a robotic system developed to enhance automation in construction workflows through advanced AI and computer vision technologies. The system integrates a robotic arm with a 3D point cloud camera and state-of-the-art 2D pre-trained Deep Learning models, such as GroundingDINO and SegmentAnything, to detect and segment construction elements in 3D dynamic, unstructured environments. By processing point cloud data from the camera and aligning it with real-world coordinates, the system achieves precise object localization, enabling tasks such as element handling and assembly. Designed to address challenges like clutter, occlusion, and variability in construction sites, this system bridges the gap between controlled laboratory conditions and real-world applications. Experimental evaluations highlight its potential to improve efficiency and adaptability in construction tasks.

Keywords -

Construction Automation; Robotic Vision; Object Detection; 3D Vision; Deep Learning

1 Introduction

The integration of artificial intelligence (AI) with robotics offers transformative potential for the construction industry. Construction workflows, requiring precision and efficiency, are well-suited for robotic automation, which can address critical challenges like labor shortages, high costs, and safety risks by automating repetitive tasks such as element handling, assembly, and inspection [1]. However, deploying robotics effectively in the chaotic and unpredictable environments of real construction sites remains a significant hurdle.

The transition from controlled laboratory settings to dynamic construction sites presents formidable challenges. Real-world construction environments are inherently complex, characterized by clutter, inconsistent lighting, diverse materials, and unforeseen obstacles. Unlike controlled experiments, these sites involve overlapping objects, occlusions, and constantly changing conditions that can severely limit the effectiveness of conventional robotic systems. While robotic automation has advanced, adapting these systems to the visual complexities of real-world construction remains a critical gap.

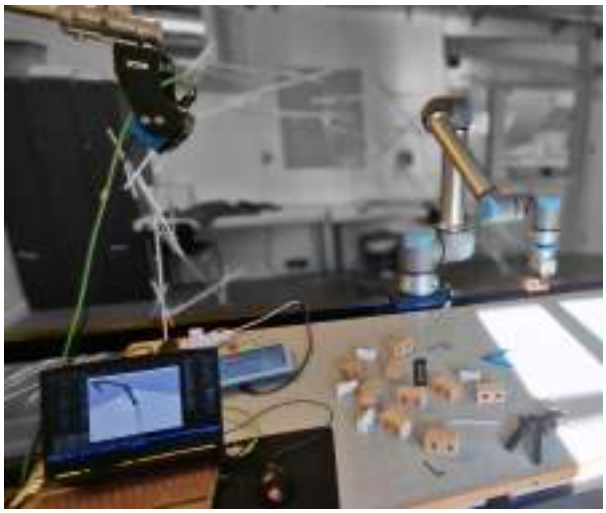


Figure 1. Robotic environment setup for object manipulation in a construction scenario. The system integrates a UR10e robotic arm and a Visionary-S point cloud scanner.

Vision is paramount for robotic automation in these environments. Although 2D image recognition models have advanced significantly [2], they often lack the spatial understanding essential for robots in complex 3D spaces. Trained primarily for 2D analysis, they struggle with depth perception and spatial relationships crucial for manipulation and assembly in cluttered scenes. In contrast, 3D vision systems, capable of capturing detailed spatial data through point clouds, offer a more comprehensive understanding of the environment, enabling robots to perceive and interact with objects more effectively, even under challenging conditions. Our work seeks to leverage the power of these advanced 2D image recognition models and apply them within a 3D context. Figure 1 illustrates our system setup.

Recent AI breakthroughs in object detection and segmentation have bolstered robotic vision. Models such as GroundingDINO and SegmentAnything leverage pre-trained deep learning techniques to detect and segment

unseen objects with high accuracy, even in cluttered, complex environments [3, 2]. These models use transformer networks to process image data, achieving high accuracy and robustness. Additionally, real-time advancements like YOLO-World [4] and EfficientSAM [5] make them suitable for dynamic environments. Extensions of such models for 3D point cloud processing [6] enable robust frameworks for detecting, segmenting, and manipulating objects under real-world constraints. However, directly applying these powerful 2D models within a comprehensive 3D robotic perception system for construction remains largely unexplored.

The integration of robotics and AI in construction has seen significant progress in object detection, manipulation, and automation of tasks. Early methods relied on feature-based approaches such as SIFT and HOG descriptors [7], which struggled with clutter and occlusions. The introduction of deep learning architectures, including transformer-based models, revolutionized object detection and segmentation [8]. Joint vision-language models like GroundingDINO and SegmentAnything demonstrated remarkable performance, enhancing applications in cluttered construction settings. Several successful applications of these techniques in construction-related tasks include ROI-based detection systems for equipment [9] and advanced point cloud segmentation using networks like PointRCNN [10]. Despite these gains, a cohesive framework effectively using 2D AI models within a 3D robotic system for construction tasks is still needed.

Robotic manipulation in construction settings also requires robust planning and control mechanisms. Recent research expanded into reinforcement learning and AI-based control strategies, as shown in [1]. Grasp planning with point clouds, advanced by tools like PointNetGPD [11], and tactile feedback integration [12] have enabled precise manipulation in real-world scenarios. Transformer-based architectures, such as RT-1 and RT-2 [13], have further shown promising results in robotic manipulation through vision-language understanding. However, realizing the potential of these manipulation advances in unstructured construction requires enhanced 3D perception, which our work addresses.

3D vision systems play a vital role in enabling robots to understand and interact with their surroundings. Point cloud data provides detailed spatial information for accurate object localization and manipulation. Enhanced registration techniques, such as Colored ICP [6], incorporate color information valuable in construction scenes. Recent depth estimation methods, such as Depth Anything V2 [14], and fusion techniques like BEVFusion [15], continue to advance 3D perception capabilities. Building on these, our system aims for a more integrated and practical approach for construction robotics by leveraging both 2D AI models and 3D point cloud data.

This paper presents a novel system designed to bridge the gap between controlled lab settings and the challenges of real construction sites. Our core novelty is the effective application of state-of-the-art 2D image recognition models within a 3D point cloud framework. This integration enhances robotic automation in construction by combining 2D AI object recognition with 3D spatial understanding. By equipping a robotic arm with a cloud scanner and advanced object detection and segmentation algorithms, the system achieves:

- Accurate detection and segmentation of unseen construction elements and tools using AI-based pretrained models, originally designed for 2D images, but adapted for 3D point cloud understanding, even in cluttered scenes.
- Precise alignment of 3D data with real-world coordinates for effective robotic manipulation.

The experimental setup, shown in Figure 1, demonstrates the system's ability to adapt to the complexities of real-world construction environments. This work is an initial step in a controlled lab setting mimicking construction scenarios. While simplified, our experiments demonstrate the feasibility and potential of our approach to tackle visual challenges in real-world construction. Future work will validate and extend these findings in more complex field conditions. By addressing challenges such as cluttered scenes, occlusions, and dynamic conditions, the proposed system aims to enhance efficiency and adaptability in construction workflows.

2 System Overview

This section provides a comprehensive overview of the hardware and software components used in the proposed robotic system, focusing on its adaptability to dynamic construction tasks. The integration of advanced AI models with a collaborative robotic arm for efficient execution of operations such as element handling, assembly, and inspection.

2.1 Hardware Setup

Our experimental setup, designed to mimic real-world construction scenarios, comprises a UR10e robotic arm, an OnRobot 2FG7 gripper, a Visionary-S point cloud scanner, and a high-performance workstation. The workspace is set on a 1000 mm x 700 mm table where construction elements are placed.

The UR10e (Universal Robots) offers a 10 kg payload capacity, 1300 mm reach, and ±0.03 mm repeatability, making it suitable for tasks like manipulating construction elements and assembling components. Its collaborative

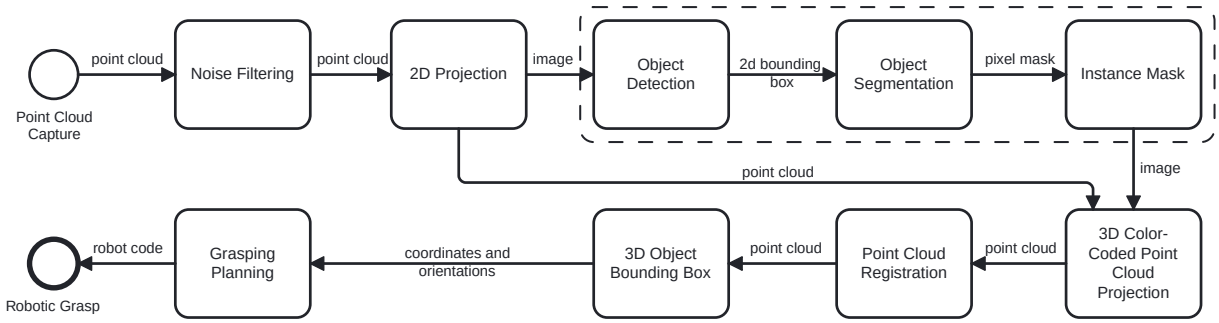


Figure 2. Flowchart of the process: integrating AI-based object detection, point cloud processing, and robotic manipulation.

design, including force sensing, enables safe human-robot interaction. Attached to the UR10e, the OnRobot 2FG7 two-finger gripper provides adjustable gripping force and width for handling various elements, such as bricks and tools.

The SICK Visionary-S point cloud scanner captures high-resolution 3D data of the environment. It boasts a pixel resolution of 640×512, depth accuracy of $\leq 0.25\,\text{mm}$, a detection angle of $60° \times 50°$, and operates within a camera distance of 0.5 to 65 meters. Long distance capture allows us to operate in large construction environments. The scanner is positioned at one end of the workspace to capture the scene from a distance of approximately 1 meter from the table. To handle the computational demands of the AI models and point cloud processing, a high-performance workstation is used. This workstation is equipped with an NVIDIA RTX 3070 GPU, an Intel i7-12700H processor, and 64 GB of RAM, providing the necessary processing power for real-time operations.

2.2 Software Architecture

The software architecture integrates advanced AI models with point cloud processing and robotic control to enable seamless operation. The Visionary-S scanner captures 3D point clouds and RGB images, which undergo preprocessing to enhance data quality. Techniques such as noise filtering using DBSCAN and voxel downsampling reduce the computational load while preserving geometric details, ensuring the accuracy and reliability of the data for subsequent stages.

AI models are employed to process the preprocessed data for object detection and segmentation. These models run within Docker containers to provide consistent and isolated environments. GroundingDINO and SegmentAnything models form the backbone of the detection and segmentation pipeline. The processed data, including the detected objects and their segmentation masks, is then used to calculate target positions and orientations for robotic manipulation. The robot executes these commands through a Python-based control system, which translates high-level instructions into precise joint movements.

2.3 AI Models

The system relies on advanced AI models, GroundingDINO and SegmentAnything, to identify and segment construction elements and components accurately. These models are optimized to handle the diverse and cluttered nature of construction environments to provide reliable performance in a wide range of tasks.

GroundingDINO is used for object detection in 2D images derived from the 3D point cloud. Using its language-guided detection capabilities, the model can identify objects based on natural language prompts, such as “brick” or “wooden block.” This adaptability allows the system to handle a wide variety of construction elements. GroundingDINO generates precise 2D bounding boxes for detected objects, even in complex and cluttered environments, as compared with YOLO-World [4] in Figure 3.

SegmentAnything complements GroundingDINO by providing detailed instance masks for the detected objects. The versatility of the model allows it to segment a diverse range of objects without requiring additional training, which makes it suitable for dynamic construction scenarios. The high-quality instance masks generated by SegmentAnything allow the robot to distinguish between overlapping objects and perform precise manipulations.

3 Methodology

The proposed system integrates advanced AI models with robotic control to achieve accurate detection, segmentation, and manipulation of objects in dynamic construction environments. This section outlines the methodological framework shown in Figure 2 for data acquisition, preprocessing, point cloud registration, AI-driven detection and segmentation, object localization, grasp planning,

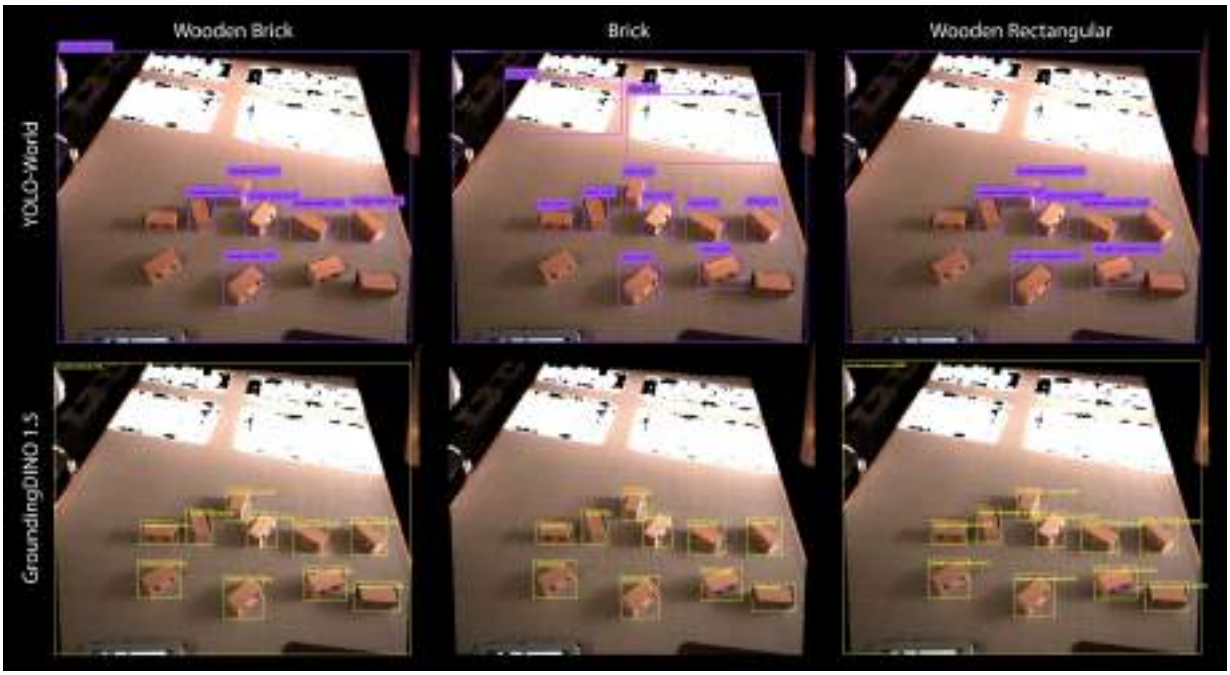


Figure 3. Comparison between GroundingDINO and YOLO-World with different prompts.

and robotic execution.

3.1 Data Acquisition and Preprocessing

The data acquisition process begins with the Visionary-S point cloud scanner capturing 3D data of the construction workspace. The scanner provides a high-resolution depth map and corresponding RGB image. Noise filtering is applied to the raw point cloud $P = \{p_1, p_2, ..., p_n\}$ using the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm. DBSCAN classifies points into core, border, and noise points based on their local density. The ϵ-neighborhood of a point p is defined as:

$$N_\epsilon(p) = \{q \in P \mid \text{dist}(p, q) \leq \epsilon\} \tag{1}$$

where $\text{dist}(p, q)$ is the Euclidean distance between points p and q, and ϵ is a predefined radius. A point p is a core point if its ϵ-neighborhood contains at least minPts points:

$$|N_\epsilon(p)| \geq \text{minPts} \tag{2}$$

A point q is a border point if it is within the ϵ-neighborhood of a core point p but is not a core point itself. A point that is neither a core point nor a border point is classified as a noise point. The filtered point cloud P' consists of core and border points, excluding noise points. The filtered point cloud is then projected onto a 2D image plane using the intrinsic parameters of the Visionary-S scanner. The camera intrinsic matrix K is defined as:

$$K = \begin{matrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{matrix} \tag{3}$$

where f_x and f_y are the focal lengths in the x and y directions, and (c_x, c_y) is the principal point. For a point $p = (x, y, z)$ in the 3D point cloud, its projection $p' = (u, v)$ onto the 2D image plane is given by:

$$\begin{matrix} u \\ v \\ 1 \end{matrix} = K \begin{matrix} x/z \\ y/z \\ 1 \end{matrix}, \quad u = f_x \frac{x}{z} + c_x, \quad v = f_y \frac{y}{z} + c_y \tag{4}$$

This projection provides image format required for object detection. (Figure 4)



Figure 4. Projection from 3d pointcloud to 2d image.

3.2 Point Cloud Registration

Point cloud registration aligns the captured point cloud with a predefined reference frame, enabling accurate localization of objects within the workspace. The Iterative Closest Point (ICP) algorithm is used to refine the alignment between the source point cloud (captured data) and the target point cloud (reference frame). The ICP algorithm iteratively minimizes the difference between two point clouds. Let P_s be the source point cloud and P_t be the target point cloud. The algorithm aims to find the optimal transformation T that aligns P_s with P_t. The algorithm starts with an initial transformation T_0. For each point $p_i \in P_s$, the closest point $q_i \in P_t$ is found. This forms a set of corresponding pairs $C = \{(p_i, q_i)\}$. The transformation T_k that minimizes the mean squared error between the corresponding pairs is computed:

$$E(T) = \frac{1}{N} \sum_{i=1}^{N} ||q_i - T_k p_i||^2 \tag{5}$$

where N is the number of corresponding pairs. The optimal transformation can be found using Singular Value Decomposition (SVD). The source point cloud P_s is transformed using T_k:

$$P'_s = \{T_k p \mid p \in P_s\} \tag{6}$$

These steps are repeated until convergence. For improved registration accuracy the Colored ICP variant is used. Colored ICP incorporates color information in addition to spatial coordinates. The error function for Colored ICP is modified to include a color difference term:

$$E(T) = \frac{1}{N} \sum_{i=1}^{N} \left(||q_i - T_k p_i||^2 + \lambda ||c(q_i) - c(T_k p_i)||^2 \right) \tag{7}$$

where $c(p_i)$ and $c(q_i)$ are the color vectors of the corresponding points, and λ is a weighting factor that balances the geometric and color terms.

3.3 AI-Driven Detection and Segmentation

The preprocessed and registered data is fed into a pipeline comprising GroundingDINO and SegmentAnything (SAM) models. GroundingDINO processes the 2D image projection of the registered point cloud, generating bounding boxes for objects based on language prompts. The output of GroundingDINO is a set of bounding boxes $B = \{b_1, b_2, ..., b_n\}$, where each bounding box b_i is represented by its coordinates $(x_{min}, y_{min}, x_{max}, y_{max})$ and a confidence score s_i. The bounding boxes generated by GroundingDINO are refined using Non-Maximum Suppression (NMS) to eliminate overlapping and redundant detections. The Intersection over Union (IoU) between two bounding boxes b_i and b_j is computed as:

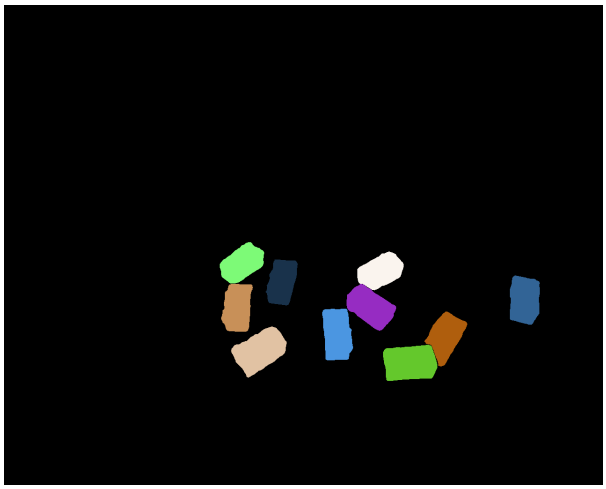


Figure 5. Segmentation mask created with SAM.

$$\text{IoU}(b_i, b_j) = \frac{\text{Area}(b_i \cap b_j)}{\text{Area}(b_i \cup b_j)} \tag{8}$$

If the IoU between two boxes exceeds a predefined threshold, the box with the lower confidence score is suppressed. The refined bounding boxes are passed to the SegmentAnything model, which generates precise instance masks for each detected object (Figure 5). The output of SAM is a set of instance masks $M = \{m_1, m_2, ..., m_n\}$, where each mask m_i corresponds to a detected object and has the same dimensions as the input image. The 2D segmentation masks are mapped back onto the 3D registered point cloud shown in Figure 6. For each point $p = (x, y, z)$ in the point cloud, its corresponding pixel (u, v) in the 2D image is determined using the projection equations derived from the camera intrinsic matrix K. The value of the segmentation mask $m_i(u, v)$ at that pixel is then assigned to the point p as shown in equation (4).

3.4 Object Localization and Grasp Planning

After detection and segmentation, localized objects are analyzed for spatial attributes using Principal Component Analysis (PCA). For a segmented point cloud P_i, the covariance matrix C is computed to determine its principal axes:

$$C = \frac{1}{n-1} \sum_{i=1}^{n} (p_i - \bar{p})(p_i - \bar{p})^T \tag{9}$$

where n is the number of points, p_i is a point, and $\bar{p}$ is the centroid. The eigenvectors of C represent the principal axes, forming the basis for the Oriented Bounding Box (OBB). The extents of the OBB are defined by projecting

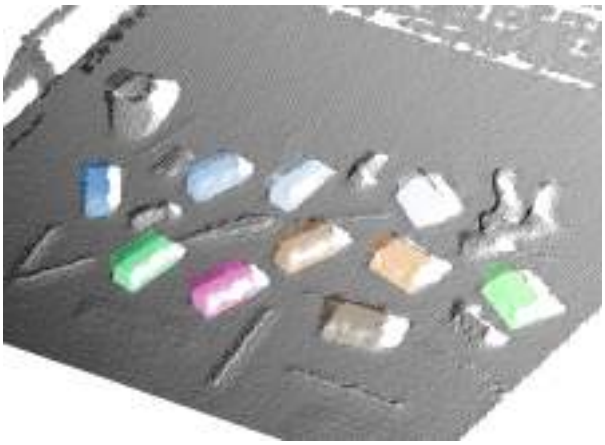


Figure 6. Mapped segmentation mask.

P_i onto the principal axes. Gripping points are selected by evaluating the longest edges of the OBB, with midpoints prioritized as candidate gripping lines:

$$m = \frac{p_j + p_k}{2} \tag{10}$$

where p_j and p_k are endpoints of the edge. Suitability of gripping points is based on gripper dimensions and workspace accessibility. Figure 7 illustrates the OBB and selected gripping locations.



Figure 7. Calculated OBB and grasping locations.

3.5 Robotic Execution

The gripping and placement actions are translated into joint commands for the UR10e robotic arm. Inverse kinematics (IK) computes the joint angles θ required to achieve the desired end-effector pose $\tilde{T}$, ensuring smooth, collision-free movements. The desired pose $\tilde{T}$ is derived from the calculated gripping points and desired placement locations. The relationship between joint angles and end-effector pose is given by:

$$FK(\theta) = \tilde{T} \tag{11}$$

where $FK(\theta)$ represents the forward kinematics function of the robot, mapping joint angles θ to the end-effector pose.

Trajectory planning generates intermediate waypoints between the initial and target poses, optimizing for speed and safety. Smooth joint-space trajectories are interpolated using cubic splines:

$$\theta(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3 \tag{12}$$

where $\theta(t)$ is the joint angle at time t, and a_0, a_1, a_2, a_3 are the spline coefficients.

The trajectory is validated in a PyBullet simulation and converted into robot-specific motion and gripper control commands. These commands are executed by the UR10e robot to perform the desired tasks, such as picking up a detected object and placing it at the target location.

4 Experimental Evaluation and Results

4.1 Experimental Setup

The experiments were carried out in the setup described in the System Overview section 2, representing typical construction scenarios. A total of 150 test scenarios were executed, with varying object arrangements of items like wooden blocks, cables and other common construction items, and environmental conditions to evaluate the robustness and adaptability of the system.

4.2 Evaluation Metrics

The performance of the system was evaluated using the following metrics:

- **Detection Accuracy:** The percentage of correctly detected objects compared to the ground truth, measured for both bounding boxes and segmentation masks.
- **Segmentation Precision:** The Intersection over Union (IoU) between predicted and ground-truth masks, averaged across all detected objects.
- **Manipulation Success Rate:** The percentage of successful grasp-and-place operations out of the total attempts.
- **Processing Time:** The average time required for data acquisition, detection, segmentation, and command generation.

4.3 Results

The system achieved high performance across all evaluated metrics, demonstrating its robustness and adaptability

Table 1. Performance Metrics of the Robotic System

Metric	Average Value
Detection Accuracy	95.8%
Segmentation Precision (IoU)	89.6%
Manipulation Success Rate	93.2%
Average Processing Time	6.3 s

to diverse scenarios. The results, averaged over the 150 test scenarios, are summarized in Table 1.

The integrated use of GroundingDINO and SegmentAnything resulted in an average detection accuracy of 95.8% and an average segmentation precision (IoU) of 89.6%. These results demonstrate the system's ability to accurately detect and segment objects even in cluttered environments with varying object arrangements. Figure 8 illustrates the detection and segmentation outputs for a sample scene, highlighting the precision of bounding boxes and instance masks generated by the AI models.

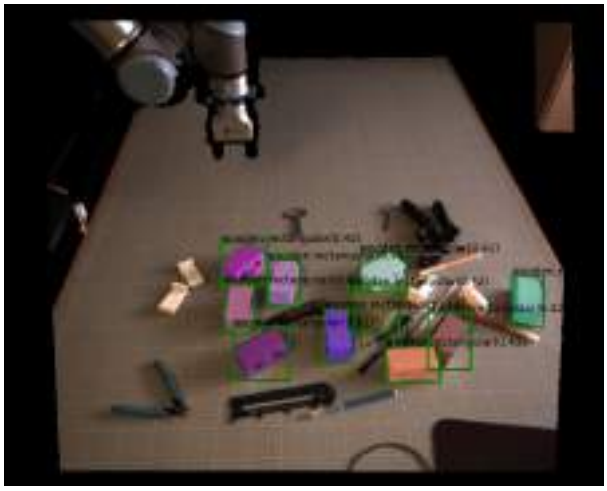


Figure 8. Sample outputs showing object detection and segmentation results.

The robotic arm successfully completed 93.2% of the grasp-and-place tasks across the 150 test scenarios. This high success rate demonstrates the system's capability to handle objects with varying shapes, sizes, and orientations. The few instances of failure were primarily attributed to limitations in the accuracy of the point cloud data for objects located further away from the scanner or those with highly reflective surfaces. These inaccuracies occasionally led to slight errors in the generated bounding boxes, making grasping more challenging.

The average processing time from data acquisition to command execution was 6.3 seconds. This includes 2.1 seconds for point cloud preprocessing and registration, 1.5 seconds for detection and segmentation using GroundingDINO 1.0 and SAM, and 2.7 seconds for object localization, grasp planning, trajectory generation, and command execution. Experiments were also conducted with GroundingDINO 1.5, which resulted in a slightly higher processing time of 9.5 seconds. The increased time is attributed to its API-based object detection approach, which added latency. However, GroundingDINO 1.5 provided a marginal improvement in detection precision (+2.5%).

4.4 Discussion

The lab evaluation demonstrates the system's potential for construction robotics, achieving high accuracy in detection and segmentation using 2D AI models in a 3D context. However, the controlled lab environment differs significantly from complex real-world construction sites, which involve dynamic lighting, dust, and larger scales. Sensor limitations were observed with reflective surfaces and distant objects, suggesting lab performance may not directly translate to real-world scenarios. Processing time, while acceptable for lab demonstrations, could limit real-time applications in dynamic environments. Validation metrics are encouraging for lab experiments, but failure analysis highlights areas for improvement. Manipulation failures were mainly due to perception inaccuracies from degraded point clouds, leading to grasping errors. Even with accurate detection, subtle 3D localization errors impacted grasping, emphasizing manipulation's dependence on robust perception in challenging conditions. Qualitatively, compared to methods like PointNetGPD, [11], RT-2 [13], PointRCNN [10] our 2D AI-in-3D approach, using pre-trained GroundingDINO and SAM, offers a balance of generalization and performance. It leverages 2D vision advancements but involves a pipeline that may introduce latency. Quantitative comparisons are needed to validate our approach against state-of-the-art methods. The successful manipulation rate in the lab confirms perception-to-action feasibility in a controlled setting.

5 Conclusion and Future Works

This work demonstrates a robotic system for object detection, segmentation, and manipulation in a controlled lab, showing promise for construction automation. Leveraging 2D AI models like GroundingDINO [3] and SegmentAnything [2] for 3D vision was effective in simulations, and manipulation success validated grasp planning strategies under these conditions. Future efforts will prioritize robustness, versatility, and real-time performance for real-world deployment. Rigorous virtual testing using game engines and datasets is crucial to assess robustness and bridge the gap to real-world validation. Future research should generalize AI models to diverse construction elements and tools, exploring adaptation strategies.

Improving robustness requires better sensing solutions, including sensor fusion and advanced point cloud processing, to address challenges like reflective surfaces and distance. Future directions include integrating advanced depth estimation models like Depth Anything V2 [14], exploring tactile sensing [16] for dexterous manipulation, and optimizing efficiency for real-time performance. Expanding to multi-robot collaboration using multi-RGBD camera setups [6], addressing occlusion, and investigating intuitive human-robot interaction via AR [17], NLP, and voice commands using reasoning LLMs like OpenAI's o1 [18] are also key. Furthermore, continual learning will be explored for autonomous adaptation. Building upon the findings of this research, our next step involves implementing the system on a larger KUKA robot and deploying it in a real construction site environment to further evaluate its performance and address real-world challenges. Pursuing these paths aims to create a highly adaptable and efficient robotic solution to enhance construction productivity, safety, and flexibility, enabling broader robotic automation and transforming building practices.

References

[1] Sulabh Kumra, Shirin Joshi, and Ferat Sahin. Antipodal robotic grasping using generative residual convolutional neural network. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 9626–9633, 2020. doi:10.1109/IROS45743.2020.9340777.

[2] Alexander Kirillov, Eric Mintun, Nikhila Ravi, et al. Segment anything. *arXiv preprint arXiv:2304.02643*, 2023. doi:10.48550/arXiv.2304.02643.

[3] Shilong Liu, Zhaoyang Zeng, Tianhe Ren, et al. Grounding dino: Marrying dino with grounded pre-training for open-set object detection. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 38–55, 2024. doi:10.1007/978-3-031-72970-6_3.

[4] Tianheng Cheng, Lin Song, Yixiao Ge, et al. Yolo-world: Real-time open-vocabulary object detection. *arXiv preprint arXiv:2401.17270*, 2024. doi:10.48550/arXiv.2401.17270.

[5] Yunyang Xiong, Bala Varadarajan, Lemeng Wu, et al. Efficientsam: Leveraged masked image pretraining for efficient segment anything. *arXiv preprint arXiv:2312.00863*, 2023. doi:10.48550/arXiv.2312.00863.

[6] Yunhan Yang, Xiaoyang Wu, Tong He, et al. Sam3d: Segment anything in 3d scenes, 2023.

[7] Richard Szeliski. *Computer Vision: Algorithms and Applications*. Springer, 2nd edition, 2022. doi:10.1007/978-3-030-34372-9.

[8] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, et al. End-to-end object detection with transformers. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 213–229, 2020. doi:10.1007/978-3-030-58452-8_13.

[9] Hanbo Zhang, Xuguang Lan, Site Bai, et al. Roi-based robotic grasp detection for object overlapping scenes. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019. doi:10.1109/IROS40897.2019.8967869.

[10] Shaoshuai Shi, Xiaogang Wang, Hongsheng Li, et al. Pointrcnn: 3d object proposal generation and detection from point cloud. pages 770–779, 2019. doi:10.1109/CVPR.2019.00086.

[11] Hongzhuo Liang, Xiaojian Ma, Shuang Li, et al. Pointnetgpd: Detecting grasp configurations from point sets. 2019. doi:10.1109/ICRA.2019.8794435.

[12] Yunhai Han, Kelin Yu, Rahul Batra, et al. Learning generalizable vision-tactile robotic grasping strategy for deformable objects via transformer. *IEEE/ASME Transactions on Mechatronics*, 2024. doi:10.1109/TMECH.2024.3400789.

[13] Anthony Brohan, Noah Brown, Justice Carbajal, et al. Rt-2: Vision-language-action models transfer web knowledge to robotic control. *arXiv preprint arXiv:2307.15818*, 2023. doi:10.48550/arXiv.2307.15818.

[14] Lihe Yang, Bingyi Kang, Zilong Huang, et al. Depth anything v2: Accurate depth maps using a single rgb image. 2024. doi:10.48550/arXiv.2406.09414.

[15] Zhijian Liu, Haotian Tang, Alexander Amini, et al. Bevfusion: Multi-task multi-sensor fusion with unified bird's-eye view representation. 2023. doi:10.1109/ICRA48891.2023.10161234.

[16] Georgia Chalvatzaki, Nikolaos Gkanatsios, Petros Maragos, et al. Orientation attentive robotic grasp synthesis with augmented grasp map representation. 2020. doi:10.1109/IROS45743.2020.9341200.

[17] Konstantinos Bousmalis, Alex Irpan, Paul Wohlhart, et al. Using simulation and domain adaptation to improve efficiency of deep robotic grasping. 2018. doi:10.1109/ICRA.2018.8460875.

[18] OpenAI. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774v6*, 2024.

Automated Helmet Detection in Construction Sites using UWB and Machine Learning

Amir Shahbazi Ojghaz[1], Sayeh Bayat[1], and Farnaz Sadeghpour[1]

[1]Schulich School of Engineering, University of Calgary, Canada
amir.ojghaz@ucalgary.ca, sayeh.bayat@ucalgary.ca, farnaz@ucalgary.ca

Abstract

Ensuring the proper use of personal protective equipment (PPE), particularly helmets, is crucial for enhancing safety on construction sites. This study proposes a novel approach for detecting helmet usage and worker states using ultra-wideband (UWB) localization sensors combined with machine learning algorithms. Unlike traditional sensor-based or image-based systems, the proposed method integrates helmet detection into existing proximity warning systems, offering a cost-effective solution without the need for additional hardware. Data was collected from 12 participants in a controlled environment using UWB sensors, and a machine learning model was developed to classify four worker states: standing with a helmet, standing while holding a helmet, walking with a helmet, and walking while holding a helmet. The Gradient Boosting Decision Trees (GBDT) algorithm was selected for model development due to its superior performance. The model achieved an overall accuracy of 74.46% and performed well in detecting unsafe conditions. However, variability was observed in identifying safe states. Additionally, the study explored the impact of worker height on Z-axis localization error, revealing a correlation that suggests the need for height-adjusted safety monitoring systems. This research demonstrates that UWB-based systems can enhance PPE monitoring, reduce computational costs, and address privacy concerns, while also highlighting areas for future improvement, such as expanding the model to detect other PPE and refining its ability to differentiate between worker postures.

Keywords

Personal Protective Equipment (PPE); Construction Safety; Ultra-wideband Technology (UWB); Machine Learning.

1 Introduction

Construction sites are among the most hazardous environments for workers, with many sustaining injuries or losing their lives in workplace accidents every year. Numerous solutions have been proposed to enhance safety on construction sites, one of which is the use of personal protective equipment (PPE). The use of PPE can significantly reduce workplace risks and fatal accident [1, 2]. Previous studies have demonstrated that safety behaviors, including the consistent use of PPE, are strongly associated with a reduction in work-related injuries. Conversely, unsafe behaviors, such as the failure to wear PPE, are key risk factors in workplace accidents [3].

One of the challenges related to construction worker safety is the ineffective use of PPE. Factors such as discomfort, inadequate training, and poor communication between safety managers and workers can hinder the effective use of PPE. As a result, one of the main responsibilities of safety managers is to ensure that workers are wearing PPE on site. However, due to the dynamic and expansive nature of construction sites, manual supervision of PPE usage by safety managers is often inefficient. Consequently, efforts have been made to develop automated solutions for monitoring PPE usage by workers. Solutions include wearable devices, computer vision, and deep learning models [4-6].

These technologies enable real-time monitoring of PPE usage and can proactively warn workers of potential safety hazards, ensuring that they are properly wearing their safety gear, thereby reducing the likelihood of accidents. However, the adoption of these technologies is not without challenges [7-9]. For instance, the use of cameras on construction sites, in addition to the high costs associated with purchasing and processing equipment, raises concerns about worker privacy.

One technology used to improve worker safety is real-time localization through Ultra-wideband (UWB) sensors. This technology is typically employed in proximity warning systems to locate workers on construction sites and provide alerts when they approach hazardous areas. While these systems generally offer accurate two-dimensional localization of workers, three-dimensional localization, which involves the z-axis, often results in significant errors [10]. Consequently, using the z-axis for tasks that require high precision may not yield

satisfactory results. However, if machine learning can be used to detect helmet usage based on the relatively imprecise z-axis data from these systems, it would be possible to extend their functionality without additional equipment.

Therefore, the objective of this research is to develop a model for detecting helmet usage by workers on construction sites using machine learning and UWB technology. By developing this model, it will be possible to monitor both the location of workers and their PPE status using current proximity warning systems without the need for additional equipment. This method not only preserves worker privacy but also reduces the financial and computational costs associated with system development.

2 Literature Review

Given the importance of safety in construction sites, numerous previous studies have proposed various solutions to ensure the appropriate use of PPE by workers. These solutions typically follow two main approaches: sensor-based methods and image processing techniques. Sensor-based studies often use proximity sensors, such as Bluetooth and RFID, to monitor the use of PPE [8, 11-13]. However, due to the limited applicability of these solutions, using such equipment solely to track PPE usage can increase initial costs. Another limitation of these methods is their inability to detect individuals outside of the predefined system. Moreover, the need for separate hardware to address different safety-related applications ultimately reduces the feasibility of implementing these solutions in construction sites.

On the other hand, image processing and deep learning models have gained significant attention in recent years. In this line of research, scholars typically collect images from construction site environments and process them using deep learning models. These models first detect target equipment, such as helmets and vests, and then assess whether workers are wearing the equipment by analyzing the relationship between the worker and the detected gear [4, 6, 9, 14-16]. Researchers have worked to incorporate equipment of various colors into their models and evaluate the performance of different models, such as YOLO and YOLACT-based approaches. However, this method also presents challenges.

One such challenge is the dependence of cameras on direct line of sight. Additionally, using cameras to monitor workers on construction sites raises concerns about worker privacy. Furthermore, many studies that employ this approach often integrate several deep learning models to achieve their goals, which significantly increases computational load. As a result, to achieve near-real-time performance, costly hardware is required to run the models. Considering the advantages and disadvantages of these approaches, there is a need to develop a model that retains the strengths of both methods while addressing their associated challenges.

3 Methodology

This research followed a three-step process to achieve its objective of developing a model for detecting helmet usage in construction sites using UWB sensors. These steps included data collection, machine learning model training, and performance evaluation. In the following sections, the scope of the research will first be outlined, followed by a detailed description of each step in the methodology.

3.1 Research Scope

Construction workers engage in various tasks on site, during which wearing a helmet is mandatory. Since this study aims to explore the use of UWB sensors and machine learning models for helmet usage detection, four possible worker states were considered: (1) standing while wearing a helmet (CL1), (2) standing while holding a helmet (CL2), (3) walking while wearing a helmet (CL3), and (4) walking while holding a helmet (CL4). These states represent typical conditions encountered in construction sites. Accordingly, the necessary data for developing a machine learning model to differentiate between these states were collected, followed by the model's development.

3.2 Data Collection

Data collection was conducted in a controlled laboratory environment. UWB Real-Time Location Systems (RTLS) generally consist of anchors and tags, where the anchors serve as reference points and the tags represent moving objects. In this study, the DWM1001C 6.5 GHz transceiver was used, which can function as both tags and anchors. Four anchors were placed at a height of 1.9 meters, covering a square area of 7 by 7 meters. Additionally, a tag was attached to a helmet, which was used for data collection in this study (Figure 1c).

To gather the data, participants were asked to perform each of the states outlined in the research scope for a duration of two minutes, during which their localization data, including x, y, and z coordinates, were recorded. The UWB modules were set to a frequency of 10 Hz, ensuring that data points were collected at 10 readings per second. Prior to data collection, participants' demographic information, such as height and gender, was recorded. A total of 12 participants took part in the experiment, resulting in the collection of 66,446 data records. Their collective demographic information is presented in Table 1.



Figure 1. (a) Laboratory setup with UWB anchors. (b) UWB anchor at 1.9 m. (c) Helmet with UWB tag.

Table 1. Participant demographics: gender, count, height stats.

Gender	Count	Avg. Height (m)	Min. Height (m)	Max. Height (m)	Std. of Height
F	3	1.64	1.62	1.69	0.03
M	9	1.80	1.68	1.93	0.07
Total	12	1.76	1.62	1.93	0.09

3.3 Data Preprocessing

3.3.1 Data Cleaning

As previously mentioned, the aim of using machine learning algorithms in this study to detect helmet usage by workers is addressing the significant error in the z-axis measurements provided by UWB sensors. Therefore, outlier removal must be handled with extra care, as removing a substantial portion of the data could limit the model's generalizability. Additionally, given the wide range of data values due to variations in participant heights, removing outliers based on aggregated user data could introduce bias into the model. To avoid this, the Isolation Forest [17] algorithm was applied individually to each participant's data, with a contamination rate of 2% used to identify and remove outliers. The data preprocessing was implemented using Python 3.8.9 with Scikit-learn 1.3.2 and Pandas 2.0.3. After performing outlier detection and cleaning, the total number of data records decreased from 66,446 to 64,380.

3.3.2 Feature Engineering

Feature selection and extraction are critical steps in developing a machine learning model when working with location data collected in a laboratory environment. Since the laboratory space is limited, the data range in all three dimensions is also constrained, which can lead to overfitting. In such cases, the model may associate different states with the participants' specific locations in the lab, rather than identifying meaningful relationships between the data points. For instance, if participants remain in one specific area of the lab while performing a particular state, the model might incorrectly link that state to those specific coordinates. However, this would not hold true in real-world construction sites.

To mitigate overfitting, instead of using the raw x and y coordinates, the changes in these coordinates relative to previous steps were used as features for model training. The specific time intervals (10, 50, 100, and 150 steps) were selected based on temporal considerations. Since the UWB modules were set to a frequency of 10 Hz, each 10 steps correspond to approximately 1 second of data collection. The selected intervals therefore represent time windows of approximately 1, 5, 10, and 15 seconds, respectively. This range was chosen to balance the trade-off between model accuracy and real-time performance capabilities. Shorter intervals (e.g., 10 steps) capture immediate movements but may miss longer patterns, while longer intervals (e.g., 150 steps) capture more comprehensive movement patterns but increase the latency of detection in real-time applications. Table 2 lists the features selected for developing the machine learning model. It should be noted that, in addition to the rationale mentioned above, these features were also chosen based on their significance as determined through feature importance analysis.

3.4 Machine Learning Model Development

3.4.1 Algorithm Selection

The model training process involved training multiple models based on different machine learning algorithms, evaluating the performance of each, selecting the best-performing algorithm, and subsequently optimizing the chosen model and reporting its performance. In the initial selection of algorithms, a range of options was considered, from less complex algorithms such as SVM to more complex ones like deep learning algorithms. The rationale behind this approach was to identify an algorithm with lower computational cost and greater transparency. All models were implemented using Scikit-learn 1.3.2, with the Gradient Boosting Decision Trees (GBDT) implementation utilizing XGBoost 2.1.1. Additional libraries used in the analysis include NumPy 1.24.4 and Seaborn 0.13.2 for data manipulation and visualization. After evaluating the models developed in the first step, the GBDT algorithm was selected for further development and optimization due to its superior performance compared to other algorithms. A list of the evaluated algorithms and their performance is presented in Table 3.

Table 2. Selected features for ML model to mitigate overfitting

Feature	Description
z	Z coordinate values
z_norm	Z coordinate values normalized by the participant's height
dz_10	Change in Z coordinate compared to the value 10 steps earlier
dist_10	Distance change compared to the position 10 steps earlier (using X, Y, and Z)
dz_norm_10	Normalized Z changes compared to 10 steps earlier
dz_50	Change in Z coordinate compared to the value 50 steps earlier
dz_norm_50	Normalized Z change compared to 50 steps earlier
dist_50	Distance change compared to the position 50 steps earlier (using X, Y, and Z)
dz_100	Change in Z coordinate compared to the value 100 steps earlier
dz_norm_100	Normalized Z changes compared to 100 steps earlier
dist_100	Distance change compared to the position 100 steps earlier (using X, Y, and Z)
dz_150	Change in Z coordinate compared to the value 150 steps earlier
dz_norm_150	Normalized Z changes compared to 150 steps earlier
dist_150	Distance change compared to the position 150 steps earlier (using X, Y, and Z)

Table 3. Performance metrics of ML algorithms.

Model	AUC	F1	Prec.	Recall
Random Forest	0.945	0.712	0.721	0.724
SVM	0.763	0.398	0.415	0.395
Decision Tree	0.803	0.671	0.674	0.682
AdaBoost	0.773	0.626	0.629	0.635
Logistic Regression	0.928	0.585	0.674	0.615
kNN	0.843	0.489	0.49	0.491
Gradient Boosting	0.946	0.717	0.731	0.728
Naive Bayes	0.891	0.571	0.596	0.596

3.4.2 Model Training and Evaluation

The XGBoost library was used to implement the GBDT algorithm. One critical aspect to consider was the division of the available data for training, validation, and testing of the model's performance. In this study, instead of merging the data from all participants and splitting it based on a percentage of the total data, the data were divided based on the number of participants, without merging. Specifically, data from 6 out of the 12 participants were used for training the model, 3 participants' data were used for testing, and 3 participants' data were reserved for validation during training.

The model training process involved optimizing the model's parameters to achieve the best possible results. This was done using a Grid Search approach with 5-fold cross-validation to find the optimal hyperparameters. The parameter grid explored ranges for learning rate (0.05-1), max depth (3-9), number of estimators (50-150), subsample ratio (0.5-0.9), column sample by tree (0.5-1), and gamma (0-0.1). The XGBoost model was configured to utilize GPU acceleration via CUDA, and hyperparameter tuning was performed by selecting the parameter combination that achieved the highest accuracy. The training and optimization process was performed on an Nvidia RTX 4070 Ti GPU. Once the optimal parameters were obtained, the model was trained and tested 100 times to assess its performance. In each iteration, the participants were randomly assigned to training, testing, and validation sets. The results of the model evaluation will be discussed in the next section.

4 RESULTS

After repeating the training and evaluation process 100 times, the model's performance for each class was recorded in every iteration. The average performance metrics from these evaluations can be seen in Table 4 and Figure 2. The overall accuracy of the model, considering all classes, was 0.744. The developed model demonstrated satisfactory capability in distinguishing between safe and unsafe conditions. In Classes 2 and 4, where workers are not wearing helmets and are thus considered in an unsafe state, the model performed well. The average accuracy for Class 2 was 69%, with a precision of 74.2%, while for Class 4, the accuracy was 83%, and the precision was 86.7%. These results indicate that the model effectively identifies unsafe conditions.
However, the model's performance was more variable for Classes 1 and 3, where workers are wearing helmets and are considered to be in a safe state. Class 1 achieved an average accuracy of 66.6% and a precision of 66.4%, while Class 3 had a higher accuracy of 79.1%. Overall, the model was able to adequately differentiate between safe and unsafe conditions, but improvements in accurately detecting safe workers could help reduce false negatives and enhance the system's overall accuracy.

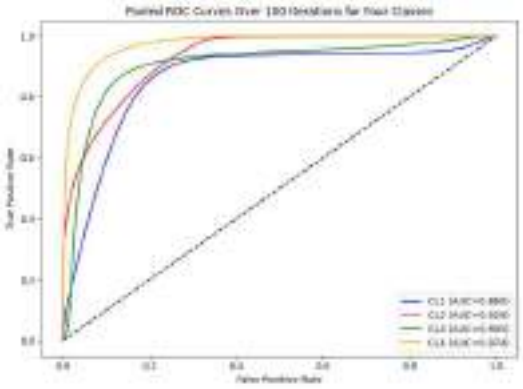


Figure 2. One-vs-Rest ROC curve showing AUC values across four classes.

Table 4. Performance metrics for each class.

Class	Acc.	Precision	Recall	F1	AUC
CL1	0.666	0.664	0.666	0.652	0.860
CL2	0.690	0.742	0.690	0.696	0.929
CL3	0.791	0.750	0.791	0.766	0.905
CL4	0.830	0.867	0.830	0.845	0.974

As the model demonstrated, it not only performed well in distinguishing safe conditions—where workers are wearing helmets—from unsafe ones, but it also showed the ability to differentiate between various worker states, such as standing or walking. Based on this capability, it was decided to assess the model's performance with the addition of another class. A new class was introduced, which involved the worker sitting on a chair while wearing a helmet. The entire training and testing process was repeated to evaluate the model's performance with this additional class. The results of this extended evaluation can be found in Table 5 and Figure 3.

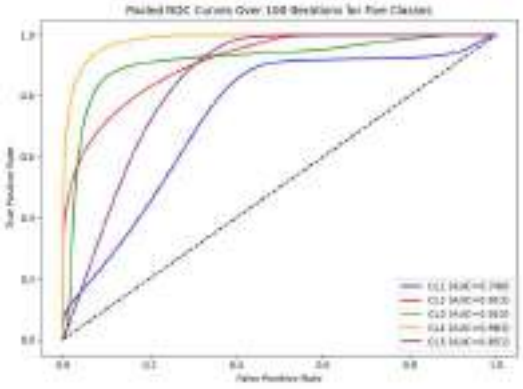


Figure 3. One-vs-Rest ROC curve showing AUC values across five classes.

Table 5. Performance metrics for five classes.

Class	Acc.	Precision	Recall	F1	AUC
CL1	0.318	0.415	0.318	0.341	0.740
CL2	0.673	0.733	0.673	0.684	0.913
CL3	0.817	0.792	0.817	0.801	0.915
CL4	0.852	0.888	0.852	0.868	0.984
CL5	0.601	0.474	0.601	0.521	0.851

5 DISCUSSION

5.1 Advantages and Limitations of the Proposed UWB and Machine Learning Approach

This study aimed to develop a model that not only detects whether workers are wearing helmets but also identifies various worker states using UWB RTLS, which are commonly used for tracking workers and issuing warnings when they approach hazardous areas or to prevent collisions. Machine learning algorithms were employed to achieve this goal.

The solution presented in this study offers both advantages and limitations compared to previous research. One of the key advantages of this approach over other sensor-based methods is that the UWB RTLS are not limited solely to PPE detection. Because this method can be integrated as an additional layer of safety monitoring within proximity warning systems, due to its software-based nature, it addresses the challenge of high initial costs typically associated with sensor-based methods. Moreover, the proposed model demonstrated that this approach can also be used to recognize different worker gestures, which could provide additional information to safety systems, facilitating their further development.

Additionally, the proposed model, being independent of cameras, avoids the disadvantages of image processing-based methods, such as privacy concerns, reliance on direct line of sight, and heavy computational loads. This independence allows for near-real-time performance in the dynamic environments of construction sites. However, it is important to note that this study focused solely on detecting helmet usage. In contrast, image processing methods can potentially detect the use of various other safety equipment, such as vests.

Moreover, the proposed model was tested in a controlled laboratory environment. However, previous studies have demonstrated that the accuracy of UWB RTLS can be impacted by various factors present in real construction sites, such as metal structures, equipment, and environmental conditions. Therefore, it is important

to validate the developed model in construction environments in future studies. Future work should also extend this model to include the detection of additional PPE to further enhance its applicability.

5.2 Impact of Participant Height on Z-Axis Localization Error

As demonstrated in the results section, while the models presented in this study are capable of distinguishing between safe conditions—indicated by helmet usage—and unsafe conditions, this capability fulfills the primary research objective of providing a model for differentiating these states using UWB. However, to improve the model's ability to differentiate between similar states, it is essential to identify the root cause of the existing error. One potential factor contributing to the increase in sensor error in the 3D space is the placement of the tags and anchors on the same plane (at the same height), which can reduce localization accuracy along the Z-axis.

To investigate this issue further, localization errors were evaluated based on the participants' heights. As shown in Figure 4, as the participants' height increases, the absolute localization error along the Z-axis also increases. The correlation between participant height and absolute error is 0.75, and the statistical significance of this relationship, with a p-value of 0.0052, suggests that as participants' height approaches the height at which the anchors are placed (1.9 meters), the error significantly increases. This finding is important because individuals with varying heights experience differing levels of error, and using the same system for individuals of different heights would result in varying safety levels. Therefore, it is necessary to adopt more flexible approaches, such as machine learning models, that can account for these variations.

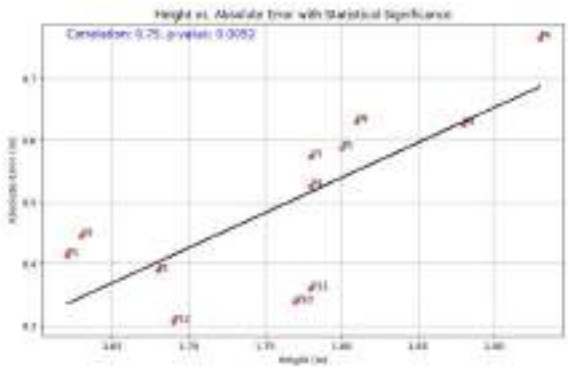


Figure 4. Correlation between participant height and Z-axis error.

6 Conclusion

This study successfully demonstrated the feasibility of using UWB sensors and machine learning algorithms to detect helmet usage and differentiate between various worker states on construction sites. The proposed approach offers significant advantages over existing sensor- and image-based methods by integrating PPE detection into existing proximity warning systems, thereby reducing costs and addressing privacy concerns. The model proved capable of accurately identifying unsafe conditions, though further improvements are necessary to enhance the detection of safe worker states, particularly in distinguishing between specific postures. Additionally, this study highlighted the impact of worker height on localization errors, suggesting that future systems should account for this factor to improve overall performance.

Future research should focus on several directions. First, expanding the model to detect other types of PPE, such as safety vests, would enhance its applicability in construction safety monitoring. Second, field validation in real construction site environments is needed to test the solution's reliability and performance under authentic conditions. Finally, as this study built its solution on top of existing proximity warning systems to decrease adoption costs, future studies can explore this approach to expand current safety systems and increase the practicality of proposed solutions for different applications in construction.

References

[1] Sawacha, E., S. Naoum, and D. Fong, *Factors affecting safety performance on construction sites.* International Journal of Project Management, 1999. **17**(5): p. 309-315.

[2] Li, J., et al., *Evaluating the impact of mental fatigue on construction equipment operators' ability to detect hazards using wearable eye-tracking technology.* Automation in Construction, 2019. **105**: p. 102835.

[3] Wong, T.K.M., S.S. Man, and A.H.S. Chan, *Critical factors for the use or non-use of personal protective equipment amongst construction workers.* Safety Science, 2020. **126**: p. 104663.

[4] Lee, Y.-R., et al., *Deep learning-based framework for monitoring wearing personal protective equipment on construction sites.* Journal of Computational Design and Engineering, 2023. **10**(2): p. 905-917.

[5] Han, K. and X. Zeng, *Deep Learning-Based Workers Safety Helmet Wearing Detection on Construction Sites Using Multi-Scale Features.* IEEE Access, 2022. **10**: p. 718-729.

[6] Ahmed, M.I.B., et al., *Personal Protective Equipment Detection: A Deep-Learning-Based Sustainable Approach.* Sustainability, 2023. **15**(18): p. 13990.
[7] Chen, H., et al., *The Impact of Wearable Devices on the Construction Safety of Building Workers: A Systematic Review.* Sustainability, 2023. **15**(14): p. 11165.
[8] Yang, X., et al., *Automated PPE-Tool pair check system for construction safety using smart IoT.* Journal of Building Engineering, 2020. **32**: p. 101721.
[9] Nath, N.D., A.H. Behzadan, and S.G. Paal, *Deep learning for site safety: Real-time detection of personal protective equipment.* Automation in Construction, 2020. **112**: p. 103085.
[10] Maalek, R. and F. Sadeghpour, *Accuracy assessment of Ultra-Wide Band technology in tracking static resources in indoor construction scenarios.* Automation in Construction, 2013. **30**: p. 170-183.
[11] *Automated PPE Misuse Identification and Assessment for Safety Performance Enhancement*, in *ICCREM 2015*. p. 204-214.
[12] Barro-Torres, S., et al., *Real-time personal protective equipment monitoring system.* Computer Communications, 2012. **36**(1): p. 42-50.
[13] Kelm, A., et al., *Mobile passive Radio Frequency Identification (RFID) portal for automated and rapid control of Personal Protective Equipment (PPE) on construction sites.* Automation in Construction, 2013. **36**: p. 38-52.
[14] Wang, Z., et al., *Fast Personal Protective Equipment Detection for Real Construction Sites Using Deep Learning Approaches.* Sensors, 2021. **21**(10): p. 3478.
[15] Lo, J.-H., L.-K. Lin, and C.-C. Hung, *Real-Time Personal Protective Equipment Compliance Detection Based on Deep Learning Algorithm.* Sustainability, 2023. **15**(1): p. 391.
[16] Chen, S. and K. Demachi, *Towards on-site hazards identification of improper use of personal protective equipment using deep learning-based geometric relationships and hierarchical scene graph.* Automation in Construction, 2021. **125**: p. 103619.
[17] Liu, F.T., K.M. Ting, and Z.H. Zhou. *Isolation Forest*. in *2008 Eighth IEEE International Conference on Data Mining*. 2008.

Advanced Vision-guided Robotic Tying with 6-DoF Rebar Pose Estimation

Mi Liu[1], Jingjing Guo[1,2], Lu Deng[1,2,*], Xiaoyi Lyu[1], Linzhen Nie[3]

[1]College of Civil Engineering, Hunan University, China
[2]Key Laboratory of Damage Diagnosis for Engineering Structures of Hunan Province, Hunan University, China
[3]School of Automotive Engineering, Wuhan University of Technology, China
*Corresponding Author
liumi@hnu.edu.cn, guojingjing@hnu.edu.cn, *denglu@hnu.edu.cn, lyuxiaoyi@hnu.edu.cn, linzhen_nie@whut.edu.cn

Abstract –

Rebar tying is a time-consuming and labor-intensive process that involves repetitive bending and hand motions to secure rebar with wire, often leading to muscular and skeletal injuries. To address these challenges, rebar tying robots have been developed to automate the process. However, existing studies primarily focus on tying point localization for horizontal rebar mesh, neglecting the 6 Degrees of Freedom (DoF) pose estimation necessary for tying complex reinforcement skeletons where rebar may be arranged in various spatial orientations. Therefore, this study proposes a 6-DoF rebar pose estimation method based on keypoint detection and point cloud registration. A deep learning-based keypoint detection algorithm is employed to extract point cloud data from each rebar intersection. Additionally, to tackle challenges posed by texture-less and self-occluding features of rebar point clouds, a coarse registration method utilizing geometric features and a fine registration method based on adaptive displacement correction is introduced to ensure accuracy and stability in rebar pose estimation. The proposed method enables the robot to autonomously understand the relative spatial relationship between itself and the tying objects, facilitating precise tying across a broader range of reinforcement skeletons, thereby significantly reducing the demand for manual labor.

Keywords –
Rebar tying; Rebar pose estimation; Keypoint detection; Point cloud registration; Reinforcement skeleton.

1 Introduction

Rebar tying is a crucial process in reinforced concrete construction. Its primary purpose is to secure the position of rebar, ensuring that it remains correctly positioned according to the design drawings during concrete pouring. However, manual rebar tying is a labor-intensive task that requires workers to bend repeatedly and tie each rebar crosspoint by hand [1].

To overcome this problem, vision-guided robotic systems have been applied to automate rebar tying tasks in recent years. For example, Tybot [2], a gantry-mounted tying robot shown in Figure 1 (a), navigates along the screed or porta-rail and uses a dual-camera system to extract tying information for bridge deck operations. Similarly, the rebar tying robot developed by the Beijing Institute of Technology [3], as shown in Figure 1 (b), moves steadily in four directions across the rebar mesh and employs a vision system to detect crosspoints and guide the tying process. A similar robot, T-iROBO [4], is shown in Figure 1 (c). In both types of robots, the vision system plays a critical role in tying operations by extracting spatial information about tying points through visual perception technology, thereby guiding the robot to complete tying tasks efficiently.



(a) (b) (c)

Figure 1. Rebar tying robots: (a) Tybot [2], (b) the rebar tying robot developed by the Beijing Institute of Technology [3], (c) T-iROBO [4].

For tying information extraction in the vision systems of rebar tying robots, Jin et al. [3] proposed using a deep learning-based keypoint detection algorithm to identify pixel coordinates of the rebar crosspoint and then calculate the rebar crosspoint's spatial coordinates by integrating the depth value of the point. Similarly, Cheng and Deng [5] introduced a rebar crosspoint recognition

method based on MobileNetV3 and the SSD algorithm. Previous research on rebar tying robots has primarily focused on detecting rebar crosspoints, enabling the robotic arm to perform three linear motions to position the end effector at the tie point. However, estimating the tying orientation has been less explored, despite its importance in guiding the end effector to rotate to the appropriate angle. This ensures that the tying wire correctly wraps around the rebar and facilitates subsequent tying operations. In construction practice, rebar planes are often positioned in various orientations on construction sites, such as the reinforcement skeletons of beams and walls, which can be vertical or inclined as well as horizontal. In these scenarios, a robotic arm with at least 6 Degrees of Freedom (DoF) is typically required to perform 6-DoF tying operations, enabling flexible positioning and orientation to meet complex rebar tying requirements. Consequently, merely detecting tying position information in previous research is insufficient for addressing rebar tying tasks in reinforcement skeletons, and it is desired to develop advanced vision perception methods to enable rebar tying robots to adapt to various spatial configurations and perform 6-DoF tying operations effectively.

Object pose estimation plays a crucial role in 6-DoF robotic manipulation, enabling robots to determine the spatial position and orientation of objects for accurate task execution [6]. Traditional methods determine the object pose by aligning the template point cloud with the scene point cloud using point cloud registration techniques [7]. The Iterative Closest Point (ICP) algorithm is one of the most commonly used point cloud registration methods for object pose estimation. It is typically employed to refine the pose after an initial rough estimate, as ICP is prone to get trapped in local optimal without a good starting pose [7]. Coherent Point Drift (CPD) is another widely used registration method capable of handling rigid, non-rigid, and affine transformations, making it suitable for pose estimation of deformable objects like cables and ropes [8,9]. Additionally, feature-based registration techniques, including point feature histograms (PFH) and fast point feature histograms (FPFH) have also been adapted for object pose estimation [10]. However, despite their widespread use, traditional methods exhibit limitations when applied to texture-less objects, as similar surface features can lead to incorrect matches between corresponding and non-corresponding points during the registration process [11]. Rebar point cloud models often exhibit texture-less surfaces, and occlusions caused by outer rebar obscuring inner rebar further complicate point cloud registration. These challenges increase the risk of inaccurate rebar pose estimation, potentially leading to collisions between the robot and the rebar. Therefore, there is a pressing need for a novel rebar pose estimation method capable of addressing the challenges posed by texture-less and self-occluding features to enhance the accuracy of robotic tying operations.

To enhance the robot's capability to perform tying operations on complex reinforcement skeletons, where rebar may be arranged in various spatial orientations, this paper proposes a rebar pose estimation method that using deep learning-based keypoint detection and point cloud registration. The proposed method includes several key innovations: (1) Introducing a tying point localization method based on keypoint detection, generating input data with prior displacement information. (2) Proposing a coarse-to-fine registration approach to handle texture-less and self-occluding rebar point clouds for rebar pose estimation. This innovative method enables the robot to adapt to diverse spatial configurations of the reinforcement skeleton and perform tying operations from various orientations and positions.

2 Method

To apply robotic rebar tying technology to diverse reinforcement skeletons in three-dimensional (3D) space, this paper develops a rebar pose estimation method to guide a 6-DoF robot in performing flexible tying operations from various orientations and positions. As shown in Figure 2, the proposed method comprises three modules: data generation, coarse registration, and fine registration. The data generation module functions as a preprocessing component, generating input data and prior displacement information for rebar pose estimation. The coarse registration and fine registration modules implement a coarse-to-fine point cloud registration method to address the challenges of texture-less and self-occluding, enabling accurate alignment of the template point cloud from the template frame to the camera frame for precise rebar position estimation. A detailed explanation is provided below.

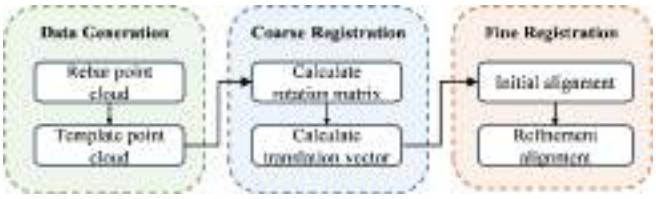


Figure 2. Flowchart of proposed formwork

2.1 Data Generation

2.1.1 Rebar Point Cloud

Keypoint detection is a critical task in computer vision, aimed at identifying essential features such as edges, corners, and textured areas within images or videos. In this study, the recently proposed deep learning-based keypoint detection model YOLOv8 [12] is employed to generate rebar point cloud data and extract the coordinate information of tying points for rebar pose estimation. This is achieved by detecting the bounding boxes and rebar crosspoints, followed by extracting the point cloud from the depth map of the tying area.

Reinforcement skeletons often contain multiple rebar layers at varying depths, presenting a visual challenge for keypoint detection algorithms in accurately identifying the top-layer crosspoints. To mitigate visual interference from rebars at different depths, this study uses a preprocessing technique [13] that applies a point cloud plane fitting algorithm to segment the point cloud at various depths and extract data corresponding to the nearest plane. The resulting point cloud data is transformed into a 2D color image containing only the pixels from the top-layer rebar. By focusing exclusively on the relevant layer, the preprocessing step facilitates accurate localization of tying points on the top layer, ensuring precise coordinate and point cloud data in the crosspoint area. Throughout this paper, the rebar point cloud in the area of the rebar crosspoint is referred to as the "rebar point cloud".

2.1.2 Template Point Cloud

Traditional object pose estimation method determines the pose of an object by aligning a template point cloud, which represents the object's shape and geometric features, with the object point cloud using point cloud techniques. In this study, the template point cloud used for rebar pose estimation is designed based on the typical layout of two intersecting cylindrical rebars in the tying area. The rebar length in the template point cloud is set to *3d*, where *d* is the diameter of the largest rebar in the intersection. Furthermore, to ensure a clear interpretation of registration results, the template's coordinate frame is aligned with the camera's coordinate frame with the Z-axis normal to the rebar plane and the X- and Y-axes aligned with the horizontal and vertical rebars, respectively.

2.2 Coarse Registration

The goal of the coarse registration module is to achieve a rough alignment between the template point cloud and the rebar point cloud. As shown in Figure 3, this is accomplished by determining the appropriate rotation matrix R and translation vector t based on the geometric feature of the rebar point cloud.

The rotation matrix R is computed by aligning the template point cloud parallel to the rebar point cloud using the normal vectors of the planes associated with both point clouds. Let n_t and n_r represent the normal vectors of the planes corresponding to the template point cloud and the rebar point cloud, respectively. In this study, we assume that n_t=(0,0,1). The normal vector n_r can have two possible orientations, which can result in different rotation axes and angles. To ensure consistent alignment, we control the orientation of n_r by ensuring its Z-component is positive, thereby aligning it with n_t. The rotation angle and axis can then be uniformly determined using the following equations:

$$n_r^{'} = \begin{cases} n_r, if\ (n_r)_z > 0 \\ -n_r, if\ (n_r)_z < 0 \end{cases} \tag{1}$$

$$\theta = \arccos\left(\frac{|\, n_r^{'} n_t \,|}{\| n_r^{'} \| \| n_t \|}\right) \tag{2}$$

$$k = \frac{n_t \times n_r^{'}}{\| n_t \times n_r^{'} \|} \tag{3}$$

where $(n_r)_z$ represents the Z-component of the vector n_r, θ refers to the rotation angle from n_t to $n_r^{'}$, and k represents the rotation axis vector. To determine the desired rotation matrix, an extended Rodrigues vector rotation formula is employed, which is widely utilized to describe rotations involving known rotation axes and specified angles [14].

$$R = E\cos\theta + (1-\cos\theta)\begin{pmatrix} k_x \\ k_y \\ k_z \end{pmatrix}(k_x, k_y, k_z) + \sin\theta \begin{pmatrix} 0 & -k_z & k_y \\ k_z & 0 & -k_x \\ -k_y & k_x & 0 \end{pmatrix} \tag{4}$$

where k_x, k_y and k_z denote the three components of the rotation axis k.

Furthermore, the translation vector t is determined by the difference in spatial coordinates between the tying point in the rebar point cloud, calculated from the pixel coordinates $p_r(u,v)$ predicted by YOLOv8, and the corresponding predefined point P_t.in the template point cloud. To obtain the spatial coordinate of the tying point in the rebar point cloud, pixel coordinates $p_r(u,v)$ are converted into spatial coordinates $P_r(X,Y,Z)$ using the following formula:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = d \cdot K^{-1} \cdot \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \quad (5)$$

where d is the depth value of tying point, and K represents the camera intrinsic matrix. The translation vector t can be calculated as:

$$t = P_r - R \cdot P_t \quad (6)$$

2.3 Fine Registration

To further enhance the registration accuracy, the point-to-plane ICP method [15] is employed to minimize the alignment errors between the template point cloud and the rebar point cloud during fine registration. Upon completion of the registration process, the ICP algorithm yields the transformation matrix that aligns the roughly registered template point cloud with the rebar point cloud, ensuring precise alignment for subsequent processing tasks. However, the rebar point cloud often exhibits uneven density between the inner and outer rebar due to self-occlusion, which can cause the registration algorithm to prematurely converge to a local optimum. This typically occurs when the outer rebar is aligned, leaving the sparser inner rebar unaccounted for. To address this issue, an adaptive displacement correction strategy is designed to help the algorithm escape from local optima, as shown in Figure 3. Specifically, after the initial fine registration using point-to-plane ICP, the registration quality is evaluated based on the root mean square error (RMSE) threshold ε. If the RMSE between the aligned template and rebar point clouds exceeds ε, it indicates insufficient alignment and necessitates corrective action. The correction involves introducing displacement value s along both the positive and negative directions of the outer rebar axis in the template frame, followed by recalculating the RMSE for both cases. The displacement correction is then applied in the direction that minimizes the error. Point-to-plane ICP is subsequently reapplied until either the accuracy requirement is satisfied or the maximum number of iterations is reached. In this paper, the RMSE threshold ε was set to 0.9, and the displacement value s was set to $2d$, where d represents the diameter of the inner rebar. This displacement correction strategy enhances alignment robustness and improves overall registration accuracy, leading to more accurate pose estimations for the rebar tying task.

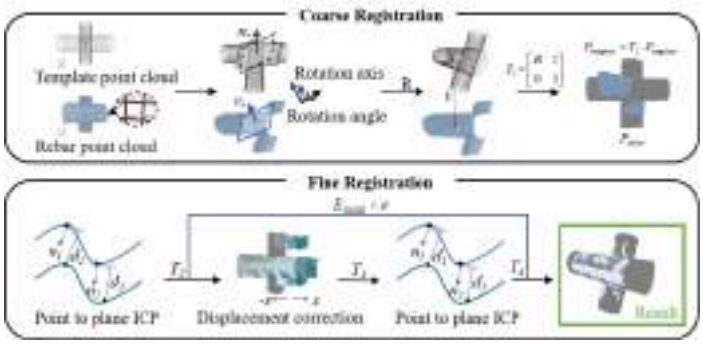


Figure 3. Coarse registration module and fine registration module.

3 Experiment Design

3.1 Experiment Device

The proposed rebar pose estimation method has been integrated into a rebar tying robot developed by our research team, as shown in Figure 4. This system comprises several key components: a 4WD mobile base with a load capacity of 150 kg for enhanced mobility; a 6-DoF robotic arm with a working radius of 1,300 mm for versatile manipulation; and a rebar tying gun as the end effector. Additionally, a 3D camera is configured in an "eye-in-hand" setup, providing a resolution of 1280×1024 and achieving a repeatability of 0.1 mm at a distance of 0.5 m. The industrial PC, operating on Ubuntu 20.04, is equipped with an NVIDIA Ampere GPU and an 8-core ARM Cortex-A78AE CPU, ensuring robust computing capabilities for efficient processing.



Figure 4. Rebar tying robot developed by our research team.

3.2 Experiment Settings

Two experiments were conducted to validate the proposed method. Experiment 1 aimed to test the effectiveness of the proposed coarse-to-fine point cloud registration algorithms for rebar pose estimation, comparing it with other mainstream registration algorithms, such as ICP [15], RANSAC [16], and FGR [17]. All algorithms were implemented using Open3D. To assess the performance of each algorithm in complex scenes, a dataset comprising 3,000 point clouds was created, with 600 point clouds captured at each of the following angles: 0°, 10°, 20°, 30°, and 40°. As shown in Figure 5, the production process involved the following steps: (1) Color maps and depth maps were captured at various angles, with 0° being the ideal camera orientation (Z-axis perpendicular to the rebar plane, X- and Y-axes are aligned parallel to the rebar). Other angles were achieved by rotating around the X-, Y-, and Z-axes. (2) The collected color images underwent preprocessing before being input into the YOLOv8 model for tying point localization. (3) Based on the predicted bounding boxes from the YOLOv8 model and the corresponding depth maps, point cloud models of crosspoint regions were extracted.

Experiment 2 aimed to explore the significance of the proposed visual perception technique by deploying it on a rebar tying robot developed by our team (as shown in Figure 4) and comparing it with other rebar tying methods, including manual tying, handheld tying machine, and other rebar tying robots. To verify the robustness of the proposed method across different scenarios, tying accuracy and efficiency are analyzed through tests involving rebars positioned horizontally, vertically, and inclined, as illustrated in Figure 6.

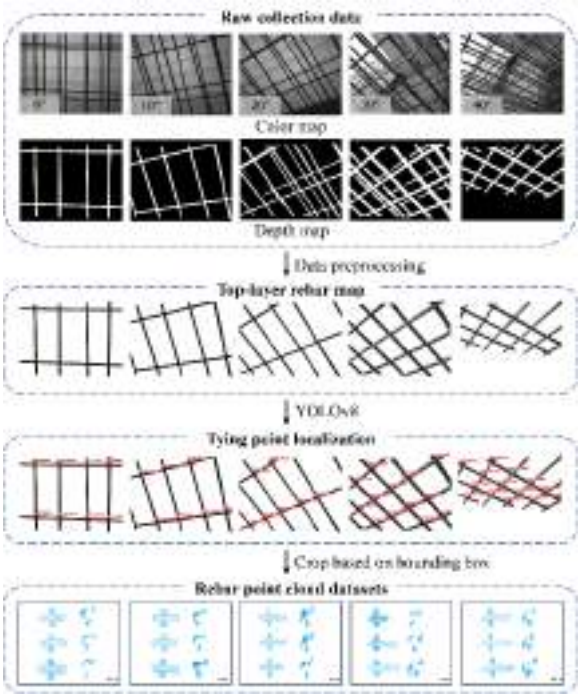


Figure 5. Point cloud model dataset under different shooting angles.



(a) (b) (c)

Figure 6. Different test scenarios: (a) horizontal tying work surface, (b) vertical tying work surface, and (c) inclined tying work surface.

4 Results

4.1 Result of Experiment 1

In Experiment 1, the average RMSE and processing time were utilized to evaluate the performance of four point cloud registration methods. To ensure a fair and reliable comparison of different registration algorithms, an initial translation vector based on the spatial coordinate difference between the tying points in the rebar and template point clouds was provided to each algorithm. This improved both the convergence speed and stability of the registration methods. The experimental results are presented in Table 1. The proposed point cloud registration method significantly reduced the average RMSE error compared to unaligned raw data, effectively addressing texture-less and self-occluding issues during template alignment for rebar pose estimation. Furthermore, the proposed method outperformed other algorithms in registration accuracy across all camera angles and is less affected by camera

angle variations. The superior performance of this method suggests its suitability for the rebar pose estimation task. According to Table 2, the proposed method was slightly slower than ICP, but higher than the other point cloud registration algorithms and still maintained high real-time performance.

Table 1. Average RMSE of different point cloud registration methods (unit: millimeter)

Method	0°	10°	20°	30°	40°
"Raw"	1.039	1.817	3.340	4.913	5.354
ICP	0.522	0.597	0.613	0.812	1.287
RANSAC	1.527	1.751	1.965	2.221	2.532
FGR	2.860	2.893	3.495	4.394	4.637
AnyDirectTying	0.480	0.551	0.526	0.602	0.650

Table 2. Average time consumption of different point cloud registration methods (unit: seconds)

Method	0°	10°	20°	30°	40°
ICP	0.009	0.009	0.010	0.013	0.014
RANSAC	0.098	0.103	0.112	0.119	0.112
FGR	0.032	0.033	0.037	0.039	0.038
AnyDirectTying	0.019	0.024	0.027	0.033	0.034

Given that the ICP algorithm generally outperforms other mainstream methods in terms of accuracy, Figure 7 presents a comparison of the point cloud registration results from the proposed method and the ICP algorithm at various shooting angles. As shown in Figure 7, when the camera angle was at 0°, the performance of ICP algorithm was excellent due to the unaligned raw data being closely aligned, providing an ideal initial state that allowed the ICP algorithm to converge quickly to the correct registration result. However, as the camera angle increased, the accuracy of the algorithm declined. At angle of 10° and 20°, the ICP algorithm tended to fall into local optima rather than achieving the global optimum. This issue was primarily due to self-occlusion, where fewer points were captured from the inner rebar, causing the algorithm to align more readily with the outer rebar while overlooking the misalignment of the inner rebar. This problem became more pronounced when the camera angle reached 30° and 40°, resulting in a rapid decrease in ICP registration accuracy. The proposed method effectively reduces interference by applying an adaptive displacement correction strategy.

In summary, current mainstream point cloud registration methods generally exhibited higher RMSE errors, with some even exceeding those of the unaligned raw data. This discrepancy suggested that existing methods were unsuitable for the task of rebar pose estimation. In contrast, the proposed coarse-to-fine point cloud registration method significantly reduced the RMSE error across different scenarios. This indicated that the proposed method effectively addresses issues related to texture-less surfaces and self-occlusion during template alignment for rebar pose estimation, enabling the robot to accurately perceive the spatial relationship between itself and the tying object, thereby facilitating tying tasks across a broader range of scenarios.

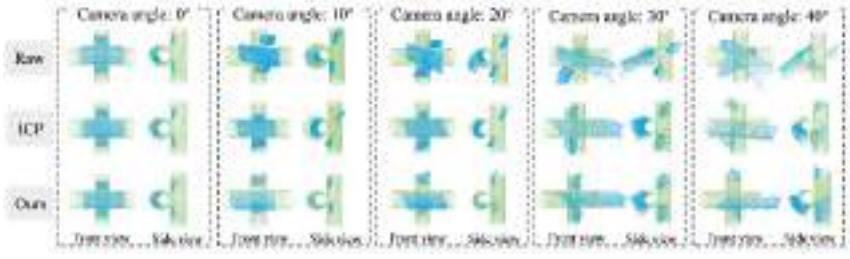


Figure 7. The registration effects of the ICP and the proposed method at different photo angles.

4.2 Result of Experiment 2

To evaluate tying accuracy, 100 rebar crosspoints were tied point by point using the rebar tying robot developed by our research team under different working conditions. As shown in Table 3, the robot achieved a high tying success rate across various working conditions.

In contrast to the visual perception methods of existing robots, which are mainly designed for horizontal rebar mesh and predict only the spatial position of tying points, the proposed method enabled robot to determine the 6-DoF rebar pose in crosspoint region. This capability allows the robot to autonomously perceive the spatial relationship between itself and the tying object, making it capable of performing tying tasks across a broader range of scenarios and demonstrating greater adaptability than existing robotic systems.

Moreover, tying efficiency was analyzed by comparing the average tying time of various tying methods. As shown in Table 4, the average tying time of the rebar tying robot developed in this study was reduced by approximately 39.71% in safe mode and 57.84% in extreme mode compared to manual tying. When compared to handheld tying machine, the robot shortened operation time by approximately 14.58% in safe mode and 40.28% in extreme mode. These results demonstrate that the robot significantly outperforms manual labor by completing more tasks within the same period, thereby boosting construction efficiency. Additionally, the average tying time of our robot was faster than that of other tying robots in extreme mode, although it was slightly slower in safe mode. Nonetheless, the robot maintained a notable overall efficiency advantage.

A further analysis of the tying time components in safe mode revealed that visual perception method accounted for only 0.05 seconds per tying point, just 2.03% of the total tying time. Most of the time was consumed by the execution of the tying action and the movement of the robot arm. This indicates that the slightly lower efficiency in safe mode is primarily due to the mechanical speed of the equipment rather than the performance of the algorithm. Future improvements could focus on deploying the proposed algorithms onto faster, more robust tying robots to enhance their adaptability to multi-directional tying and further improve operational efficiency.

Through the above experiments, it was found that the proposed 6-DoF rebar pose estimation method enabled the robot to perform tying operating across a broader range of scenarios accurately. Additionally, the rebar tying robot deployed with the proposed method demonstrated significant efficiency advantages compared to manual tying and other automatic tying techniques, highlighting the potential of the proposed method to improve construction productivity by guiding the robot towards more efficient and accurate robotic operations.

Table 3. Success rates of our rebar tying robot under different working conditions

Scenario	horizontal	vertical	inclined
Ours	99%	98%	98%

Table 4. Comparison of efficiency of different rebar tying techniques [3,5]

Rebar tying techniques	Average tying time
Manual tying	4.08
Handheld tying machine	2.88
Beijing Institute of Technology	2.34
T-iROBO	2.1
TyBot	2.01
Ours (extreme mode)	1.72
Ours (safe mode)	2.46

5 CONCLUSIONS

Traditional visual perception algorithms for rebar tying robots are often limited to tying objects on horizontal rebar meshes. This study proposes a 6-DoF rebar pose estimation method to guide robotic arms in executing tying operations at arbitrary positions and orientations in 3D space, thereby broadening the application of automated tying technology to various rebar configurations. The key innovations include using a deep learning-based keypoint detection algorithm to extract rebar point data, along with a coarse registration method based on geometric features and a fine registration method utilizing adaptive displacement correction strategy. These innovations effectively address the challenges of rebar pose estimation caused by weak texture and self-occlusion.

Two experiments validated the effectiveness of the proposed algorithm. Experiment 1 demonstrated that the proposed coarse-to-fine point cloud registration method outperformed other mainstream algorithms (ICP, RANSAC, and FGR) in accuracy across different camera angles, achieving an average RMSE of 0.650 mm in complex environments and 0.480 mm in ideal conditions. The algorithm maintains real-time performance, with processing time ranging from 0.019 to 0.034 seconds, demonstrating its capability to handle the challenges posed by the weak texture and self-occlusion in rebar point clouds.

Experiment 2 revealed that the tying robot equipped with the proposed algorithm achieved success rates of 99%, 98%, and 98% for horizontal, vertical, and inclined rebar, respectively, demonstrating improved flexibility and adaptability. The robot also exhibited notable efficiency advantages, completing ties in 1.72 seconds (extreme mode) to 2.46 seconds (safe mode), compared to manual methods and other robots.

In summary, this study enhances the tying capabilities of vision-guided rebar tying robots by developing a 6-DoF rebar pose estimation method, thereby extending their applicability to diverse rebar configurations. This research fills a research gap in 6-DoF rebar pose estimation for robotic rebar tying,

offering an advanced visual perception method that reduces reliance on manual labor. Future work will focus on improving the accuracy and stability of the algorithm, particularly regarding variations in rebar diameter or spacing.

ACKNOWLEDGMENTS

The authors would like to acknowledge the financial support provided by the National Key Research and Development Program of China (Grant No. 2023YFC3806800), the National Natural Science Foundation of China (Grant No. 52308312), the Hunan Province funding for leading scientific and technological innovation talents (Grant No. 2021RC4025), and the Postgraduate Scientific Research Innovation Project of Hunan Province (Grant No. CX20240462).

REFERENCE

[1] C.A. Issa and A. Nasr. An experimental study of welded splices of reinforcing bars. *Building and Environment*, 41(10):1394–1405, 2006.

[2] N. Melenbrink, J. Werfel, and A. Menges. On-site autonomous construction robots: Towards unsupervised building. *Automation in Construction*, 119:103312, 2020.

[3] J. Jin, W. Zhang, F. Li, M. Li, Y. Shi, Z. Guo, and Q. Huang. Robotic binding of rebar based on active perception and planning. *Automation in Construction*, 132:103939, 2021.

[4] Future Robotics Technology Center. Chiba Institute of Technology. On-line: https://furo.org/en/works/t_irobo_rebar/t_irobo_rebar.html, Accessed: 06/08/2024.

[5] B. Cheng and L. Deng. Vision detection and path planning of mobile robots for rebar binding. *Journal of Field Robotics*, 41(6):1864–1886, 2024.

[6] D.-C. Hoang, A.-N. Nguyen, V.-D. Vu, T.-U. Nguyen, D.-Q. Vu, P.-Q. Ngo, N.-A. Hoang, K.-T. Phan, D.-T. Tran, V.-T. Nguyen, Q.-T. Duong, N.-T. Ho, C.-T. Tran, V.-H. Duong, and A.-T. Mai. Graspability-aware object pose estimation in cluttered scenes. *IEEE Robotics and Automation Letters*, 9(4):3124–3130, 2024.

[7] J. Guo, X. Xing, W. Quan, D.-M. Yan, Q. Gu, Y. Liu, and X. Zhang. Efficient center voting for object detection and 6D pose estimation in 3D Point Cloud. *IEEE Transactions on Image Processing*, 30:5072–5084, 2021.

[8] C. De Farias, B. Tamadazte, R. Stolkin, and N. Marturi. Grasp transfer for deformable objects by functional map correspondence. In: *International Conference on Robotics and Automation (ICRA)*, pages 735–741, Philadelphia, PA, USA, 2022.

[9] T. Tang, C. Wang, and M. Tomizuka. A framework for manipulating deformable linear objects by coherent point drift. *IEEE Robotics and Automation Letters*, 3(4): 3426-3433,2018.

[10] Y. Zhu, M. Li, W. Yao, and C. Chen, A review of 6D object pose estimation. In: *2022 IEEE 10th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*. pages 1647–1655, Chongqing, China, 2022.

[11] C.-M. Lin, C.-Y. Tsai, Y.-C. Lai, S.-A. Li, and C.-C. Wong. Visual object recognition and pose estimation based on a deep semantic segmentation network. *IEEE sensors journal*, 18 (22): 9370–9381, 2018.

[12] A. Aboah, B. Wang, U. Bagci, and Y. Adu-Gyamfi. Real-Time Multi-Class Helmet Violation Detection Using Few-Shot Data Sampling Technique and YOLOv8. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5350–5358, Vancouver, BC, Canada, 2023.

[13] S. Wang, L. Deng, J. Guo, M. Liu, and R. Cao. Automatic Quality Inspection of Rebar Spacing Using Vision-Based Deep Learning with RGBD Camera. In *International Symposium on Automation and Robotics in Construction*, pages 57–64, Lille, France, 2024.

[14] Z. Zou, H. Lang, Y. Lou, and J. Lu. Plane-based global registration for pavement 3D reconstruction using hybrid solid-state LiDAR point cloud. *Automation in Construction*, 152:104907, 2023.

[15] K.-L. Low. Linear least-squares optimization for point-to-plane icp surface registration. Chapel Hill, University of North Carolina, 4(10):1-3, 2004.

[16] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[17] Q.-Y. Zhou, J. Park, and V. Koltun. Fast Global Registration. In *Proceedings of the European Conference on Computer Vision*, pages 766–782, Cham, Switzerland, 2016.

Adaptive Kinematic and Dynamic Control for Hexapod Robots in Complex Construction Environments

Xiayu Zhao[1], Houtan Jebelli[1]

[1]Department of Civil and Environmental Engineering,
University of Illinois Urbana-Campaign, Champaign, IL, USA
xiayuz3@illinois.edu, hjebelli@illinois.edu

Abstract –

Robotic systems for inspection have advanced significantly in construction. Hexapod robots offer unique advantages for navigating complex construction terrains, yet their performance on sloped and irregular surfaces is hindered by limitations in existing control strategies. . In this work, we present an integrated control framework that combines a dynamically tuned fuzzy PD controller with a Bézier curve-based trajectory generator. This unified approach addresses the critical research gap by simultaneously ensuring precise posture regulation and smooth gait execution, thus enhancing hexapod stability and efficiency. Simulations in the Gazebo environment validated the system's performance on various roof pitches (2/12, 4/12, 6/12) and structural transitions (ridges, valleys, flashings). Results showed a 21% improvement in stability margin on 2/12 slopes, precise body orientation control within ±5° roll, ±10° pitch, and ±1.3 m/s² vertical acceleration variations. Also, this system achieved 80-100% success rates in three different types of surface transitions. This research provides a robust framework for deploying hexapod robots in challenging architectural contexts, advancing autonomous inspection capabilities in construction environments.

Keywords –

Robotic system; Hexapod Robots; Roof Inspection; Adaptive Kinematic Control; Complex Construction Environment; Autonomous Construction Inspection

1 Introduction

The construction industry is undergoing a transformative shift, propelled by the integration of robotic systems that streamline workflows and boost operational efficiency [1]. Among these applications, robotic systems developed for inspection tasks have advanced significantly, with researchers focusing on innovative locomotion methods that enable robots to navigate complex architectural structures [2,3]. Robots equipped with AI and computer vision are increasingly outperforming human capabilities in detecting and analysing maintenance needs, providing more accurate and timely assessments [4,5].

Despite these technological advancements, deploying robots for high-precision inspection tasks on construction sites remains challenging, often limiting their practical application [6]. Roof inspections, for instance, are particularly demanding due to diverse types of visible damage—curling, cracking, flaking, granule loss, and missing shingles—as well as less apparent issues, such as water infiltration, loosening, and blistering, which further complicate assessment [7]. To address these specialized inspection challenges, researchers have explored various robotic platforms with different locomotion mechanisms. Aerial robots, including Unmanned Aerial Vehicles (UAVs), offer potential solutions [8] but are constrained by payload capacity, battery life, and regulatory restrictions in urban settings, along with limited close-up inspection capabilities. Unmanned Ground Vehicles (UGVs) offer a viable alternative, as they support heavier payloads and can sustain longer operational times, making them more suitable for close-range inspections. However, standard UGVs often lack the adaptive locomotion needed to navigate the complex, sloped, and discontinuous surfaces commonly found on roofs, emphasizing the need for UGVs specifically designed for safe, slope-adaptive movement [9].

Among the various UGV configurations available for construction inspection tasks, multi-legged designs offer distinct advantages for complex terrain navigation. UGVs encompass various configurations, including wheeled, tracked, multi-legged, and hybrid forms, each with distinct advantages. Multi-legged UGVs excel on rough and uneven terrain [10]. Their ability to step over or around obstacles allows them to maneuver effectively in unstructured environments, making them well-suited for complex terrains. Multi-legged robots also offer greater agility, enabling movement in multiple directions, such as sideways and diagonally, which is particularly beneficial on sloped surfaces. Additionally, these designs

ensure stability on downhill slopes by distributing load more effectively and provide resilience, allowing continued operation even with a compromised leg. Among multi-legged designs, hexapod robots stand out for their optimal balance between stability and complexity [11]. With six legs, they maintain static stability by keeping the Center of Gravity (CoG) within a support triangle formed by three legs, enabling a tripod gait that is both stable and efficient. This configuration provides redundancy, allowing the robot to function even if one or more legs malfunction, while avoiding the added complexity and energy demands of designs with more legs. Compared to quadrupeds, hexapods offer superior stability without requiring sophisticated dynamic balancing, as well as enhanced ground clearance and obstacle traversal. Furthermore, the hexapod design distributes load effectively, enabling the robot to carry heavier payloads while maintaining stability on uneven or inclined surfaces.

Despite these inherent advantages, hexapod robots still face challenges when navigating sloped and irregular terrains, as they often lack specialized kinematic and dynamic control mechanisms needed for such environments [12]. In applications like roof inspection, where safe and adaptive movement is critical, these gaps can lead to risks, including fall hazards [13], increased maintenance costs, and elevated energy consumption. This paper addresses this gap by proposing a novel integrated control framework that enhances both the kinematic and dynamic performance of hexapod robots. Our primary contributions include:

1. Providing an integrated control framework that synergistically combines an adaptive fuzzy PD controller and a Bézier curve-based trajectory generator. This unified approach dynamically adjusts leg motor angles based on center-of-gravity errors while generating smooth, energy-efficient swing-phase trajectories, resulting in superior posture regulation and enhanced stability on variable roof slopes and irregular terrains.
2. Validating the approach through extensive simulations, demonstrating a 21% improvement in stability margin on standard roof slopes and robust performance during complex structural transitions.

2 Methodology

This section presents a comprehensive description of the module-enhanced hexapod robot platform developed for automated roof inspection tasks.. This method adapt the traditional fuzzy PD controller with dynamic gain adjustments tailored for real-time leg motor control. Our study also extends the conventional use of Bézier curves by tuning the parameters for swing-phase trajectory generation. This results in smoother, energy-efficient leg movements that significantly improve gait stability over traditional semicircular trajectories. Furthermore, the novel aspect of our work is the integration of these two modified components into a unified control framework. This synergy not only leverages the strengths of both techniques but also produces superior performance,

2.1 Hexapod Kinematic Model

The hexapod robot platform utilized in this study is based on the model Muto [14], an 18 degrees-of-freedom (DoF) six-legged walking robot. The robot is equipped with a Jetson NANO board [15] for on-board processing. Locomotion is achieved through 18 high-torque servo motors. For precise posture measurement and adjustment, the robot incorporates an inertial measurement unit (IMU), capable of measuring acceleration, gyroscope data, magnetometer readings, and temperature. In default standing position, the hexapod robot measures 465 mm in length (forward direction) and 514 mm in width (lateral direction). Each of the six legs is composed of three segments: coxa (50 mm), femur (75 mm), and tibia (140 mm). The robot's coordinate frame is defined as shown in Fig.1(a): **{c}** is the mass center coordinate frame of hexapod robot determined by right-hand rule, with X_c along the body transverse direction, Y_c along the forward direction, and Z_c perpendicular to the ground. As shown in Fig.2(b), $\mathbf{\{O_{ij}\}}$ is the coordinate frame of the j_{th} joint of the i_{th} leg of the robot, $i \in \{1,2,3,4,5,6\}$, $j \in \{1,2,3,4\}$. Note when j=4, $\mathbf{\{O_{i4}\}}$ presents the coordinate frame of the foot end. Axis $\mathbf{Z_{i,j}}$ is the rotation axis of the joint, the $\mathbf{X_{i,j}}$ is chosen along the common normal from axes $\mathbf{Z_{i,j}}$ and $\mathbf{Z_{i,(j-1)}}$, and the $\mathbf{Y_{i,j}}$ is determined by righthand rule. Respectively, L1, L2, L3 represent the length of coxa (C_i), femur (F_i), and tibia (T_i), $i \in \{1,2,3,4,5,6\}$.

2.1.1 Forward Kinematics

The forward kinematics of the 18-DOF hexapod robot is based on the Denavit–Hartenberg (DH) convention. This approach allows for a systematic representation of the spatial relationships between adjacent link coordinate frames, facilitating the calculation of the end-effector position (in this case, the foot position) given the joint angles. The transformation matrix between two adjacent leg coordinate systems is

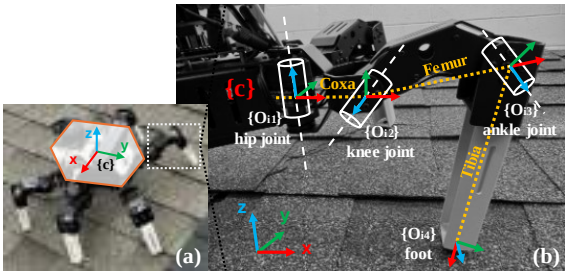


Figure 1. Kinematic Model of a Hexapod Robot Leg: Joints, Links, and Coordinate Frames

expressed in the DH convention:

$$\mathbf{M}_{\mathrm{j}}^{\mathrm{j-1}} = \begin{bmatrix} c_{\theta_j} & -s_{\theta_j}c_{\alpha_j} & s_{\theta_j}c_{\alpha_j} & \alpha_j c_{\theta_j} \\ s_{\theta_j} & c_{\theta_j}s_{\alpha_j} & -c_{\theta_j}s_{\alpha_j} & \alpha_j s_{\theta_j} \\ 0 & s_{\alpha_j} & c_{\alpha_j} & d_i \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (1)$$

where c_{θ_j}, s_{θ_j} denotes cos (θ_j) and sin (θ_j), respectively, and α, a, θ, d is the DH parameters with the values listed in Table 1.

Table 1. Denavit–Hartenberg Parameters Between **{c}** and the Hexapod's foot $\mathbf{p_i}$

Transformation	α_i	a_i	θ_i	d_i
{c}→{O$_{i1}$}	0	l_1	θ_{i1}	0
{O$_{i1}$}→{O$_{i2}$}	$\pi/2$	l_2	θ_{i2}	0
{O$_{i2}$}→{O$_{i3}$}	0	l_3	θ_{i3}	0
{O$_{i3}$}→{O$_{i4}$}	0	l_4	θ_{i4}	0

In Table 1, $\boldsymbol{\alpha}$ is the angle about common normal, from old z axis to new z axis; $\boldsymbol{a}$ is the length of the common normal; $\boldsymbol{\theta}$ is the angle about previous z from old x to new x; $\boldsymbol{d}$ is the offset along previous z to the common normal. Let $\mathbf{p^c}$ be the foot-tip position of the leg in the coordinate system relative to the mass center and $\mathbf{p^4}$ be the foot-tip position in the foot-tip coordinate system. Then the mapping between the body coordinate system and the foot-tip coordinate system is given by the following kinematic chain:

$$\mathbf{p}_{\mathrm{i}}^{\mathrm{c}} = \mathbf{T}_i^c \mathbf{M}_2^1 \mathbf{M}_3^2 \mathbf{M}_4^3 \mathbf{p}_{\mathrm{i}}^4 \quad (2)$$

where $\mathbf{T}^{\mathrm{i}}$ is the transformation matrix between the body center frame **{c}** and the coxa coordinate frame of the $\mathrm{i_{th}}$ leg, given as a rigid body transformation:

$$\mathbf{T}^i = \begin{bmatrix} c_{\beta^i} & -s_{\beta^i} & 0 & p_x^i \\ s_{\beta^i} & -c_{\beta^i} & 0 & p_y^i \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3)$$

where β^i, $\mathrm{p_x^i}$, and $\mathrm{p_y^i}$, are the body parameters of the $\mathrm{i_{th}}$ link which are known by default setting.

2.1.2 Inverse Kinematics

The inverse kinematics for the hexapod robot is simplified by its joint configuration, allowing for an analytical solution. Given the desired foot-tip coordinates in the global frame, the inverse kinematics calculates the required joint angles for each leg. For a given foot-tip position $\mathbf{p^c} = [\mathrm{p_1^c}, \mathrm{p_2^c}, \mathrm{p_3^c}]$ in the global coordinates, the hip joint angle θ_1^i of the $\mathrm{i_{th}}$ is obtained as:

$$\theta_1^i = \begin{cases} 0, i = 1 \\ arctan\left(\frac{l_a}{l_b}\right), i \in \{2,3,5,6\} \\ \pi, i = 4 \end{cases} \quad (4)$$

where $\mathrm{l_a}$ and $\mathrm{l_b}$ are illustrated in Figure 2(a).

The angle θ_2^i is calculated with respect to the fixed reference frame {**h**} located at the **hip** joint (Fig. 1b):

$$\theta_2^i = \tan^{-1}\left(\frac{x_p^h}{y_p^h}\right) \quad (5)$$

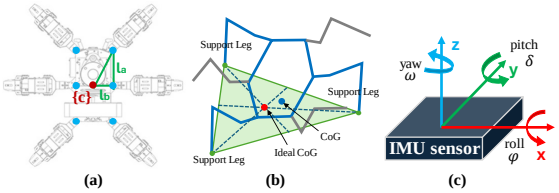


Figure 2. (a) Hip Joint Relative to the Center Frame, (b) Diagram Explaining CoG and CSP, (c) Rotation Angles Measured by the IMU Sensor

where $\mathbf{p}_{\mathrm{i}}^{\mathrm{h}} = \{x_p^h, y_p^h, z_p^h\}^T$ is the position of $\mathbf{p}_{\mathrm{i}}$ relative to {**h**}, obtained by:

$$\mathbf{p}_{\mathrm{i}}^{\mathrm{h}} = (\mathbf{T}_1^0)^{-1}\mathbf{p}_{\mathrm{i}}^{\mathrm{c}} \quad (6)$$

in which $\mathbf{T}_1^0$ is the transformation matrix between the center frame {**c**} and the reference frame {**h**}. To calculate the remaining angles, we need the position of reference frame {**k**} at the knee joint, defined by the Denavit-Hartenberg convention as:

$$\mathbf{k}_{\mathrm{i}}^{\mathrm{h}} = (\mathbf{T}_1^0)^{-1}\mathbf{T}_2^0 \quad (7)$$

Considering the vector $\mathbf{s}_{\mathrm{p}}$ between the knee frame {**k**} and the foot, the angle θ_3^i and the θ_4^i are calculated as:

$$\theta_3^i = \cos^{-1}\left(\frac{{l_3}^2 + \|\boldsymbol{s_p}\|^2 - {l_4}^2}{2l_3\|s_p\|}\right) - \sin^{-1}\left(\frac{z^k - z^p}{\|\boldsymbol{s_p}\|}\right) \quad (8)$$

$$\theta_4^i = \pi + \cos^{-1}\left(\frac{{l_3}^2 + {l_4}^2 - \|\boldsymbol{s_p}\|^2}{2l_3l_4}\right) \quad (9)$$

z^k and z^p represent the vertical positions of the knee frame {k} and the foot, respectively. These equations form the foundation for the robot's motion control and gait design in subsequent sections.

2.2 Integrated Control System for Slope and Terrain Stability

2.2.1 Adaptive Control System for Slope Navigation

The stability of a hexapod robot during locomotion on both inclined and uneven surfaces is crucial for effective roof inspection tasks. This section presents an integrated approach that combines slope-adaptive posture control with enhanced gait stability mechanisms to ensure reliable operation across diverse terrain conditions. The foundation of the system's stability is based on the principle that the hexapod robot maintains static stability when the vertical projection of its Center of Gravity (CoG) falls within the Conservative Support Polygon (CSP) formed by its supporting legs (Fig. 2b). While this principle is straightforward on flat surfaces where the D-H convention and forward kinematics can directly calculate support points and CoG projection, inclined surfaces present additional complexity as the robot's reference frame **{c}** becomes misaligned with the global reference frame **{s}.** To address this misalignment,

a rotation matrix is introduced to transform positions from the robot's reference frame **{c}** to the space reference frame **{s}:**

$$\mathbf{R_{sc}} = \begin{bmatrix} c_\omega c_\delta & c_\omega s_\delta s_\varphi - s_\omega c_\varphi & c_\omega s_\delta c_\varphi + s_\omega s_\varphi & 0 \\ s_\omega c_\delta & s_\omega s_\delta s_\varphi + c_\omega c_\varphi & s_\omega s_\delta c_\varphi - c_\omega s_\varphi & 0 \\ -s_\delta & c_\delta s_\varphi & c_\delta c_\varphi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (10)$$

where the rotation angles φ around X-axis, δ around Y-axis, and ω around Z-axis are measured by the IMU sensor (Fig. 2c) positioned at the robot's center of mass.

Stability on inclined surfaces is further improved using a fuzzy Proportional-Derivative (PD) controller, which dynamically adjusts the supporting leg motors' angles. This study chooses the fuzzy PD controller for three main reasons: (1) it handles the nonlinearities and uncertainties of slope navigation without requiring a precise dynamic model, (2) it dynamically adjusts control parameters based on error magnitude and rate of change, ensuring better adaptation to varying slopes, and (3) it strikes an optimal balance between performance and computational efficiency, essential for real-time embedded systems in field robotics. The controller maintains the real CoG's projection close to the ideal CoG within the CSP by calculating errors e_x and e_y, representing differences between the actual and ideal CoG positions. The compensation angles $\Delta\mu_x$ and $\Delta\mu_y$ are calculated as:

$$\Delta\mu_x = K_{p,x}e_x + K_{d,x}\dot{e}_x \quad (11)$$

$$\Delta\mu_y = K_{p,y}e_y + K_{d,y}\dot{e}_y \quad (12)$$

where $K_{p,x}$ and $K_{p,y}$ are the proportional gains, $K_{d,x}$ and $K_{d,y}$ are the derivative gains, all of which are dynamically adjusted by the fuzzy PD controller based on the magnitude of errors and their derivatives [16]. This approach allows the system to respond appropriately to varying conditions. For instance, when errors are large, smaller gains are applied to prevent the robot from overturning. Conversely, when errors are small, larger gains are used for fine-tuning motor adjustments, enhancing stability (Table 2,3., Figure 3).

Table 2. Consequent membership Functions

Parameter \ Value	↑	↑↑	↑↑↑	↑↑↑↑
$K_{p,x}$	0.05	0.1	0.15	0.2
$K_{p,y}$	0.2	0.3	0.4	0.5
$K_{d,x}$	0.05	0.07	0.09	0.15
$K_{d,y}$	0.05	0.07	0.09	0.15

Table 3. Fuzzy rule table for the proportional and derivative gains

$K_{p,x}$ $K_{d,y}$	$K_{p,y}$ $K_{d,x}$	e_y				
		− −	−	0	+	+ +
$\dot{e}_y$	− −	↑	↑↑	↑↑↑	↑↑	↑
	−	↑↑	↑↑↑	↑↑↑↑	↑↑↑	↑↑
	0	↑↑↑	↑↑↑↑	↑↑↑↑	↑↑↑↑	↑↑↑
	+	↑↑	↑↑↑	↑↑↑↑	↑↑↑	↑↑
	+ +	↑	↑↑	↑↑↑	↑↑	↑

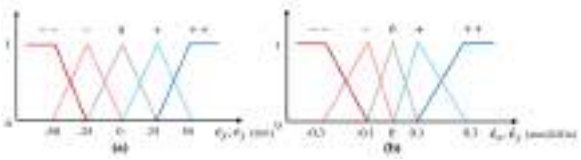


Figure 3. Fuzzy rule plots for e_x, e_y and $\dot{e}_x, \dot{e}_y$

In practice, the global reference frame rotates with the robot's reference frame around the Z-axis to ensure that compensation angles remain effective regardless of the robot's orientation. The fuzzy PD controller's ability to adapt dynamically outperforms traditional linear PD controllers, particularly on inclined surfaces, by maintaining stability without detailed dynamic modelling. This adaptive control strategy was validated through experiments on roof pitches of varying slopes, detailed in the *Experimental Validation* section.

2.2.2 Stability Enhancement for Uneven Terrain

The robot employs a tripod gait as its primary locomotion strategy, where three legs remain in the stance phase while the other three are in the swing phase. The legs are divided into two sets: Set A (legs R1, L2, and R3) and Set B (remaining legs). This configuration enables a four-step sequence: initial six-foot contact (Fig. 4a), Set A forward movement (Fig.4b), Set B forward movement with Set A thrust (Fig.4c), and Set A lift returning to initial position (Fig.4d). The swing leg trajectory is the curve connecting the lifting point and the landing point of the swing leg's tip.

Traditionally, a semicircular trajectory algorithm is used for the swing phase, generating uniform foot arcs. While computationally efficient, this method produces abrupt transitions between the stance and swing phases, increasing energy consumption and destabilizing the robot's posture. To mitigate these issues, this study introduces a Bézier curve-based trajectory for the swing phase. This method, modelled as a third-degree curve, ensures more consistent height control during the swing phase. This study chose Bézier curves over polynomial, spline, and semicircular trajectories for three main reasons: (1) control points enable smoother stance-to-swing transitions, (2) Bézier curves avoid the oscillations typical of high-order polynomials and maintain stability under varying terrain conditions, and (3) they strike an optimal balance between trajectory quality and computational efficiency.

During the swing phase, the feet follow a 3rd-degree

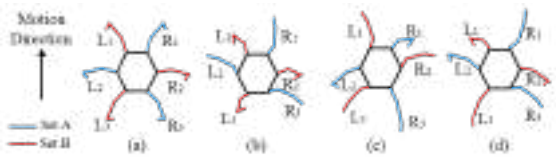


Figure 4. Diagram of four-step tripod gait

Bézier trajectory described by:

$$\mathbf{p}_{i,j}^{c}(t) = \sum_{j=0}^{3} \binom{3}{j} t^{j} (1-\mathrm{t})^{3-j} \mathbf{p}_{i,j}^{c},$$

$$t \in [0, t_{int}], i \in [1,6], j = 0,1,2,3 \quad (13)$$

$$\binom{3}{j} = \frac{3!}{j!\,(3-j)!} \quad (14)$$

where 3 denotes the curve degree, j denotes the number of joint, t_{int} is the trajectory's time interval, and $\mathbf{p}_{i,j}^{c} = \{x_p^c, y_p^c, z_p^c\}^T$ denote the control points of the trajectory.

This approach provides superior stability compared to traditional methods, particularly in terms of roll and pitch velocity variation and height stability. The control points for the Bézier curve are strategically positioned to ensure smooth transitions between phases while maintaining stability. The first control point corresponds to the foot's initial position $\mathbf{p}_{i,0}^{c}$ is defined by Eq. (2), which can also be derived as:

$$\mathbf{p}_{\mathrm{i,0}}^{\mathrm{c}} = \begin{Bmatrix} c_{(\theta_0+\theta_1)}(L_3 c_{(\theta_2+\theta_3)} + L_2 c_{\theta_2} + L_1) + L_0 c_{\theta_0} \\ s_{(\theta_0+\theta_1)}(L_3 c_{(\theta_2+\theta_3)} + L_2 c_{\theta_2} + L_1) + L_0 s_{\theta_0} \\ L_3 s_{(\theta_2+\theta_3)} + L_2 s_{\theta_2} \end{Bmatrix} \quad (15)$$

The remaining control points are determined as:

$$\mathbf{p}_{\mathrm{i,1}}^{\mathrm{c}} = \left\{0, \frac{s}{4}, \frac{h}{3}\right\}^{\mathrm{T}} + \mathbf{p}_{\mathrm{i,0}}^{\mathrm{c}} \quad (16)$$

$$\mathbf{p}_{\mathrm{i,2}}^{\mathrm{c}} = \left\{0, \frac{3s}{4}, \frac{h}{3}\right\}^{\mathrm{T}} + \mathbf{p}_{\mathrm{i,0}}^{\mathrm{c}} \quad (17)$$

$$\mathbf{p}_{\mathrm{i,3}}^{\mathrm{c}} = \{0, s, 0\}^{\mathrm{T}} + \mathbf{p}_{\mathrm{i,0}}^{\mathrm{c}} \quad (18)$$

where h is the maximum height, and s denotes the stride length. During the stance phase, where the feet maintain ground contact while propelling the torso forward, the foot trajectory is defined as:

$$\mathbf{p}_{\mathrm{i}}^{\mathrm{c}}(\mathrm{t}) = \left\{0, -\frac{s}{t_{int}} t, 0\right\}^{\mathrm{T}} + \mathbf{p}_{\mathrm{i,0}}^{\mathrm{c}}, t \in [0, t_{int}] \quad (19)$$

By integrating slope-adaptive posture control with a stability-enhanced gait system, this approach addresses the dual challenges of inclined and uneven terrain navigation. The adaptive slope system ensures that the CoG remains properly aligned within the CSP, while the Bézier curve trajectory smooths leg movements, enhancing stability and energy efficiency.

3 Experimental Validation

The experimental validation of the proposed control system was conducted through two comprehensive simulation tests implemented in the Gazebo environment, each designed to evaluate different aspects of the hexapod robot's performance. The simulation environment was configured using Robot Operating System (ROS) and Gazebo, providing a physics-accurate testing platform. The same hexapod robot URDF (Unified Robot Description Format) model was employed across both test setups, ensuring consistency in the robot's physical parameters and joint configurations throughout the experimentation process. Gazebo and ROS together provide a comprehensive and flexible setup for simulating and measuring with the appropriate sensors, plugins, and ROS nodes.

3.1 Test 1: Diverse Slopes

Test 1 evaluated the control system's stability across various roof inclines using the tripod gait, which balances stability and speed at 0.1 m/s. Three standard residential roof pitches were tested: 2/12 (9.5°), 4/12 (18.4°), and 6/12 (26.6°). The simulation used a 3.0m × 2.0m platform with high-friction coefficient (μ = 0.8) to simulate roofing material. For each slope, the hexapod performed five ascending trials covering 3.0m distance.

The robot model incorporated an IMU sensor at its center of mass, providing orientation data through ROS "sensor_msgs" for accurate CoG projection calculations. Performance metrics included stability margin, body orientation angles, foot placement accuracy, and power consumption. The simulation featured realistic physics parameters to ensure accurate representation of the robot's behavior on inclined surfaces. For all experiments, we implemented a comparative analysis between our integrated control system and a traditional control approach. The baseline system utilized a standard PD controller with fixed gains (Kp = 0.5, Kd = 0.15) for posture regulation and conventional semicircular trajectory generation for leg movements.

3.2 Test 2: Surface Transitions

Test 2 evaluated the control system's ability to navigate common roof structural transitions using the same hexapod model from Test 1. The simulation focused on three critical elements: ridge lines (30° angular transition), valleys (135° angles), and flashings (10mm elevation change).

For each transition type, ten trials were conducted at 0.1 m/s from identical start positions to ensure statistical reliability. Data was collected using the robot's integrated sensors, including IMU readings for stability assessment and foot pressure sensors for contact force analysis.

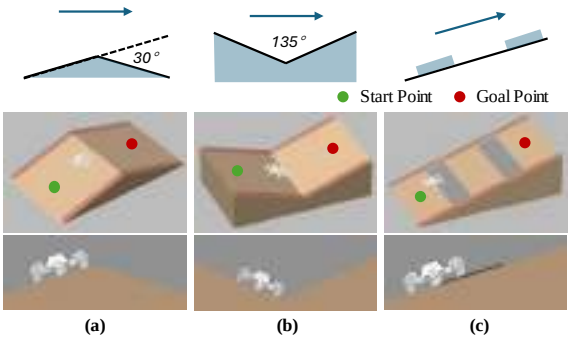


Figure 6. Three different surface transition trials in Test 2: (a) Ridge (30 degree), (b) Valley (135 degree), (c) Flashing (10mm elevation).

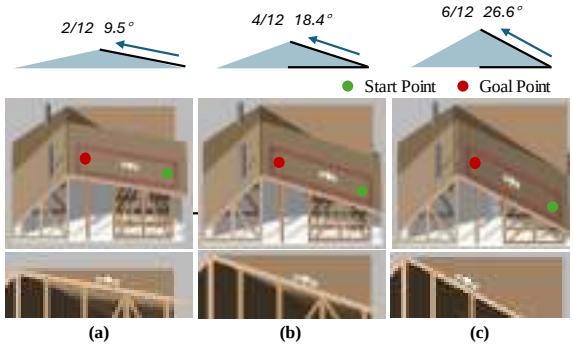


Figure 5. Three different slope angle trials in Test 1: (a) Pitch 2/12 (~9.5 degree), (b) Pitch 4/12 (~18.4 degree), (c) Pitch 6/12 (~26.6 degree).

Performance evaluation concentrated on two key metrics: transition success rate and crossing time.

3.3 Performance Metrics

To comprehensively evaluate the hexapod robot's performance across both slope navigation and structural transition tests, a multi-metric evaluation framework was implemented. The primary stability indicator was the stability margin (SM), defined as the shortest distance from the horizontally projected Center of Gravity (CoG) to the edge of the support polygon (McGhee and Frank 1968):

$$S_m \equiv \min(D_1, D_2, D_3)$$

where D_i represents the perpendicular distance from the projected CoG to each edge of the support triangle formed by the supporting legs. This metric has been extensively validated in previous studies for analysing hexapod stability [17] and designing fault-tolerant gaits [16]. Body orientation control was monitored through roll (φ) and pitch (θ) angles, measured in degrees from the horizontal plane. The maximum allowable deviations were established at 15° for roll angle, 20° for pitch angle on slopes, and 12° for pitch angle during structural transitions. These measurements were crucial for assessing the robot's ability to maintain proper posture throughout locomotion.

For Test 1 (Diverse Slopes), specific performance metrics included: (1) Average stability margin maintained during continuous slope walking; (2) Maximum and minimum body orientation angles during slope traversal; (3) Vertical Acceleration variations. For Test 2 (Surface Transitions), additional metrics were implemented: (1) Transition success rate, defined as the percentage of crossings completed without stability failures such as turnover or keep stepping in place; (2) Crossing time, measured from the initiation of transition approach to completion of crossing from fixed start and end positions. All metrics were recorded through the ROS framework using custom message types, with data synchronized through the ROS time stamp system. The results are analysed in the following section.

4 Results and Discussion

4.1 Results Analysis

4.1.1 Test 1: Diverse Slopes Performance

The experimental results revealed significant stability gains with our control system for all tested slope angles (Figure 7), namely 2/12 (9.5°), 4/12 (18.4°), and 6/12 (26.6°). On the gentlest slope, the system maintained an average SM of 115 mm, 21% higher than the default. Although this margin dropped to 84 mm at 4/12 pitch, it still surpassed the default by 15%. On the steep 6/12 pitch, our system recorded 44 mm—over double the 20 mm minimum for stable locomotion—while the default system repeatedly overturned or stepped in place. These results highlight the inherent challenges of steep inclines, which our adaptive control strategy successfully mitigated. Overall, we achieved a 100% success rate, with only brief interruptions during initial slope engagement.

Body orientation control also showed strong performance (Figure 8). Our system maintained body height between 255 and 272 mm, while the uncontrolled approach varied from 228 to 289 mm, especially on steeper slopes. Roll angles stayed within ±2°, ±3°, and ±5° at 9.5°, 18.4°, and 26.6°, respectively—well below the ±15° limit. By contrast, the default approach reached ±15° on the 6/12 pitch. Pitch angles similarly stayed within ±10° under our system, far beneath the 20° cutoff, while the default reached ±25°. The fuzzy controller's adaptive nature minimized sudden orientation changes, containing vertical acceleration within ±1.3 m/s², whereas the uncontrolled approach hit ±2.5 m/s². These findings confirm the robust stabilization capabilities of our system on inclined surfaces.

Such improvements in stability and orientation are critical for preventing slip or rollover events. By minimizing abrupt shifts, our system reduces mechanical stress and power consumption, enabling safer, more efficient navigation and real-world roof inspections. Hence, it effectively supports steep-roof applications requiring reliable locomotion.

4.1.2 Test 2: Surface Transition Performance

As shown in Table 4, the system demonstrated high success rates across all structural surface transitions, with 80% successful completions for ridge line crossings (30° angle) among 10 trials, compared to 50% success rate without our control system. Under another 10 trails on valley transitions (135° angle), trails under our control system gained 100% success, while only 6 of another 10 trails succeeded without our control system. For flashing interface navigation (10mm elevation change), both trails

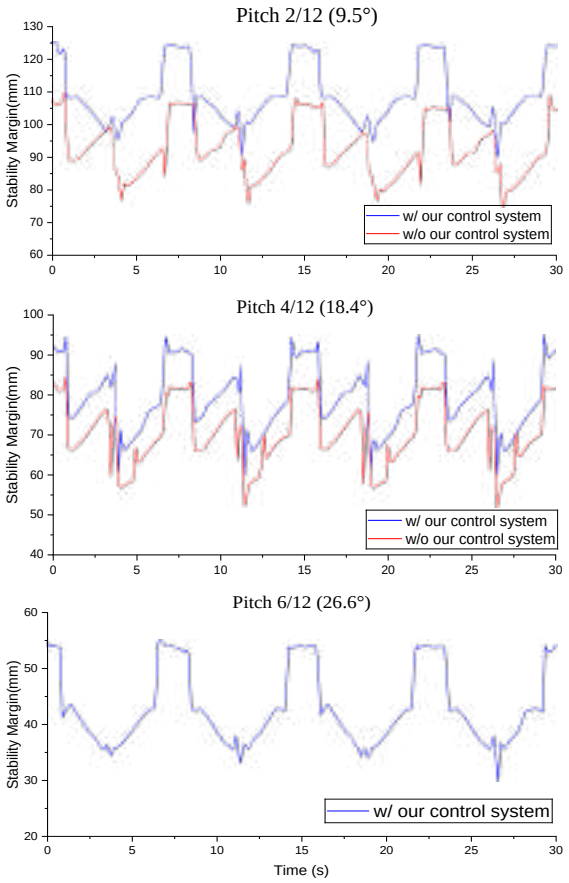


Figure 7. Stability Margin Comparison w/ and w/o Our Control System on Pitch 2/12, 4/12, and 6/12.

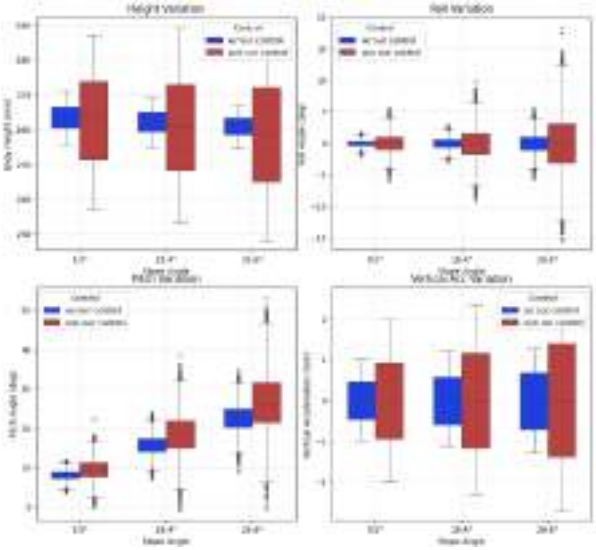


Figure 8. Orientation (Height, Roll, and Pitch) Variation and Acceleration Variation on Pitch 2/12, 4/12, and 6/12

w/ and w/o our control system achieved 100% success rate, but the trails without our control system had trouble with stepping in place for a while before going through the transition area, which lead to longer task complete time. The average crossing times shown in Table 1 varied significantly by transition type: 8.6 seconds (w/ our control system) and 11.5 seconds (w/o our control system) for ridge lines; 5.7 seconds (w/ our control system) and 6.8 seconds (w/o our control system) for valleys; 7.7 seconds (w/ our control system) and 8.9 seconds (w/o our control system) for flashing interfaces.

Table 4. Result Metrics of Test 2

Subtest	Transition Type	Success Rate (%)		Average Crossing Time (s)	
		w/ ours	w/o ours	**w/ ours**	w/o ours
#1	ridge	**80%**	50%	**8.6**	11.5
#2	valley	**100%**	60%	**5.7**	6.8
#3	flashing	**100%**	100%	**7.7**	8.9

4.2 Discussion

The results of Test 1 and Test 2 demonstrate that our integrated control system significantly enhances hexapod robot performance in both slope navigation and structural transition scenarios.

In Test 1, the system demonstrated enhanced stability across different roof slopes, with notable improvements in stability margin and body orientation control compared to the default control system. The use of a fuzzy PD controller was instrumental in maintaining consistent body orientation, particularly in challenging inclines. The improvement becomes more crucial at steeper angles, as demonstrated by the system's ability to maintain stable locomotion on 6/12 pitch slopes where default controls failed. The ability to maintain roll and pitch angles within acceptable limits highlights the system's robustness, allowing for effective traversal of various roof pitches without compromising stability. The Bézier curve-based gait trajectory further contributed to a smoother and more energy-efficient movement, reducing abrupt transitions that could destabilize the robot.

Test 2 focused on the hexapod robot's ability to navigate structural transitions commonly encountered during roof inspections, such as ridges, valleys, and flashings. The high success rates achieved across different transition types (80% for ridge lines, 100% for valleys) represent a marked improvement over conventional control methods. The reduced crossing times for all transition types indicate that our system not only improves success rates but also enhances efficiency. This is particularly evident in flashing interface navigation, where both control systems achieved 100% success rates, but our system demonstrated superior efficiency in transition completion times. The fuzzy PD controller's dynamic adjustments ensured that the robot could adapt its posture in real-time, thereby mitigating risks associated with sudden changes in terrain. The Bézier curve-based trajectory also played a crucial role in ensuring smooth foot placement during these transitions, reducing the likelihood of slips or missteps.

However, certain limitations were observed during testing. The slight decrease in stability margin with increasing slope angle suggests that additional optimization may be needed for extremely steep slopes. Furthermore, the 80% success rate in ridge line crossings, while superior to conventional methods, indicates room for improvement in handling sharp angular transitions.

5 Conclusion

This research presents an adaptive control system for hexapod robots that improves stability in complex construction environments. By integrating a fuzzy PD controller with a Bézier curve-based trajectory generator, our system maintains stability across varied terrain. This work advances construction robotics in three significant ways: (1) it demonstrates how intelligent control strategies can overcome traditional limitations of ground robots on variable construction surfaces, (2) it provides a validated framework for autonomous inspection in previously challenging environments, reducing human exposure to hazardous conditions, and (3) it establishes quantifiable performance benchmarks for robotic systems in construction inspection tasks, creating a foundation for future developments in this domain. The findings highlight the potential for deploying hexapod robots in challenging construction environments, extending beyond roof inspections to other applications that require adaptive locomotion on irregular or inclined surfaces, such as structural health monitoring of bridges, post-disaster building assessment, and maintenance of complex industrial facilities.

Future work will focus on further refining the control algorithms, including exploring machine learning-based adaptations to enhance real-time decision-making and developing a more robust physical prototype for field testing. Additionally, we plan to investigate energy efficiency improvements to extend operational time, thereby enhancing the practicality of hexapod robots for prolonged inspection and maintenance tasks. Although our evaluations were performed in the Gazebo simulation environment, these results offer a robust preliminary assessment of our method while minimizing associated risks and costs. Future work will involve validating our approach through controlled real-world experiments to further substantiate its practical applicability and performance.

References

[1] S. Shayesteh, H. Jebelli, Toward Human-in-the-Loop Construction Robotics: Understanding Workers' Response through Trust Measurement during Human-Robot Collaboration, Construction Research Congress 2022: Computer Applications, Automation, and Data Analytics - Selected Papers from Construction Research Congress 2022. 2-B (2022) pp. 631–639. https://doi.org/10.1061/9780784483961.066.

[2] D. Lattanzi, G. Miller, Review of Robotic Infrastructure Inspection Systems, Journal of Infrastructure Systems. 23 (2017) pp. 04017004. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000353.

[3] H. Peel, S. Luo, A.G. Cohn, R. Fuentes, Localisation of a mobile robot for bridge bearing inspection, Automation in Construction. 94 (2018) pp. 244–256. https://doi.org/10.1016/j.autcon.2018.07.003.

[4] X. Zhao, H. Jebelli, A computational method for real-time roof defect segmentation in robotic inspection, Computer-Aided Civil and Infrastructure Engineering. (2025). https://doi.org/10.1111/MICE.13471.

[5] A. Khan, C. Mineo, G. Dobie, C. Macleod, G. Pierce, Vision guided robotic inspection for parts in manufacturing and remanufacturing industry, Journal of Remanufacturing. 11 (2021) pp. 49–70. https://doi.org/10.1007/s13243-020-00091-x.

[6] Y. Liu, H. Jebelli, Enhanced Robotic Teleoperation in Construction Using a GAN-Based Physiological Signal Augmentation Framework, Lecture Notes in Civil Engineering. 239 (2023) pp. 295–307. https://doi.org/10.1007/978-981-19-0503-2_24.

[7] H. Miura, T. Aridome, M. Matsuoka, Deep Learning-Based Identification of Collapsed, Non-Collapsed and Blue Tarp-Covered Buildings from Post-Disaster Aerial Images, Remote Sensing. 12 (2020) pp. 1924. https://doi.org/10.3390/rs12121924.

[8] P. Sestras, G. Badea, A.C. Badea, T. Salagean, S. Roșca, S. Kader, F. Remondino, Land surveying with UAV photogrammetry and LiDAR for optimal building planning, Automation in Construction. 173 (2025) pp. 106092. https://doi.org/10.1016/J.AUTCON.2025.106092.

[9] X. Zhao, T. Ren, H. Jebelli, Adaptive Intelligence for Robot Navigation Efficiency with a Deep-Reinforcement-Learning-Based Cyber-Physical System, in: 2024 International Conference on Computing in Civil Engineering (I3CE), 2024.

[10] L. Yang, P. Li, S. Qian, H. Quan, J. Miao, M. Liu, Y. Hu, E. Memetimin, Path Planning Technique for Mobile Robots: A Review, Machines 2023, Vol. 11, Page 980. 11 (2023) pp. 980. https://doi.org/10.3390/MACHINES11100980.

[11] J. Coelho, F. Ribeiro, B. Dias, G. Lopes, P. Flores, Trends in the Control of Hexapod Robots: A Survey, Robotics 2021, Vol. 10, Page 100. 10 (2021) pp. 100. https://doi.org/10.3390/ROBOTICS10030100.

[12] Y. Liu, C. Wang, H. Zhang, J. Zhao, Research on the Posture Control Method of Hexapod Robot for Rugged Terrain, Applied Sciences 2020, Vol. 10, Page 6725. 10 (2020) pp. 6725. https://doi.org/10.3390/APP10196725.

[13] A. Ojha, Y. Liu, S. Shayesteh, H. Jebelli, W.E. Sitzabee, Affordable Multiagent Robotic System for Same-Level Fall Hazard Detection in Indoor Construction Environments, Journal of Computing in Civil Engineering. 37 (2023). https://doi.org/10.1061/(ASCE)CP.1943-5487.0001052.

[14] Muto S2 Robot, (n.d.). http://www.yahboom.net/study/Muto-S2 (accessed November 16, 2024).

[15] NVIDIA, Jetson Nano, (n.d.). https://developer.nvidia.com/embedded/jetson-nano (accessed November 16, 2024).

[16] Z.Y. Wang, X.L. Ding, A. Rovetta, Analysis of typical locomotion of a symmetric hexapod robot, Robotica. 28 (2010) pp. 893–907. https://doi.org/10.1017/S0263574709990725.

[17] Jung-Min Yang, Jong-Hwan Kim, Fault-tolerant locomotion of the hexapod robot, IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics). 28 (1998) pp. 109–116. https://doi.org/10.1109/3477.658585.

Drone-Assisted Progress and Quality Tracking for Surface Finishing in Construction

Tianyu Ren[1] and Houtan Jebelli[1]

[1] Department of Civil and Environmental Engineering, University of Illinois Urbana-Champaign, USA
tianyur2@illinois.edu, hjebelli@illinois.edu

Abstract

In the construction industry, achieving high standards in surface finishing is essential for structural integrity and aesthetic quality, yet traditional inspection methods can be slow and miss subtle defects. This paper introduces a drone-based system that utilizes advanced sensor processing for automated surface quality assessment in construction finishing tasks. The system integrates Light Detection and Ranging (LiDAR) and high-resolution RGB cameras to collect detailed surface data. The LiDAR module applies Gaussian filtering for noise reduction, followed by bicubic interpolation to generate a smooth and continuous elevation map, allowing for precise flatness estimation across the scanned surface. For visual defect detection, a Convolutional Neural Network based de-blurring algorithm is first applied to captured images to remove motion blur caused by drone movement. This process enhances image sharpness by passing the blurred image through multiple convolutional layers, where each layer reconstructs finer details, resulting in a deblurred image. Subsequently, a zero-shot transformer-based segmentation model uses multi-head self-attention to analyze spatial relationships within the deblurred image, allowing for effective identification and segmentation of defect regions without requiring extensive training data. Testing was conducted in both a ROS-based simulation and on real-life surface work. The LiDAR-based flatness estimation achieved a Root Mean Square Error (RMSE) of 0.2 mm in simulation and 0.8 mm in real-life tests, with Mean Absolute Error (MAE) values of 0.4 mm and 1.1 mm, respectively. The camera-based defect detection showed substantial improvement with the de-blurring module, increasing Intersection over Union (IoU) for defect segmentation to over 95% with de-blurring.

Keywords – Construction Robotics, Computer Vision, Unmanned Aerial Vehicle, Machine Learning

1 Introduction

Ensuring consistent surface quality is an essential aspect of construction that significantly impacts the functionality of buildings [1]. Surface finishing operations—such as plastering, painting, floor leveling, and coating—are not merely cosmetic but serve as critical elements for achieving structural performance and regulatory compliance. These processes contribute to the durability, safety, and appearance of finished structures, underscoring their importance across various construction projects, from residential homes to industrial facilities. However, maintaining high standards in surface quality remains a complex task [2], particularly as project scales and technical requirements continue to grow. The ability to accurately monitor surface conditions and promptly identify defects is crucial, not only for achieving desired design outcomes but also for reducing potential maintenance issues and ensuring the longevity of the finished surface [3].

Despite the fundamental role of surface finishing, current methods for tracking progress and assessing quality still rely heavily on manual inspections. Traditional approaches are predominantly labor-intensive, involving visual assessments and basic measurement tools like spirit levels, tape measures, and straight edges [4]. These methods, while useful for identifying major defects, are limited in their capacity to detect subtle surface imperfections, such as minor texture inconsistencies, fine cracks, or small deviations in flatness. As a result, errors can be missed, leading to compromised surface quality that may only be detected during later stages of construction or post-completion inspections. Additionally, manual inspection processes are time-consuming and often contribute to delays, especially in large-scale projects where extensive surface areas must be monitored [5]. The variability in human assessment further complicates achieving consistent quality, making it difficult to maintain uniform standards across diverse site conditions.

Given these limitations, the adoption of automation in construction has gained momentum, with drones emerging as a promising tool for progress monitoring and quality assessment in surface finishing [6]. Drones offer advantages that extend beyond traditional methods, including rapid data collection, broad area coverage, and

the ability to access elevated or hard-to-reach locations without scaffolding or other support structures [7]. Equipped with high-resolution cameras and LiDAR sensors, drones can capture detailed surface data, enabling comprehensive analysis and monitoring in real-time [8]. This capability allows for faster and more systematic inspections, improving both accuracy and consistency in quality control. By integrating drone technology into surface finishing workflows, construction teams can reduce reliance on manual processes, increase inspection speed, and enhance defect detection.

However, despite the growing use of drones for surface monitoring, existing systems still face significant challenges that limit their effectiveness. Current drone-based systems often depend on basic imaging techniques that lack the sensitivity needed to identify fine surface defects. Issues such as slight texture variations, minor cracks, and subtle flatness deviations may go undetected due to inadequate image resolution and algorithm limitations. Additionally, many drone systems are not designed to handle real-time data analysis, making them less effective in dynamic conditions where factors like changing light levels, reflective surfaces, or varying environmental conditions can influence inspection accuracy. As a result, these systems may fail to provide the precision required for complex surface assessments, prompting the need for more advanced solutions that can adapt to diverse monitoring conditions.

To overcome the limitations of traditional inspection methods in surface finishing, this paper proposes an advanced drone-assisted system for real-time progress tracking and quality assessment. The system integrates high-resolution cameras and LiDAR sensors to capture comprehensive data on surface characteristics, including smoothness, evenness, and potential defects such as cracks and texture inconsistencies. To enhance measurement accuracy, the raw sensor data undergoes pre-processing with Gaussian filtering to reduce noise, followed by calibration for reliable results.

For defect detection, a Convolutional Neural Network (CNN)-based de-blurring module first restores clarity in motion-blurred images captured by the drone, ensuring sharper details. This deblurred data is then processed by a zero-shot transformer-based segmentation model, which uses multi-head self-attention to examine spatial relationships, accurately segmenting defect areas without the need for extensive labeled training data.

In parallel, the system employs bicubic interpolation on the LiDAR data to generate a smooth elevation map, facilitating precise flatness estimation. The system's combination of enhanced sensor processing and adaptive algorithms offers a robust and efficient solution for monitoring surface quality in complex construction settings. This method minimizes the risk of rework, improves accuracy, and supports consistent adherence to quality standards, contributing to more reliable and efficient surface finishing in construction environments. Additionally, the system employs a vision transformer-based segmentation model, which reduces the need for extensive training across different test environments, enhancing adaptability in diverse construction scenarios while maintaining high defect detection accuracy.

2 Current Developments in Surface Finishing and Drone-Based Quality Tracking in Construction

2.1 Limitations of Traditional Surface Quality Monitoring Methods

Traditional methods for monitoring surface quality in construction primarily rely on manual inspections [9]. These techniques often involve visual evaluations and the use of handheld tools to measure attributes such as smoothness, flatness, and texture consistency [10]. While these approaches can be effective in identifying major defects, they are inherently limited by their reliance on human observation. Subtle surface imperfections, such as minor texture variations, small cracks, or slight deviations in flatness, are frequently overlooked during manual inspections [11], compromising the overall quality of the finished surface. Additionally, these methods are time-consuming and labor-intensive [12], making it challenging to achieve consistent results across large or complex surfaces. Irregularly shaped areas and confined spaces pose further difficulties, as inspectors may struggle to access or measure these regions accurately. This often results in uneven quality, the need for rework, and delays in project timelines [13]. As construction projects grow in scale and complexity, the demand for more accurate, efficient, and real-time quality assessment tools has become increasingly evident, necessitating a shift toward automated solutions.

2.2 Emergence of Drone Technology in Construction

Drones have increasingly been adopted in construction for a variety of applications, owing to their versatility and capability to access areas that are otherwise difficult to reach [14]. Drones have proven to be valuable tools for tasks such as site inspections, surveying, and progress tracking [15], providing real-time aerial data that can be used to monitor various aspects of a project. Equipped with high-resolution cameras, LiDAR sensors, and GPS, drones can rapidly collect comprehensive visual and spatial information [16], covering large areas more quickly than manual methods [17]. Drones offer non-invasive data collection

capabilities, allowing for detailed topographical mapping, structural inspections, and general site monitoring without the need for direct contact [18].

Despite these advantages, the use of drones for surface-specific quality monitoring in finishing operations remains relatively underdeveloped. While some research has explored drone manipulation of tools or objects for surface tasks, achieving the precision and stability needed for delicate surface analysis and adjustments has proven challenging.

2.3 This Paper's Contribution to Drone-Based Surface Monitoring and Quality Tracking

This paper addresses a critical need in construction quality control by introducing a drone-based system that integrates high-resolution LiDAR and RGB camera sensors with advanced machine learning algorithms to monitor surface quality. The proposed system improves upon existing methods by combining multiple innovations to enhance accuracy, adaptability, and efficiency in surface assessment tasks.

To improve flatness estimation, the system incorporates a LiDAR-based mapping module enhanced with Gaussian filtering to reduce noise and bicubic interpolation to generate a smooth and continuous elevation map. This approach enhances measurement precision, improving surface elevation mapping accuracy and ensuring consistent results across diverse conditions.

For defect detection, our system integrates a CNN-based de-blurring module that effectively restores motion-blurred images captured during drone operation. This enhancement improves image clarity, ensuring sharper details that are crucial for identifying subtle surface defects such as cracks and texture inconsistencies.

Additionally, the system employs a vision transformer-based segmentation model, which requires significantly less training across diverse test environments compared to conventional CNN-based models. This reduced training requirement improves the system's adaptability and scalability, minimizing the need for extensive data collection and retraining when applied to different construction scenarios. Together, these advancements improve the precision of flatness mapping, enhance defect detection through improved visual data clarity, and increase the system's adaptability with reduced training demands. The resulting framework offers a robust, efficient, and scalable solution for real-time surface quality assessment in construction environments.[19].

3 Drone-Based Real-Time Surface Quality Assessment and Defect Detection

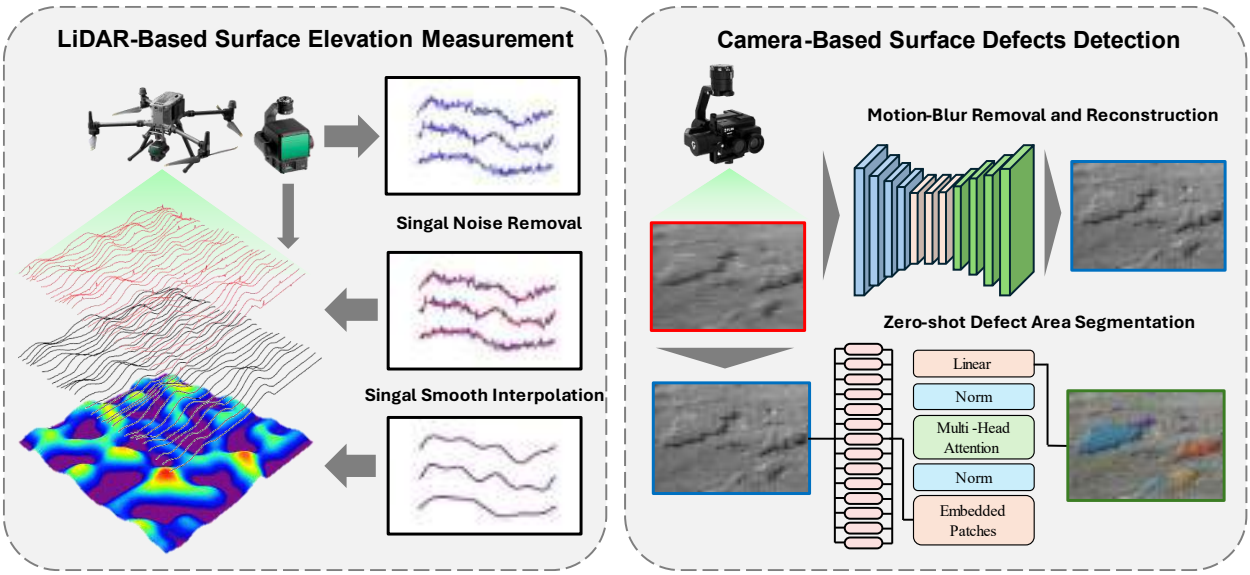


Figure 1. Drone-Based Real-Time Surface Quality Assessment and Defect Detection System Overview

The proposed UAV-based system for surface quality assessment and defect detection in construction applications is divided into two main components: LiDAR-Based Surface Elevation Measurement and Camera-Based Surface Defect Detection as shown in Figure 1. Each component is designed to address different aspects of surface monitoring, enabling the

system to capture both topographical and visual defects with high accuracy.

3.1 LiDAR-Based Surface Elevation Measurement

The LiDAR module scans the target surface to obtain detailed elevation data. This initial scan captures the surface topography, producing a series of elevation profiles that can identify height variations indicative of surface irregularities. The LiDAR data is represented as a 3D point cloud, with each point denoted by coordinates (x_i, y_i, z_i), where:

$$L = \{(x_1, y_1, z_1), (x_2, y_2, z_2), \ldots, (x_n, y_n, z_n)\} \quad (1)$$

(x_i, y_i, z_i) represents the spatial position and height of each LiDAR data point on the surface.

3.1.1 Noise Removal Module

The raw LiDAR data often contains noise due to environmental factors or sensor inaccuracies, which can lead to errors in surface mapping. To enhance accuracy, a Gaussian filtering approach is applied to the LiDAR data, effectively smoothing out noise while preserving the underlying topography. The Gaussian filter calculates a weighted average of neighboring points, giving higher importance to points closer to the target point and gradually reducing the influence of more distant points. The filtered height value $z_{\text{filtered}}(x, y)$ at each position (x, y) is computed as:

$$z_{\text{filtered}}(x, y) = \frac{1}{2\pi\sigma^2} \sum_{i=-k}^{k} \sum_{j=-k}^{k} z(x-i, y-j) \exp\left(-\frac{i^2 + j^2}{2\sigma^2}\right) \quad (2)$$

where σ is the standard deviation of the Gaussian kernel, controlling the extent of smoothing, and k is the filter size determining the neighborhood around each point (x, y). This Gaussian filtering process reduces noise effectively while preserving the surface's structural features, ensuring a more accurate representation of the surface elevation for subsequent analysis.

3.1.2 Smooth Interpolation for Elevation Estimation

After noise removal, a bicubic interpolation method is applied to fill gaps between scanned points, creating a continuous and smooth elevation map of the surface. Bicubic interpolation estimates the elevation between data points by fitting a smooth surface through the existing points, calculated as:

$$z(x, y) = \sum_{i=0}^{3} \sum_{j=0}^{3} c_{ij} x^i y^j \quad (3)$$

where c_{ij} are coefficients derived from the neighboring data points. This interpolation process generates a smooth elevation profile, allowing the system to visualize height changes continuously across the surface and detect any deviations from expected flatness, such as undulations or depressions.

3.2 Camera-Based Surface Defect Detection

The Camera Module captures high-resolution RGB images of the surface, allowing the system to detect visual defects such as cracks, texture inconsistencies, or other surface imperfections. This component involves two main steps: motion-blur removal and defect area segmentation.

3.2.1 Motion-Blur Removal Using Convolutional Neural Networks (CNN)

During UAV operations, motion blur can be introduced due to drone movement, compromising the clarity and accuracy of captured images. A Convolutional Neural Network (CNN) is applied to each image to remove any motion blur, reconstructing the image with sharp details. The blurred image $I_{\text{blurred}}(x, y)$is processed through several convolutional layers, where the output deblurred image $I_{\text{deblurred}}(x, y)$ is generated as:

$$F_l(I) = \sigma(W_l * I + b_l) \quad (4)$$

where $F_l(I)$ denotes the feature map at layer l, W_l and b_l represent the weights and biases of the filters, $***$ denotes the convolution operation, and σ is the activation function (commonly ReLU). We adopted a Multi-Scale Convolutional Neural Network architecture [20]. This architecture was chosen due to its improved stability during training, reduced computational overhead, and lower risk of generating artifacts in low-texture regions, which are common in construction environments.

The CNN model employs three parallel convolutional branches operating at different scales to effectively capture both fine textures and large structural patterns. The network integrates residual connections to improve gradient flow and enhance detail preservation, particularly in texture-rich environments such as concrete and brick surfaces. The multi-scale outputs are fused through a concatenation layer, followed by additional convolutional layers for image reconstruction. The final output layer adopts a ReLU activation function to enhance image contrast and preserve sharp details.

3.2.2 Zero-Shot Defect Area Segmentation Using Transformer Networks

After deblurring, a zero-shot transformer-based

network is used to segment defect areas without requiring extensive labeled data for training. This network utilizes a multi-head self-attention mechanism to highlight areas likely to contain defects, based on spatial features in the image. The segmented defect map $M_{\text{def}}(x, y)$ is generated by passing the deblurred image through the transformer network, resulting in a probability map that indicates the likelihood of defects at each pixel:

$$M_{\text{def}}(x, y) = \text{softmax}\left(\sum_{h=1}^{H} \text{Attention}(Q_h, K_h, V_h)\right) \quad (5)$$

where H is the number of attention heads, and Q_h, K_h, V_h represent the query, key, and value matrices for each head. This multi-head attention mechanism allows the transformer to capture nuanced spatial relationships within the image, enabling it to effectively identify defects such as cracks, discolorations, or texture inconsistencies on the surface.

4 Implementations

This section describes the implementation and testing of the two primary components of the proposed system: LiDAR-Based Surface Elevation Detection and Camera-Based Surface Defect Detection. Testing is conducted in both a ROS (Robot Operating System) simulation environment to provide controlled, reproducible conditions and on a compact real-life setup with an uneven concrete surface for real-world evaluation. These tests aim to assess the system's performance in de-blurring and segmenting surface defects, as well as accurately estimating surface flatness.

4.1 Training and Testing for Surface Defect Detection Methods

The CNN-based deblurring model was initially trained using the GoPro Blur Dataset [20] and further fine-tuned using 500 images collected from construction sites. This fine-tuning dataset captured materials such as concrete, brick, and plastered walls under varying lighting and motion conditions to better simulate realistic UAV inspection scenarios. The zero-shot transformer segmentation model was adapted from a pre-trained Vision Transformer model [21]. During evaluation, the model was tested on images collected from three construction sites featuring varied building materials (e.g., concrete, brick, and plastered walls) to assess adaptability. The vision transformer model enhances the system's adaptability by reducing the need for scenario-specific retraining. By leveraging pre-trained weights and multi-head attention mechanisms, the model efficiently identifies surface defects in environments it has not been explicitly trained on. During tests, the system successfully identified defects without additional fine-tuning.

4.2 ROS-Based Simulation for Surface Defect Detection and Flatness Estimation

The ROS environment serves as an initial testing ground for the proposed UAV system, simulating real-world construction surfaces under controlled conditions. This simulation is designed to evaluate both the LiDAR-based surface elevation detection and the camera-based defect detection modules, allowing for precise performance assessment before moving to real-world testing.

In the ROS environment, a simulated construction site surface is created with a variety of textures, irregularities, and defects. The simulated surface includes flat regions, textured areas, and sections with artificial cracks and bumps to mimic typical construction site conditions. This environment allows for thorough testing of both the LiDAR and camera modules under a variety of scenarios.

The virtual UAV in the ROS simulation is equipped with simulated RGB camera and LiDAR sensors, both of which are configured to replicate real-world data acquisition processes. The RGB camera captures high-resolution images of the surface from a consistent altitude of 2 meters. To evaluate the de-blurring module, synthetic motion blur is introduced to the images to simulate the drone's movement during flight. The LiDAR sensor, positioned alongside the camera, continuously scans the surface to create a 3D point cloud that represents surface elevation data, including controlled noise to test the noise removal and interpolation modules.

In the first phase of testing, the UAV captures both RGB images and LiDAR data as it hovers over the surface at a designated altitude.

4.3 Real-Life Test for Motion Blur Removal and Defect Segmentation

Following the ROS simulation, a real-life test is conducted to validate the de-blurring and segmentation capabilities on an actual uneven concrete surface. This real-life setup is designed to closely replicate the types of surfaces commonly encountered on construction sites, including intentional irregularities such as cracks, bumps, and texture inconsistencies. This research also used open-access dataset to validate the effectiveness and accuracy of the defect detection methodology [22]. This experiment provides insights into the system's performance under realistic lighting, material, and environmental conditions.

The experimental setup employed a quadcopter drone,

equipped with a compact Runcam Phoenix 2 camera (1000 TVL resolution, 2.1mm lens) and an Ouster OS1-32 LiDAR sensor with ±0.01° angular sampling accuracy for both vertical and horizontal directions. The UAV maintained a flight speed of 0.5 – 1.0 m/s and followed a predefined grid pattern at a height of 1.5 meters to ensure comprehensive data coverage with a 75% forward overlap and 60% lateral overlap. Data collection occurred in both indoor and outdoor construction environments under varied lighting conditions to evaluate system adaptability and robustness. Due to unavoidable minor movements during hover, moderate motion blur is naturally introduced into the images. The captured images are processed in real-time by the de-blurring module, which uses a CNN architecture to remove the motion blur and enhance clarity. This de-blurred image is then fed into the zero-shot transformer-based segmentation model to detect and isolate defect areas. The segmentation output is visualized as a colored mask, with highlighted regions indicating the presence of cracks, texture inconsistencies, or other defects. This test not only verifies the effectiveness of the de-blurring and segmentation process but also provides a realistic evaluation of how well the system performs in identifying defects on rough, uneven surfaces like concrete.

The ROS simulation is conducted in 20 different environment and the real-life test collects 200 images to gather performance metrics. Key metrics include defect detection accuracy, de-blurring effectiveness, and the system's ability to accurately estimate flatness. In each test, the accuracy of defect detection is measured by comparing the binary defect maps with known defect locations on the test surfaces. The success of flatness estimation is assessed by measuring the deviation between the estimated surface heights from the LiDAR data and the actual surface profile, with an emphasis on identifying areas with elevation deviations exceeding acceptable tolerance levels.

Through these tests, the proposed UAV-based system demonstrates its capability to accurately detect defects, segment surface irregularities, and estimate surface flatness.

5 Results

The results of the implementation are presented in this section, focusing on the performance of the LiDAR-based surface elevation measurement and camera-based surface defect detection modules.

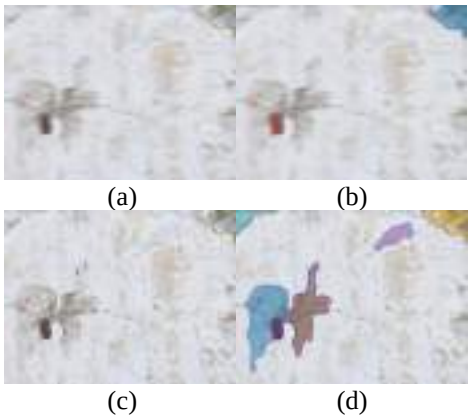
5.1 LiDAR-Based Surface Elevation Detection

To evaluate the accuracy and consistency of the LiDAR-based surface elevation detection, several metrics were calculated to compare the estimated surface elevation data with the actual physical surface measurements obtained from the test environment. The main metrics used were Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), both of which quantify the deviation between the LiDAR-generated elevation map and the actual surface profile.

Root Mean Square Error (RMSE): The RMSE was calculated to quantify the average magnitude of errors in the LiDAR-derived elevation data. This metric reflects how well the LiDAR data matches the actual surface heights across all tested points. The RMSE for the LiDAR-based flatness estimation was observed to be approximately 0.2 mm in the ROS simulation and 0.8 mm in the real-life test, indicating a high degree of accuracy for both controlled and real-world environments.

Mean Absolute Error (MAE): The MAE provides an alternative error measurement, representing the average of absolute differences between estimated and actual surface heights. The MAE for the flatness estimation was 0.4 mm in the ROS environment and 1.1 mm in the real-life setup, further demonstrating the LiDAR system's consistency in representing surface elevation.

Consistency Across Multiple Runs: To assess consistency, the LiDAR module's flatness measurements were repeated across multiple runs in both the simulated and real-life environments. The Standard Deviation of the RMSE across five independent runs was 0.1 mm in the ROS simulation and 0.4 mm in the real-life test, demonstrating consistent performance with minimal variation, even in the presence of environmental noise.



(a) (b)

(c) (d)

Figure 2. Example Real-life Image Processing. (a) Initial Motion Blurred Image. (b) Segmentation Result without Deblurring (c) Deblurred Image with CNN (d) Segmentation Result after Deblurring

5.2 Camera-Based Surface Defect Detection

The effectiveness of the de-blurring module can be observed in Figure 2. The first image (a) represents the

raw data captured by the camera, where significant motion blur obscures surface details, making it challenging to identify surface defects accurately. Texture and fine structural elements, crucial for detecting cracks or bumps, are blurred, which would likely hinder the segmentation module's ability to accurately isolate defect areas. The third image (c), after processing through the CNN-based de-blurring module, shows a marked improvement in clarity and sharpness. Fine textures and subtle surface irregularities that were previously obscured are now distinctly visible. The improved clarity allows for enhanced contrast in texture features, essential for precise defect detection. This qualitative improvement demonstrates the critical role of the de-blurring module in enhancing the fidelity of image data, ultimately supporting higher segmentation accuracy and ensuring robust defect detection.

The effectiveness of the camera-based surface defect detection module was evaluated by measuring the Precision, Recall and Intersection over Union (IoU), for the segmentation results. The experiments compared segmentation performance with and without the de-blurring module, revealing a significant improvement in defect detection accuracy when the de-blurring process was applied.

Precision was calculated to assess the proportion of correctly identified defect pixels out of all pixels classified as defects. Without de-blurring, precision was observed to be around **78%**, indicating a high rate of false positives due to motion blur. After de-blurring, precision improved to **96%**, showing that the de-blurring module reduced false positives significantly, enabling the system to accurately classify defect regions. **Recall** was also used to measure the proportion of actual defect pixels that were correctly identified. The recall score without de-blurring was around **15%**, highlighting that many true defect areas were missed due to image blurriness. With the de-blurring module, recall improved to **94%**, indicating that the system could accurately capture almost all defect areas present on the surface.

Intersection over Union (IoU) was used as the primary metric to evaluate segmentation accuracy, measuring the overlap between the predicted defect regions and the ground truth defect areas. Without applying the de-blurring module, the IoU for defect segmentation was approximately **28%**, indicating a high degree of misalignment between the segmented defects and actual defect locations due to motion blur in the images. However, with the de-blurring module applied, the IoU increased dramatically to over **95%**, showing that the de-blurred images allowed for accurate identification of defect areas, with high consistency between predicted and actual defects.

6 Conclusions

This paper presents a novel UAV-based system designed for real-time surface quality assessment in construction, integrating LiDAR for flatness estimation and a camera module for defect detection. The implementation of a CNN-based de-blurring module significantly enhanced defect segmentation accuracy, achieving an IoU improvement from 28% to over 95%, by restoring image clarity affected by motion blur. The LiDAR module demonstrated high consistency with actual surface elevations, with low RMSE and MAE values, validating its effectiveness in accurately representing surface topography. Field tests in both simulated and real-world environments confirmed the system’s robustness, achieving reliable performance in detecting surface defects and estimating flatness.

Future research directions include conducting more comprehensive real-life field tests in diverse construction environments to evaluate the system's adaptability to varying lighting, weather, and material conditions. Additionally, exploring the integration of advanced machine learning techniques, such as self-supervised learning or transformer-based models, could further improve defect detection and flatness estimation across a broader range of surface types.

References

[1] X. Pan, Z. Shi, C. Shi, T. C. Ling, and N. Li, “A review on concrete surface treatment Part I: Types and mechanisms,” *Constr Build Mater*, vol. 132, pp. 578–590, Feb. 2017, doi: 10.1016/J.CONBUILDMAT.2016.12.025.

[2] J. Liu, Y. Lee, and J. Bard, “Material characterization of workability and process imaging for robotic concrete finishing,” *Construction Robotics 2021 5:1*, vol. 5, no. 1, pp. 73–85, Feb. 2021, doi: 10.1007/S41693-021-00058-9.

[3] K. Amini, H. Ceylan, and P. C. Taylor, “Effect of finishing practices on surface structure and salt-scaling resistance of concrete,” *Cem Concr Compos*, vol. 104, p. 103345, Nov. 2019, doi: 10.1016/J.CEMCONCOMP.2019.103345.

[4] H. A. Heinemann and W. J. Quist, “Stone and concrete: A review of the coevolution of the surface finishes of two buildings materials,” *Building Knowledge, Constructing Histories*, vol. 1, pp. 141–148, 2018, doi: 10.1201/9780429506208-20.

[5] Alf Gardiner, “Solving Finishing Problems| Concrete Construction Magazine.” Accessed: Nov. 09, 2024. [Online]. Available:

https://www.concreteconstruction.net/how-to/solving-finishing-problems_o

[6] T. Ren and H. Jebelli, "Developing an Advanced Perception System for Robotic Attic Insulation Using Large Vision-Language Models and LiDAR," in *2024 International Conference on Computing in Civil Engineering (i3CE)*, Pittsburgh, 2025.

[7] T. Ren and H. Jebelli, "Efficient 3D robotic mapping and navigation method in complex construction environments," *Computer-Aided Civil and Infrastructure Engineering*, 2024, doi: 10.1111/MICE.13353.

[8] M. N. A. Gulshan Taj, R. Prema, S. Anand, A. Haneefi, R. P. Kanishka, and D. H. A. Mythra, "Monitoring and Management of Construction Sites Using Drone," *Proceedings of International Conference on Innovative Technologies for Clean and Sustainable Development (ICITCSD – 2021)*, pp. 705–720, 2022, doi: 10.1007/978-3-030-93936-6_57.

[9] J. W. Kim, A. Golabchi, S. U. Han, and D. E. Lee, "Manual operation simulation using motion-time analysis toward labor productivity estimation: A case study of concrete pouring operations," *Autom Constr*, vol. 126, p. 103669, Jun. 2021, doi: 10.1016/J.AUTCON.2021.103669.

[10] B. Bissonnette, L. Courard, and A. Garbacz, "Concrete surface engineering," *Concrete Surface Engineering*, pp. 1–244, Jan. 2015, doi: 10.1201/B19372/CONCRETE-SURFACE-ENGINEERING-BENOIT-BISSONNETTE-LUC-COURARD-ANDRZEJ-GARBACZ.

[11] B. Liu and T. Yang, "Image analysis for detection of bugholes on concrete surface," *Constr Build Mater*, vol. 137, pp. 432–440, Apr. 2017, doi: 10.1016/J.CONBUILDMAT.2017.01.098.

[12] Ivana Hermida, "Concrete Finishing: 3 Tools That Will Make Your Life Easier – Tomahawk Power," Tomahawk Power. Accessed: Nov. 10, 2024. [Online]. Available: https://tomahawk-power.com/blogs/articles/concrete-finishing-3-tools-that-will-make-your-life-easier?srsltid=AfmBOop_yy44-nkybidAgZjoeg2X_7SIW7m_KxHiMF_e4p6lqsSX_j4r

[13] A. T. Yudhistira, I. Satyarno, A. S. B. Nugroho, and T. N. Handayani, "Effect of Construction Delays and the Preventive Role of Concrete Works Optimization: Systematic Literature Review," *TEM Journal*, vol. 13, no. 2, p. 1203, May 2024, doi: 10.18421/TEM132-34.

[14] H. W. Choi, H. J. Kim, S. K. Kim, and W. S. Na, "An Overview of Drone Applications in the Construction Industry," *Drones 2023, Vol. 7, Page 515*, vol. 7, no. 8, p. 515, Aug. 2023, doi: 10.3390/DRONES7080515.

[15] Y. Li and C. Liu, "Applications of multirotor drone technologies in construction management," *International Journal of Construction Management*, vol. 19, no. 5, pp. 401–412, Sep. 2019, doi: 10.1080/15623599.2018.1452101.

[16] T. Ren and H. Jebelli, "Robotic Navigation and Obstacle Avoidance System for Advancing Attic Insulation," in *2024 International Conference on Computing in Civil Engineering (i3CE)*, Pittsburgh, 2025.

[17] S. Shin, O. Kwon, and C. B. of, "Example of Application of Drone Mapping System based on LiDAR to Highway Construction Site," *Journal of the Korean Society*, 2023, Accessed: Nov. 09, 2024. [Online]. Available: https://koreascience.kr/article/JAKO202305361450639.page

[18] P. Nooralishahi *et al.*, "Drone-Based Non-Destructive Inspection of Industrial Sites: A Review and Case Studies," *Drones 2021, Vol. 5, Page 106*, vol. 5, no. 4, p. 106, Sep. 2021, doi: 10.3390/DRONES5040106.

[19] A. Kirillov *et al.*, "Segment Anything," Apr. 2023, Accessed: Jan. 24, 2024. [Online]. Available: https://arxiv.org/abs/2304.02643v1

[20] S. Nah, T. H. Kim, and K. M. Lee, "Deep Multi-scale Convolutional Neural Network for Dynamic Scene Deblurring," *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-January, pp. 257–265, Dec. 2016, doi: 10.1109/CVPR.2017.35.

[21] M. Caron *et al.*, "Emerging Properties in Self-Supervised Vision Transformers," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 9630–9640, Apr. 2021, doi: 10.1109/ICCV48922.2021.00951.

[22] M. Mundt, S. Majumder, S. Murali, P. Panetsos, and V. Ramesh, "Meta-Learning Convolutional Neural Architectures for Multi-Target Concrete Defect Classification With the COncrete DEfect BRidge IMage Dataset," 2019.

Path tracking of bucket edge using a hydraulic excavator with a built-in angular velocity controller

Daisuke Endo[1], Genki Yamauchi[1], Takeshi Hashimoto[1], Yoshiaki Tsutsumi[2], Shinjiro Yamamoto[2], Kenta Furuuchi[2] and Shinya Imura[2]

[1]National Research and Development Agency Public Works Research Institute,
Construction Technology Research Department Advanced Technology Research Team, Japan
[2]Hitachi Construction Machinery Co., Ltd.,
Research and Development Department Advanced Development Center, Japan

endou-d177cl@pwri.go.jp, yamauchi-g573bs@pwri.go.jp, t-hashimoto@pwri.go.jp,
y.tsutsumi.ii@hitachi-kenki.com, s.yamamoto.jp@hitachi-kenki.com, k.furuuchi.ur@hitachi-kenki.com,
s.imura.hk@hitachi-kenki.com

Abstract -

This study presents the development and evaluation of a hydraulic excavator equipped with a built-in angular velocity controller to improve the precision of excavation operations. By integrating a velocity control system that takes into account the characteristics of the hydraulic system, the proposed approach enhances the path-tracking accuracy of the bucket edge during digging tasks. Experimental results demonstrate that the angular velocity controller significantly reduces path errors compared to conventional lever-based control systems, achieving improvement in path accuracy. This advancement facilitates the automation of hydraulic excavators, and provides a reliable foundation for further developments in construction robotics.

Keywords -

Motion Control; Hydraulic System Dynamics; Excavation Path Tracking;

1 Introduction

In recent years, there has been a growing demand for the automation of construction machinery, driven by labor shortages in construction and industrial plants, particularly the declining availability of skilled workers. To address this issue, technologies enabling the autonomous operation of construction machinery, such as hydraulic excavators performing excavation tasks automatically, have been developed [1].

Previous automation technologies were typically built upon conventional machines designed for manual operation. These systems relied on human-optimized interfaces—lever manipulation, pedal operation, and steering inputs—and replaced human inputs with a controller mounted externally on the machine. Moreover, feedback elements, such as rotary angle sensors, optical cameras, LiDAR, and GNSS, were often added to the machinery to determine control inputs, which required additional hardware installations. However, this approach presents several challenges:

- Nonlinear elements optimized for human operation (e.g., play, sensitivity adjustments) make it difficult to enhance servo system responsiveness[2].
- The cost of modifying the hardware of the machinery is significant for users.
- Modeling and parameter identification necessary for controlling the hydraulic excavator system are technically complex for users[3].

To overcome these issues, it is desirable for hydraulic excavator manufacturers to incorporate controllers into the machinery itself, taking into account the system's characteristics and enabling users to provide input directly through interfaces. This integration facilitates automation by treating the excavator's components as a four-degree-of-freedom manipulator and developing a control framework similar to that of industrial robotic arms. This facilitates the ease of automation for hydraulic excavator. This approach is expected to enable highly precise control of the bucket's edge path. If such technology is realized, it could automate hydraulic excavator operations that require strict conformance to design specifications, such as shaping embankment slopes. Additionally, it would facilitate autonomous operations in environments with surrounding obstacles, allowing the excavator to avoid contact with them effectively.

Hitachi Construction Machinery has previously provided machine control functions, such as the automatic control of boom movements to ensure the bucket of a hydraulic excavator follows a target surface [4]. Additionally, the company developed the ZCORE® system platform, which allows for easy function expansion and customization based on automation needs [5]. By enabling hydraulic excavators to operate using common control

messages proposed by the Public Works Research Institute [6], it is anticipated that construction operators will be able to command multiple hydraulic excavators from different manufacturers using a unified system, thereby promoting the widespread adoption of automation. In September 2022, Hitachi Construction Machinery (Hitachi CM) and the Public Works Research Institute (PWRI) launched a joint research project aimed at improving the productivity of mechanized earthwork using the OPERA autonomous construction technology open platform [6]. In this project, Hitachi Construction Machinery developed a hydraulic excavator equipped with a built-in angular velocity controller, taking into account the characteristics of the hydraulic system (Figure 1). This paper confirms the effectiveness of the developed system in advancing the automation of hydraulic excavators and is organized as follows.

In Chapter 2, the system architecture for controlling the path of the edge of the bucket is described. We present an example of a conventional approach where a servo system outputs lever operations from an external motion controller of the excavator. In contrast, an example of the proposed approach using the developed velocity controller, which outputs target joint angular velocity from an external motion controller, is also explained.

Chapter 3 provides a detailed explanation of the developed velocity controller. Chapter 4 reports on the experimental validation by performing excavation tasks using both the conventional and proposed methods, comparing the behavior of each actuator in the excavator. Finally, Chapter 5 concludes the findings of this study.



Figure 1. Developed excavator built-in velocity controller.

2 System configuration for bucket edge path control

This chapter describes the system configuration required to automate the bucket edge path control of a hydraulic excavator. The excavator used in this study is equipped with an electronically controlled hardware system capable of receiving pilot pressure commands corresponding to lever operations. Figure 2 illustrates the hardware system configuration.

2.1 Conventional method: Lever operation-based commands

In the conventional approach, the hydraulic excavator is controlled using lever operation commands as the operational input. The system architecture is shown in Figure 3-(a). A motion planner sequentially calculates the target joint angles required to achieve the desired bucket edge pose. These angles are then controlled by a Joint Position Controller, which outputs lever operation commands to the Joint Control Unit (JCU). The JCU sends the vehicle control data to the Motion Control Unit (MCU) as feedback data and generates target pilot pressure for the actuators.

2.2 Proposed method: Angular velocity-based commands

The proposed method replaces lever operation commands with angular velocity commands as the operational input. Figure 3-(b) illustrates this system configuration. The motion planner calculates the sequential target joint angles for achieving the desired bucket edge pose. These angles are converted into target angular velocities by the joint position controller and transmitted to the JCU. Within the JCU, a dedicated velocity control algorithm calculates the necessary target pilot pressure and sends commands to the solenoid valve controller, thereby driving the hydraulic system. The output of the Hydraulic Pressure Sensors in Figure 2 is used to measure the load applied to the actuator and improve the accuracy of angular velocity control. Details are described in Chapter 3.

The proposed method offers significant advantages over the conventional approach. By directly controlling the angular velocity of each joint, the system improves path tracking accuracy and responsiveness. Moreover, this method simplifies the control architecture by reducing dependency on external hardware, enabling easier integration into automated systems.

3 Joint velocity controller considering hydraulic system characteristics

This chapter details the joint velocity controller that considers the characteristics of the hydraulic system built

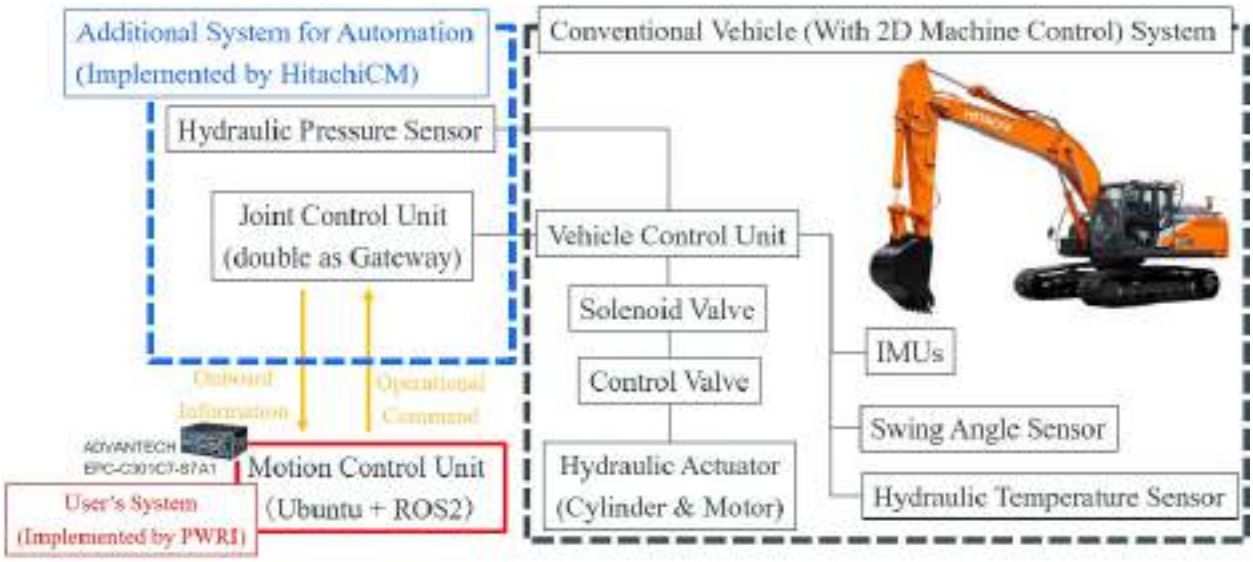


Figure 2. Hardware system configuration of the hydraulic excavator.

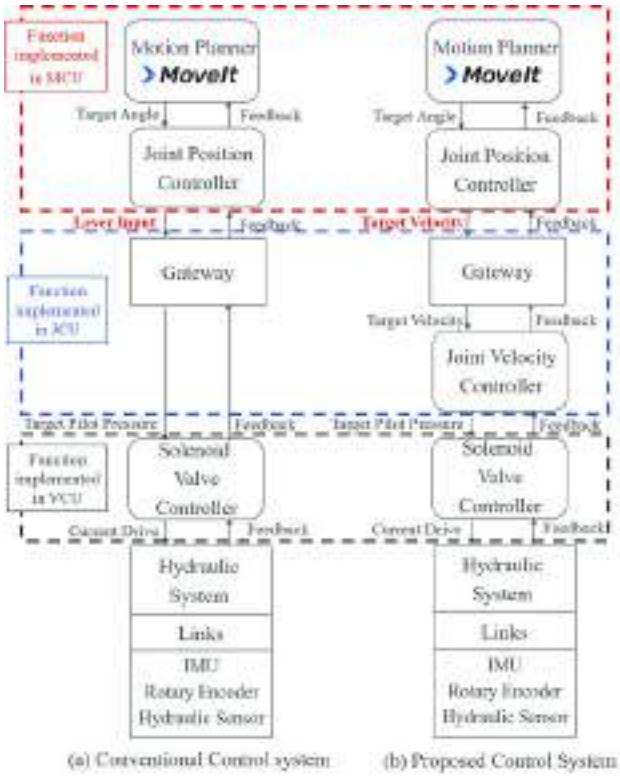


Figure 3. Schematic diagram of automated system for excavator.

into the excavator. As shown in Figure 4, this controller consists of three computational processes: (a) Link calculation, (b) Pump flow rate calculation, and (c) Valve opening calculation. Each of these processes is described in the following subsections.

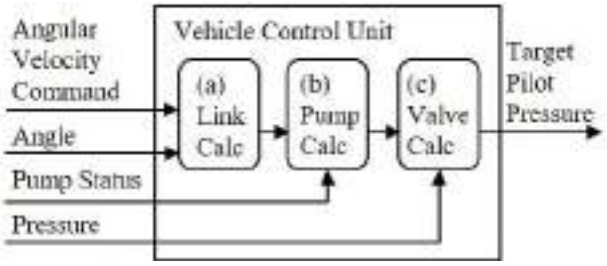


Figure 4. Schematic diagram of the Joint Velocity Controller.

3.1 Link calculation

As an example, the relationship between the boom angle and the boom cylinder length is explained. The kinematic relationship between the boom and the boom cylinder is shown in Figure 5. The relative angle $\angle BAC$ of the boom to the chassis and the boom cylinder length BC can be expressed by Equation (1).

$$\overline{BC}^2 = \overline{AB}^2 + \overline{AC}^2 - 2\overline{AB} \cdot \overline{AC} \cos \angle BAC \qquad (1)$$

The relationship between the angular velocity ω of the boom and the boom cylinder speed can be derived by differentiating Equation (1) with respect to time, resulting in Equation (2).

$$\frac{d\overline{BC}}{dt} = \frac{\overline{AB} \cdot \overline{AC} \sin \angle BAC}{\overline{BC}} \omega \tag{2}$$

Assuming the hydraulic fluid is incompressible and the cylinder has a circular cross-section, the boom cylinder speed obtained in Equation (2) can be converted to the flow rate Q of the hydraulic fluid using the cylinder cross-sectional area A_{cyl}, as expressed in Equation (3).

$$Q = A_{\text{cyl}} \frac{d\overline{BC}}{dt} \tag{3}$$

In this manner, the angular velocity command can be converted to the flow rate Q of hydraulic fluid supplied to the actuator from the pump.

When an angular velocity command is input to the JCU, a process is implemented to calculate the optimal target pilot pressure required for operation at the target angular velocity, taking into account the pump output and the load on the actuator. Figure 6 shows a schematic diagram of the power transmission to the actuator. The changes in joint angles of each link can be converted into the volume of incompressible hydraulic fluid supplied from the pump to the actuator, which is a cylinder. Based on this, the flow rate of the actuator corresponding to the commanded angular velocity can be calculated.

$$Q = CA \sqrt{\frac{2\Delta P}{\rho}} \tag{4}$$

The flow rate Q_{in} from the pump to the actuator is generally expressed by the orifice equation (4), which relates the opening area A of the spool on the path to the pressure difference before and after it. Therefore, the spool opening area A required to achieve the target angular velocity can be calculated. In equation (4), C is the flow coefficient, ρ is the fluid density, and ΔP is the differential pressure across the spool opening ($P_P - P_{\text{in}}$). The relationship between the spool opening area A and the pilot pressure is determined by the opening characteristics of each spool.

3.2 Pump flow rate calculation

Hydraulic excavators are generally designed with hydraulic circuits that allow a single pump to operate multiple actuators. Therefore, when simultaneous operation commands are input to multiple joints, calculating the target pilot pressure for each command as if operating a single joint may result in the total flow exceeding the output capacity of a single pump. In such cases, differences in the load on each actuator may lead to an uneven distribution of flow, causing one side to receive more flow and making it impossible to achieve the target angular velocity. Particularly, if the ratio of the current value to the target value differs among joints, the bucket tip may deviate from the path planned by the Motion Controller. This issue can be critical in use cases where avoiding contact with obstacles around the hydraulic excavator is required.

On the other hand, even if there is a deviation between the current value and the target value for each joint, as long as the ratio of the current value to the target value remains consistent across all joints, the planned path can be maintained. This is considered desirable in use cases that involve avoiding surrounding obstacles. To address this, when angular velocity commands are simultaneously given to multiple joints, a control algorithm is implemented in the Joint Velocity Controller to determine the target pilot pressure. This algorithm ensures that the ratio of angular velocities is maintained such that the total flow does not exceed the pump's discharge capacity.

Figure 7 shows an example of a circuit where flow is distributed to multiple actuators from a single pump. When the pressure in the hydraulic line does not reach the bleed-off pressure, the bleed-off flow Q_T flowing into the tank in the figure can be considered as $Q_T = 0$. In this case, as shown in Equation (5), the pump flow Q_P is equal to the sum of the actuator flows Q_{in1} and Q_{in2}.

$$Q_p = Q_{\text{in1}} + Q_{\text{in2}} \tag{5}$$

On the other hand, based on the linkage mechanism calculations described in Section 3.1, the required flows $Q_{\text{Req in1}}$ and $Q_{\text{Req in2}}$ for each actuator at the commanded angular velocity can be determined.

$$\begin{aligned} \text{If } Q_p &< Q_{\text{Req in1}} + Q_{\text{Req in2}}, \\ Q_{\text{in1}} &= \frac{Q_{\text{Req in1}}}{Q_{\text{Req in1}} + Q_{\text{Req in2}}} Q_p, \qquad (6) \\ Q_{\text{in2}} &= \frac{Q_{\text{Req in2}}}{Q_{\text{Req in1}} + Q_{\text{Req in2}}} Q_p, \qquad (7) \end{aligned}$$

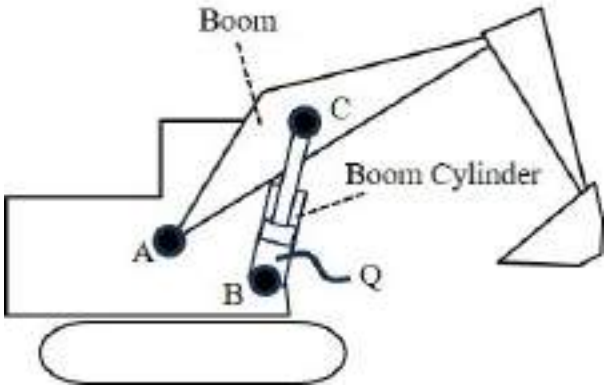


Figure 5. Kinematic relationship between the boom and the boom cylinder.

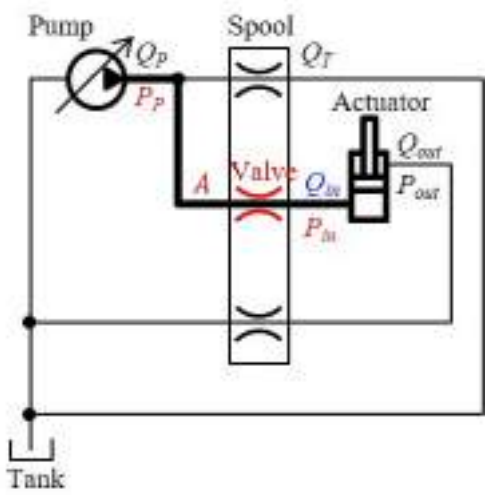


Figure 6. Hydraulic circuit diagram of power transmission from pump to actuator(Single pump to single cylinder).

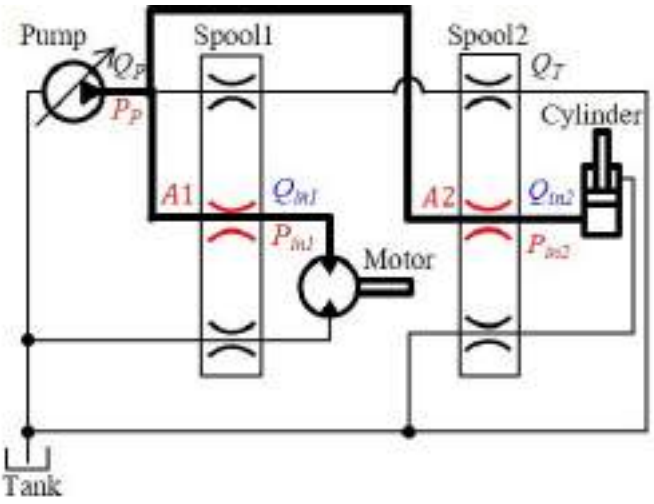


Figure 7. Hydraulic circuit diagram of power transmission from pump to actuator(Single pump to double cylinder).

As shown in Equation (6) and (7), if the sum of the required flows $Q_{\text{Req in1}}$ and $Q_{\text{Req in2}}$ exceeds the pump's discharge capacity Q_P, the target actuator flows Q_{in1} and Q_{in2} are adjusted by multiplying them by the ratio of the pump's discharge capacity to the required flow. By doing so, deviations of the bucket tip from the target path can be suppressed within the pump's discharge capacity.

$$Q_{\text{in1}} : Q_{\text{in2}} = Q_{\text{Req in1}} : Q_{\text{Req in2}}, \tag{8}$$

$$\omega_1 : \omega_2 = \omega_{\text{Req1}} : \omega_{\text{Req2}}, \tag{9}$$

$$Q_{\text{in1}} = CA1 \sqrt{\frac{2(P_p - P_{\text{in1}})}{\rho}}, \tag{10}$$

$$Q_{\text{in2}} = CA2 \sqrt{\frac{2(P_p - P_{\text{in2}})}{\rho}}, \tag{11}$$

$$A1 = \frac{1}{C} \frac{Q_{\text{Req in1}} Q_p}{Q_{\text{Req in1}} + Q_{\text{Req in2}}} \sqrt{\frac{\rho}{2(P_p - P_{\text{in1}})}}, \tag{12}$$

$$A2 = \frac{1}{C} \frac{Q_{\text{Req in2}} Q_p}{Q_{\text{Req in1}} + Q_{\text{Req in2}}} \sqrt{\frac{\rho}{2(P_p - P_{\text{in2}})}}. \tag{13}$$

3.3 Valve opening calculation

The flow rate Q_{in} flowing into the actuator from the pump is generally expressed by the orifice equation (Equation (4)) using the valve opening area A of the spool and the pressure difference ΔP before and after the valve. Therefore, the valve opening area A required to achieve the target angular velocity can be calculated. The relationship between the valve opening area A and the pilot pressure is uniquely determined by the valve characteristics.

4 Experimental verification of conventional and proposed methods

4.1 Experimental purpose

The objective of this experiment was to confirm the effectiveness of the velocity controller, which takes into account the characteristics of the hydraulic system of the excavator's working mechanism. The comparison was made with the conventional method outlined in Section 2.1, which uses simple proportional controllers (P controllers) on the external *Joint Position Controller* of the excavator. The conventional method does not consider hydraulic system characteristics and relies on lever inputs as commands.

4.2 Experimental procedure

The experiment involved excavating a flat surface using the hydraulic excavator. During this process, errors in the cutting-edge path and joint angle control (the difference between the commanded and actual values) were measured. The excavator was placed on a nearly level ground surface. To standardize the hardness of the ground, the excavation area was manually dug beforehand to loosen the soil, and the terrain was leveled to ensure that the plane of the excavation surface matched the ground plane of the excavator.

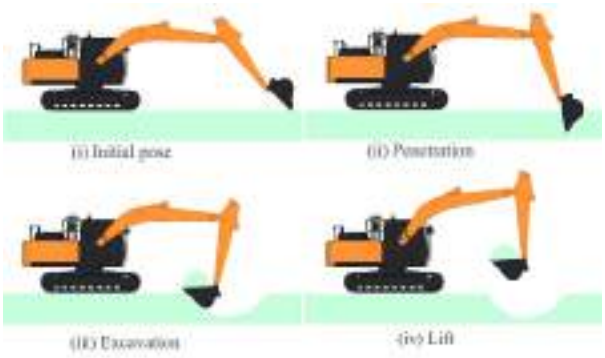


Figure 8. Display of test movements: (i) Initial pose, (ii) Penetration, (iii) Excavation, (iv) Lift.

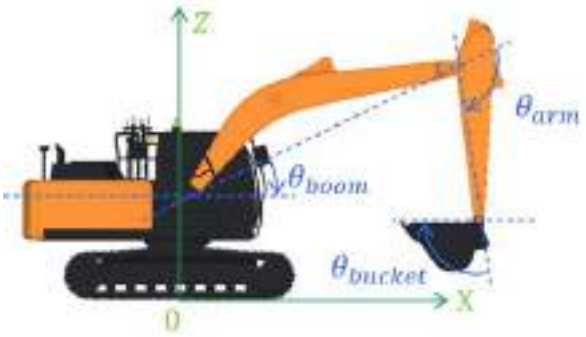


Figure 9. The definition of each coordinate.

The experimental motions were carried out sequentially as shown in Figure 8 from states (i) to (iv). The target joint angles for each state are summarized in Table 1. The definition of each joint angle and cartesian coordinates system are shown in Figure 9 .

During the transitions between these states, the commanded and actual joint angles (calculated using sensors such as IMUs) were recorded. The recorded joint angles were then used to calculate the cutting-edge path using forward kinematics. This allowed the path traversed during the experiment to be determined. Deformations caused by external forces acting on the excavator's links during motion were not considered.

Table 1. Target joint angles for each state (i)-(iv)

State	Boom	Arm	Bucket
(i) Initial Pose	-0.24 rad	1.19 rad	-0.07 rad
(ii) Penetration	-0.24 rad	1.57 rad	-0.07 rad
(iii) Excavation	-0.24 rad	1.92 rad	1.36 rad
(iv) Lift	-0.44 rad	1.92 rad	1.55 rad

4.3 Experimental conditions

Two experimental conditions were evaluated to compare the path followed by the cutting-edge:

1. **Condition 1: Conventional method (Without velocity control)**
   The method described in Section 2.1 was employed, where lever inputs were commanded to the excavator. The target angles for the cutting-edge were calculated sequentially using *MoveIt2*[7], and these values were input to the P controllers outside the excavator.

2. **Condition 2: Proposed method (With velocity control)**
   The method described in Section 2.2 was employed, where target joint angular velocities were directly commanded to the excavator. The target angular velocities required to achieve the desired cutting-edge pose were calculated and directly input to the excavator.

The various parameters related to the motion controller, MoveIt2, used in the experiment are summarized in Table 2. In this table, θ_{limit}^{upper} and θ_{limit}^{lower} represents upper or lower limit angles of the range of motion, $\dot{\theta}_{limit}$ represents angular velocity limit, and $\ddot{\theta}_{limit}$ represents angular acceleration limit, for each joint respectively. Additionally, variables K_P, K_I and K_D with the superscript "lev" and "vel" represent the P, I, and D gains of the joint position controller in conventional method or proposed method respectively. Essentially, in this experiment, the joint position controller is a pure proportional controller in both conditions. This is a measure to simplify the comparison between the conventional method and the proposed method.

The tolerance value for the control error of joint angles was set to 0.03 rad for all axes and tests. This means that the target poses from (i) to (iv) are switched only when the control errors of all joint axes fall within this allowable range. The control timeout was set to 30 seconds. This means that if 30 seconds have elapsed since the control target value stabilized, the bucket edge target position will be forcibly switched to the next target position, even if a control error exceeding the tolerance remains.

Table 2. Parameters related to the motion controller

Parameter	Boom	Arm	Bucket
θ_{limit}^{upper}	0.349 rad	2.687 rad	2.372 rad
θ_{limit}^{lower}	-1.137 rad	0.594 rad	-0.622 rad
$\dot{\theta}_{limit}$	0.3 rad/s	0.7 rad/s	1.0 rad/s
$\ddot{\theta}_{limit}$	0.2 rad/s^2	0.7 rad/s^2	1.0 rad/s^2
K_P^{lev}	1.7	1.4	0.6
K_I^{lev}	0.0	0.0	0.0
K_D^{lev}	0.0	0.0	0.0
K_P^{vel}	1.4	1.5	1.0
K_I^{vel}	0.0	0.0	0.0
K_D^{vel}	0.0	0.0	0.0

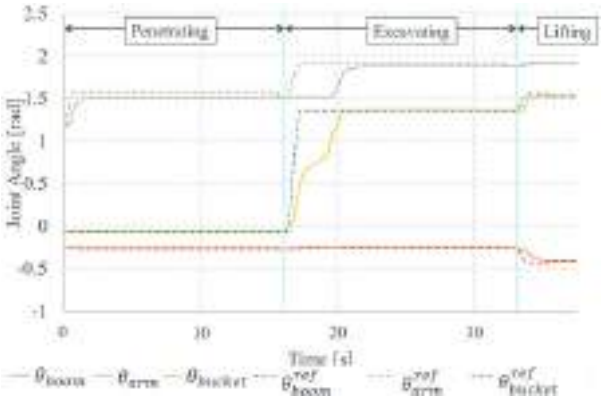


Figure 10. Angle profile in conventional control system.

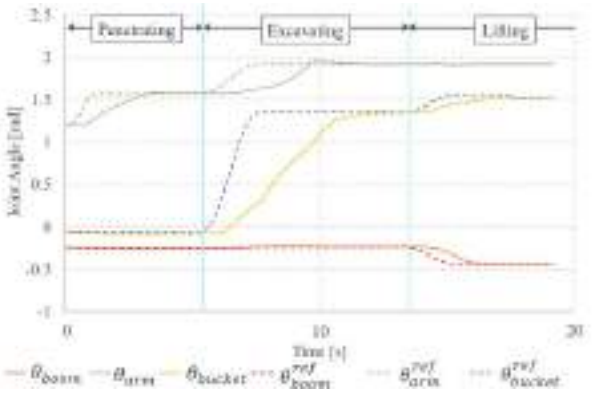


Figure 11. Angle profile in proposed control system.

4.4 Experimental results

Figure 10 and 11 present the angle profiles (commanded and actual) for the conventional and proposed methods, respectively. Figure 12 shows the path of the cutting-edge for both methods compared to the target path.

The horizontal axis (X) represents the forward direction from the excavator's crawler as projected onto the ground, while the vertical axis (Z) represents the upward direction. The line $Z = 0$ corresponds to the ground plane. The operational modes are labeled as *Penetration*, *Excavation*, and *Lift*, as indicated in Figure 8.

From Figure 12, it is evident that during *Penetration* and *Excavation*, the proposed method achieved trajectories closer to the target than the conventional method. During the *Lift* phase, both methods performed comparably, indicating that the proposed method maintained path accuracy even under reduced load conditions.

Next, Figure 13 shows the results of comparing the control errors of the bucket edge path between the conventional and the proposed method. The vertical axis represents the control error, which is the Root Mean Square (RMS) value of the bucket edge position. The calculation was performed across the entire bucket edge path. It can be observed that the proposed method reduces the error to 41% compared to the conventional method.

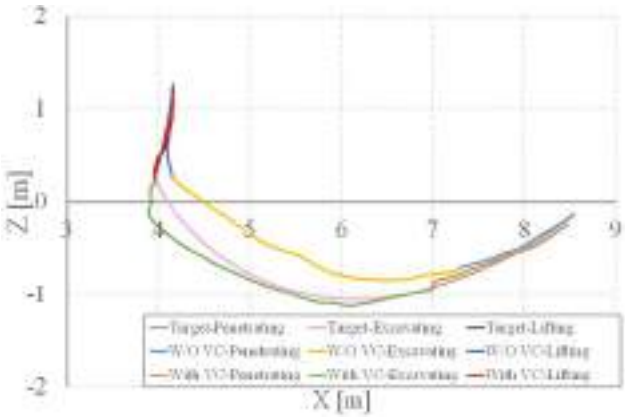


Figure 12. Cutting-edge path comparison: conventional vs. proposed methods.

4.5 Discussion

As shown in Figure 12, the proposed method effectively maintained cutting-edge path closer to the target than the conventional method, particularly during the *Penetrating* and *Excavating* phase. Analysis of the control profiles in Figs. 10 and 11 indicates that the conventional method experienced delays, particularly for the arm joint, where *Excavating* phase started, for approximately 3 seconds after receiving the command. This discrepancy can be attributed to unbalanced oil flow in the conventional method, as lever inputs cannot ensure balanced motion across multiple joints. Furthermore, at the end of both the *Penetrating* and *Excavating* phases, the arm joint exhibited a steady-state deviation exceeding the tolerance range of 0.03 rad relative to the target value, which persisted for more than 10 seconds, resulting in a prolonged operating time.

Conversely, the proposed method mitigated this issue by utilizing angular velocity control via joint velocity controller, allowing the system to compensate for variations in resistance and disturbances encountered during excavation. This feature ensured adherence to the target path, which was unattainable with the conventional proportional control approach. In the conventional method, the lever operation amount is determined by proportional (P) control based on the angle. Consequently, when external loads act on each joint during excavation, the operation halts when the load balances with the force generated by the lever operation amount output by the P controller. For

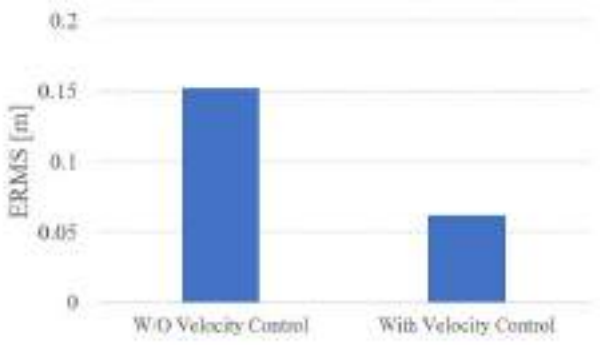


Figure 13. Error comparison: RMS cutting-edge error for conventional and proposed methods.

this reason, it is considered that the conventional method required about 30 seconds (equals to the control timeout) to reach the final target position in each operation mode.

5 Conclusion

To advance the automation of construction, we prototyped automated system for a hydraulic excavator with a velocity controller considering the characteristics of the hydraulic system. This velocity controller allows users to issue rotational velocity commands for each axis of the excavator's working device. Using this controller, the excavator performed ground excavation tasks, and the path control accuracy of the bucket cutting edge was compared between cases with and without the proposed controller.

The results confirmed that the path deviation of the cutting-edge was reduced to less than half when using the proposed velocity controller. This finding demonstrates that for automating hydraulic excavators, embedding controllers capable of realizing actuator angular velocities within the excavator and using angular velocity commands from the upper motion control system instead of lever-based commands is a a preferable approach. The proposed velocity control method is effective in achieving this goal.

Skilled human operators adeptly perform various tasks through lever operation. For example, during excavation, they adapt their operations based on the load conditions, distinguishing between hard and soft ground. This suggests that, even if control commands are given only as lever operation inputs, it may be possible to develop an autonomous system capable of executing tasks at a level comparable to humans, provided the upper-level control system is sufficiently sophisticated. However, implementing such an advanced upper-level system poses significant technical challenges for users. By contrast, if the angular velocity control utilizing hydraulic pressure information proposed in this study is pre-installed within the hydraulic excavator, the upper-level system can be simplified. In future work, we will not only aim to further enhance the robustness of angular velocity control but also explore optimal system configuration to improve user-friendliness in automation. This includes integrating a high-level controller within the excavator, allowing users to specify the desired path of the bucket's cutting edge.

References

[1] T. Bock and T. Linner. *Construction Robots*. Cambridge University Press, New York, NY, USA, 2016.

[2] T.C.S. Hsia. A new technique for robust control of servo systems. *IEEE Transactions on Industrial Electronics*, 36(1):1–7, 1989. doi:10.1109/41.20338.

[3] Chang PH. Lee, SJ. Modeling of a hydraulic excavator based on bond graph method and its parameter estimation. *Journal of Mechanical Science and Technology*, 26:195–204, 2012. doi:https://doi.org/10.1007/s12206-011-0938-2.

[4] Hitachi Construction Machinery. Ict construction solution. On-line: `https://www.hitachicm.com/global/en/solutions/solution-linkage/ict-construction/`, Accessed: 16/12/2024.

[5] Hitachi Construction Machinery. Hitachi review construction machinery products. On-line: `https://www.hitachihyoron.com/rev/archive/2021/r2021_03/16/index.html`, Accessed: 06/12/2024.

[6] G. Yamauchi, D. Endo, H. Suzuki, and T. Hashimoto. Proposal of an open platform for autonomous construction machinery development. In *Proceedings of ISARC 2023*, pages 186–191, Chennai, India, 2023.

[7] MoveIt Motion Planning Framework. Moveit motion planning framework. On-line: `https://moveit.ros.org`, Accessed: 06/12/2024.

Interoceptive Robotic Agents for Shared Control and Learning in Collaborative Construction Work

Xiaoshan Zhou[1], Carol C. Menassa[2], and Vineet R. Kamat[3]

[1]Ph.D. Candidate, Dept. of Civil and Environmental Engineering, Univ. of Michigan, Ann Arbor, MI 48109, USA.
[2]Professor, Dept. of Civil and Environmental Engineering, Univ. of Michigan, Ann Arbor, MI 48109, USA.
[3]Professor, Dept. of Civil and Environmental Engineering, Univ. of Michigan, Ann Arbor, MI 48109, USA.

xszhou@umich.edu, menassa@umich.edu vkamat@umich.edu

Abstract -

Advancing the adaptivity of autonomous mobile robots (AMRs), particularly empowering them to align with human expertise and learn from past experiences to infer optimal decisions in response to changing task demands and dynamic environment, is critical for their application in modular construction automation and reducing operational carbon footprints. Although a few learning paradigms based on human demonstrations have been introduced, robots remain rigid in structured tasks and lack internal adaptivity and authentic learning capabilities. This research explores a fundamental approach to equipping robots with an internal drive for self-reflection and conscious learning within a shared control setting, where human supervisors provide occasional corrective inputs and interventions. This approach is centered on the concept of interoception, which we factorize as "cognitive dissonance" within the robotic cognitive architecture. When significant decision divergence with human counterparts occurs, the robotic agent will actively seek decision rationale from semantic inputs, extracts knowledge from human scaffolding, encodes it into a memory system, and retrieves this "cached" knowledge for future decision-making in similar scenarios. This paper explores this new perspective to help build adaptive motion planning in AMRs. First, we propose integrating the legacy of heuristic costs from grid/graph-based path planning algorithms with a hypergraph model that caches declarative and procedural knowledge extracted from human semantic inputs. Second, we design a velocity-replay module that uses few-shot contrastive learning with a encoder-decoder architecture and compare its performance against reinforcement learning with human feedback and imitation learning. Two simulation demonstrations were conducted using the Issac Sim platform to showcase multi-robot synchronization and stacking collaboration. In addition, the proposed shared control paradigm was tested in a Gazebo environment in ROS, where human inputs were provided via relatively noisy brainwave-based control, compared to conventional peripheral devices. The results demonstrate how this approach better mitigates human input errors while facilitating autonomy for task completion compared to existing dominant shared control methods. The insights from this study resonate with recent advances in neuroscience and reinforcement learning, and pave the way toward artificial general intelligence in AMRs, fostering their progression from complexity to competence in construction automation.

Keywords -

Autonomous Mobile Robot; Shared Control; Motion Planning; Learning; Memory

1 Introduction

Automation in construction can significantly contribute to achieving sustainability goals [1]. In particular, modular construction using offsite prefabrication and onsite assembly have the potential to address climate action by reducing operational carbon footprints compared to traditional onsite methods such as concrete mixing and curing [2]. In additional to potential environmental benefits, automation can improve worker safety by mitigating exposure to hazardous conditions. For example, automated robotic machinery used for onsite tasks, such as assembly, reduces risks like falls from scaffolding, inhalation of welding fumes, and electrocution [3]. Mobile robots, including forklifts, cranes, and dozers, have been increasingly used to streamline offsite and onsite construction processes [4]. However, the integration of mobile robotics in construction introduces new challenges. Without effective coordination of multi-robot path planning that manages sequential task dependencies and precise velocity control to adapt to dynamic nature of construction workflows, risks such as struck-by hazards increase [5].

In addition, autonomous mobile robots (AMRs) reshape the power dynamics between humans and machines. These robots can operate with programmed autonomy, performing tasks such as delivering materials across a precast factory or a site to support fabrication or assembly processes [3]. Although they remain under the supervision of human workers, their autonomy shifts the relationship from subservience—where robots follow human instructions without intrinsic understanding—to augmentation, where robots act more like symbiotic co-workers with spe-

cialized capabilities. This shift, however, can provoke apprehension among industry practitioners, often fueled by fear of the unknown [6]. Concerns about the emergence of superintelligent systems, often referred to as singularity, suggest a future where robots could become uncontrollable and irreversible in their autonomy. Such fears are compounded by the opaque reasoning processes of robots and the asymmetry in knowledge between humans and machines, leading to uncertainty about whether these systems are truly beneficial and trustworthy.

To address these challenges, this research advocates for advancing the cognitive capabilities of AMRs to align with human intelligence. A critical focus is to learn from humans and learn the way how humans learn. In fact, many principles of robotics have been inspired by how humans perceive, reason, and act. For instance, robotic sensing mimics human sensory organs: cameras (RGB, depth, or infrared) parallel human vision, force sensors replicate touch, microphones simulate hearing, and joint encoders mirror proprioception. State-of-the-art robots have developed highly dexterous manipulation skills comparable to those of humans. However, a key distinction between humans and robots lies in adaptability, self-reflection, and learning. Human workers can consciously think, learn from past experiences, and modify their behaviors to adapt to changes. In contrast, robots are mechanical and purpose-built, lacking genuine situational awareness and the ability to abstract knowledge, let alone reuse it. More fundamentally, robotics research has primarily focused on replicating exteroception (sensing the external environment), while overlooking interoception (perceiving internal states). Interoception, however, plays a criticla role in human cognition, facilitating individuals to maintain homistasis, extrapolate mental schemas from observations, and make decisions based on hierarchical prioritization [7]. Inspired by this, a logical next step in robotics reseach should explore equipping mobile robots with interoceptive-like capabilities.

Interoception in mobile robots is a novel concept, and its conceptualization and implementation require a formal definition of the phenotypic characteristics—particularly the internal states—of mobile robots. Unlike human interoception, we are not suggesting that robots experience genuine emotions, pain, or self-consciousness. Instead, we define ”cognitive dissonance” as an abstract internal cognitive state variable robotic agents possess and factorize it based on the intensity of human inputs to refine robotic autonomy in shared control paradigms. This is proposed along with a new shared control architecture as illustrated in Figure 1: Rather than allowing human interventions to override robotic decisions, our proposed shared control integrates human inputs as a tributary that converges with autonomous decisions to produce a collaborative decision-making output forwarded to the robot's actuators. Simultaneously, this architecture inherently introduces a potential state of conflict between the robot's internal beliefs and its executed actions. Such conflicts create an internal drive for the robotic agent to reflect on its decisions, seeking insights from human feedback to identify potential biases or missed factors causing misalignment with human expectations. Through this reflective process, the robot engages in conscious-like learning and can gradually evolve into a meta-level understanding that reconciles discrepancies with its human counterparts.

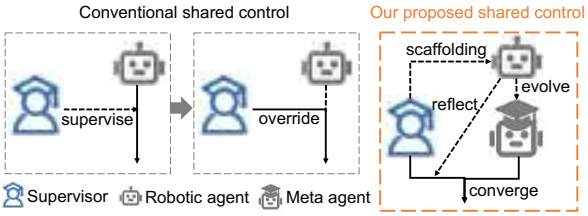


Figure 1. Shared control architectures

This framework is particularly designed for spatio-temporal-sensitive motion planning situations in multi-robot systems that navigate shared spaces and must dynamically adapt to evolving task sequences. For example, robots delivering panels must complete their task before façade installation can proceed. Precise velocity control is requrid thoughtout the process to avoid collisions and minimize downtime. To achieve this, the knowledge required by the robotic agents to make efficient motion planning decisions can be categorized into declarative knowledge and procedural knowledge. Declarative knowledge includes those facts (e.g., terrain features that were previously unavailable from the spatial grid map, traffic regulations with safety margins around obstacles, and path preferences provided by human supervisors) that inform future path-planning decisions. In our approach, this knowledge is acquired through human feedback that is prompted by an intuitive Graphical User Interface (GUI). The GUI facilitates the formalization of human knowledge while providing contextual dependency. This feedback is then encoded using natural language processing (NLP) techniques, including tokenization, tagging, and logical reasoning, and finally constructed as a knowledge graph (scaffolding). This design aligns with Lev Vygotsky's theory of the Zone of Proximal Development, which highlights bridging the gap gradually between what the learner (robotic agent) can do independently and what it can do with guidance from a more knowledgeable other [8].

Procedural knowledge pertains to how the mobile robot accurately replicates velocity profiles. Unlike dominant methods such as reinforcement learning (RL) and imitation learning—which start from a “blank slate”, rely on

random exploration in large state spaces, or require explicitly designed rewards (a challenging task)—our approach formalizes learning as a structured process of schema representation and modification. This process is regulated by goal-oriented behavior, avoiding the habituation tendencies of RL that often fail to adapt quickly to reward contingencies and thus sacrifice flexibility [9]. Our approach conceptualizes learning as a sequential production of components comprising a human-corrected velocity repertoire, implemented using an encoder-decoder architecture based on few-shot contrastive learning and Gated Recurrent Unit (GRU) to capture temporal dependencies. These designs lay the groundwork for factorizing the ontology of mobile robot motion planning and appeal to diverse optimization principles for routing decisions. An illustration of the robotic interoception concept, along with its relationship to modern topics in RL, is shown in Figure 2.

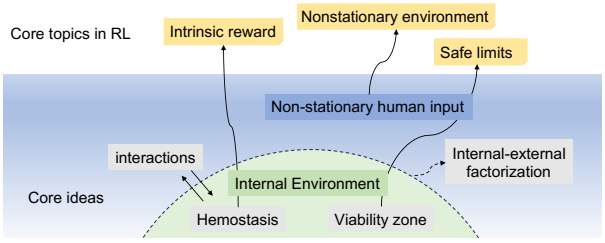


Figure 2. Overview of the interocepltive framework

In practice, the workflow of the proposed approach is illustrated in Figure 3. While acknowledging the multifaceted nature of interoceptive robotics, this paper narrows its focus to a detailed methodology for motion planning in Section 2, followed by three illustrative examples in Section 3.

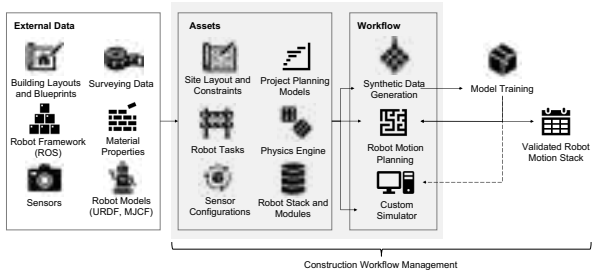


Figure 3. Simulation-enabled workflow for on-site robot motion planning

2 Methodology and Experimentation

The overarching pipeline of the designed methodology is shown in Figure 4. The motion planning consists of two components: path planning and velocity planning.

2.1 Path Planning

The path planning is defined to find a continuous path from a start location to a goal location throughout the subset of spatial configuration space consisting of traversable space. The space is created as a discretized approximation of connectivity using an occupancy grid, which can also be seen as a graph in which vertices are individual configurations and edges represent the transitions between pairs of configurations. Paths are found by searching the space using A* [10] (see Figure 5 for an abstracted formulation of the physical spatial environment and paths).

Given the designated paths, the tracking of the AMRs' positions is available (see Figure 5) with a hypothesized constant velocity. In practice, however, the human supervisors occasionally adjust the velocity to accommodate ground or floor conditions, task hierarchies, and interaction dynamics with other entities. This research proposes a shared control paradigm based on a Bayesian updating framework, leverageing robotic autonomy (scaled in absolute terms) as the prior and velocities from human agency (scaled as increments) as the likelihood. Using the integral of velocity curves (often approximated with sample-based approaches, e.g., Monte Carlo) at a pre-defined sampling rate along the designated path, the distances of the AMRs can be tracked in real time, with proximity thresholds set to trigger automated alerts.

In the proposed shared control paradigm, the factorized cognitive dissonance of AMRs' reasoning and action—as parameterized by the intensity of human inputs—is demonstrated using a contour plot (see Figure 6). This provides an intuitive representation for retrospective documentation of rationale behind human interventions. The texts inputs from human supervisors are processed with NLP open information extraction (OpenIE) focusing particularly on two attributes—terrain features and task sequences—as declarative knowledge mapped to the space vertices. Additional attributes are preserved as episodic knowledge, which is however beyond the scope of this article. The source code for this simulation is available at `https://github.com/XiaoshanZhou624/A-star-with-GUI-for-human-inputs`.

Task sequence attributes partially overlap with positional information from the spatial space but require an additional task space to define the availability state of the objects at the target workspace where robots offload or load and the occupancy state of critical pathways shared by robots approaching the workspace. We propose to use hypergraphs to model this new configuration space that requires additional complexity. A hypergraph is an extension of a graph where hyperedges (or hyperarcs) are not restricted to connecting only two vertices [11], thus allowing us to accommodate more dependence conditions. A graph-based model also facilitates to modify the heuristic

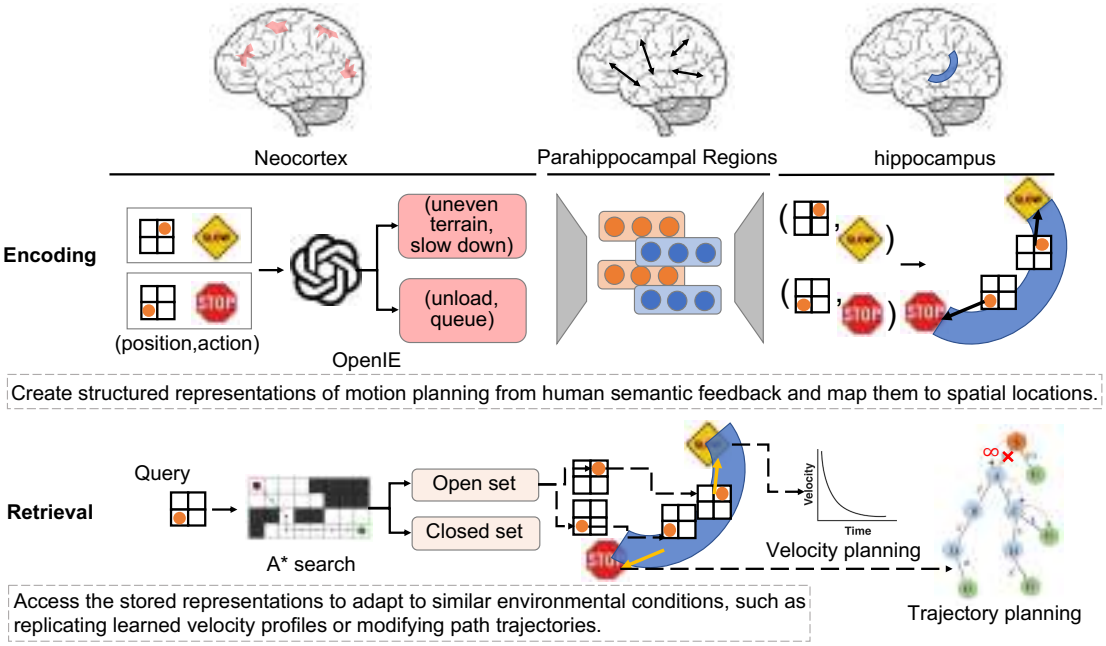


Figure 4. Methodology overview for knowledge encoding and retrieval in path and velocity planning

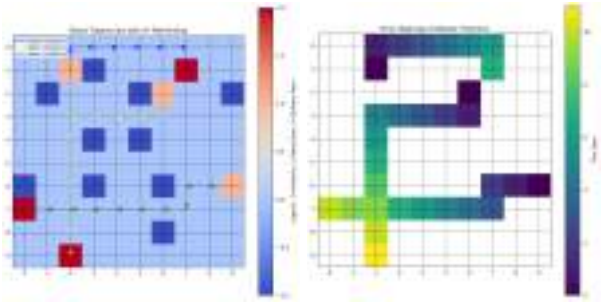


Figure 5. (left) Trajectories with A* pathfinding in a grid. (right) Heatmap of AMRs position tracking

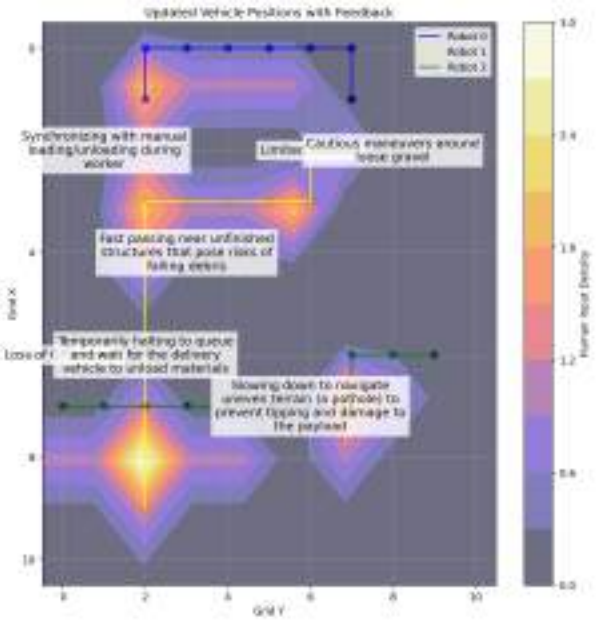


Figure 6. Contour plot of human intervention intensity and GUI for human semantic inputs in rationale

costs in A*, thus automatically influencing path planning in the algorithmic logic that determines which nodes are first expanded as leaf nodes and added to the closed set. The task space states define the preconditions for admissible moves to the explored nodes. A visual depiction of the hypergraph is shown in Figure 7, and the pseudocode for the modified A* algorithm is provided in Figure 8.

2.2 Velocity Planning

As the spatial configuration expands in the hypergraph model, terrain attributes can be coded to trigger the replay of human-refined velocity profiles. We propose an encoder-decoder network architecture, as shown in Figure 9. Given only one trial of data, we designed a sliding time window to segment the velocity profiles into smaller segments. Leveraging the similarity contrast principles of few-shot learning, we construct positive pairs from consecutive windows and negative pairs from random windows. A Gaussian filter is applied to smooth the data within each window. Subsequently, these pairs are processed through

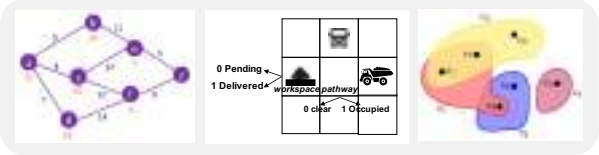


Figure 7. Hypergraph attribute visualization

Algorithm 1 Hypergraph A* with State-Dependent Costs

Input:

$H = (V, E)$: A hypergraph where:
- V: Set of nodes $v = (p, s)$, with p as position and s as binary state.
- E: Set of hyperedges $e \subseteq V$, where the cost of each node $v \in e$ depends on the s of other nodes in e, represented by $c(v, e, \text{states})$.

start: Starting node.
goal: Goal node.
Output: The optimal path from *start* to *goal*.
Procedure:

Initialize $open_set \leftarrow$ PriorityQueue() ▷ Stores nodes to be explored, ordered by f_score
$g_score[v], f_score[v] \leftarrow \infty$ for all $v \in V$ ▷ Initialize cost scores
$g_score[start] \leftarrow 0$ ▷ Cost to reach the start node is zero
$f_score[start] \leftarrow$ `heuristic_cost`$(start, goal, H)$ ▷ Estimate total cost from start to goal
$open_set.push(start, f_score[start])$ ▷ Add start node to the open set
while $open_set \neq \emptyset$ **do** ▷ Main loop: Process nodes until the open set is empty
  $current \leftarrow open_set.pop()$ ▷ Retrieve node with lowest f_score
  if $current == goal$ **then**
    return `reconstruct_path`$(current)$ ▷ Return the reconstructed optimal path
    Break ▷ Terminate the loop as the goal is reached
  end if
  for $neighbor \in neighbors(current)$ **do** ▷ Explore each neighbor of the current node
    if $neighbor \notin e$ for all $e \in E$ **then** ▷ Check if neighbor is part of any hyperedge
      $h_cost \leftarrow$ `heuristic_cost`$(neighbor, goal, \emptyset)$ ▷ No hyperedge states to consider
    else
      $connected_states \leftarrow \{n.s \mid n \in e.connected_nodes\}$ ▷ Retrieve nodes' states in hyperedge
      $h_cost \leftarrow$ `heuristic_cost`$(neighbor, goal, connected_states, e)$ ▷ Update heuristic
    end if
    $tentative_g \leftarrow g_score[current] +$ `edge_cost`$(current, neighbor)$
    $tentative_f \leftarrow tentative_g + h_cost$ ▷ Compute total estimated cost
    if $tentative_g < g_score[neighbor]$ **then** ▷ Update path if better cost is found
      $g_score[neighbor] \leftarrow tentative_g$
      $f_score[neighbor] \leftarrow tentative_f$
      $open_set.push(neighbor, f_score[neighbor])$ ▷ Re-add neighbor to the open set
    end if
  end for
end while

Figure 8. Pseudocode for hypergraph A* with state-dependent costs

a Siamese Network [12] with GRU modules to extract embeddings for each window, using the Euclidean distance as the contrastive loss for network training.

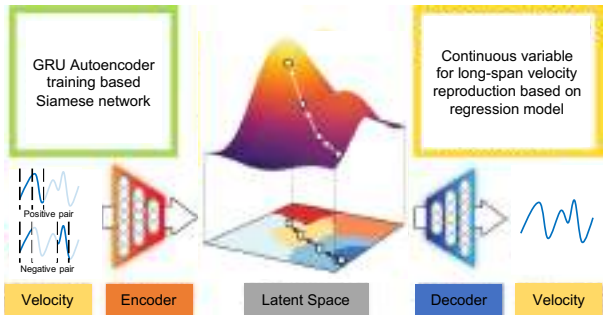


Figure 9. Technical overview of the encoder-decoder architecture for velocity replay

During retrieval, a regression model is designed and trained using Mean Squared Error as the loss function to map the embeddings back to velocity values (see Figure 13 for a demonstration). These values can then be directly broadcasted and subscribed to by robotic servo motor nodes in the Robot Operating System (ROS) environment. Detailed algorithm designs are available in the open-source code repository at `https://github.com/XiaoshanZhou624/velocity-replay`.

3 Demonstration in Simulation

3.1 Multi-Robot Coordination for Pick-and-Place

This section provides two examples of multi-robot coordination in motion planning and adaption to workpiece properties. The scenes and simulations were set up in Nvidia Omniverse Issac Sim (version 4.1.0).

As shown in Figure 10, two Jetbots navigate the workspace using a differential drive controller with physically accurate articulated joints. The task sequence dependencies are: after the first Jetbot delivers material to the workspace, it must reverse to clear the pathway. Simultaneously, the second Jetbot must wait and avoid entering the workspace prematurely to prevent blocking the pathway. A cube was added to the scene as an abstracted form of the material. To ensure realistic interactions, physical properties such as force and rigidity were applied to enable the first Jetbot to push the material toward the Franka Panda manipulator for handover. A Franka Panda manipulator was positioned in the workspace to pick up the object and rotate to hand it over to the second Jetbot. The hypergraph vertices' preconditions were incorporated into the program logic by parameterizing tasks or programming states to transition between subtasks, such as the Franka's handover or allowing the Jetbot to resume movement instead of being blocked by non-steppable areas influenced by A* heuristic costs. A video of this simulation is available at `https://youtu.be/RwKJ0zggWfE`. In real-world applications, precise localization is often achieved using land-marked AprilTags for accurate positioning of robots and manipulators. As demostrated in the video, when the goal location of the transport robot overlaps with the handover location of the manipulator's end-effector, collisions can occur. This highlights the need for research on precise spatial alignment, timing calculations, and accurate velocity control to synchronize robotic motion. These considerations are particularly crucial in large automated machinery, such as forklifts, or when dropping large-scale material panels that require significant space for maneuvering, reorientation, or flipping before placement.

Although block stacking tasks have been widely studied in robotic manipulation and have demonstrate high dexterity, real-world placement present greater complexity. Unlike uniform blocks with consistent shapes, sizes, and weights that can be easily aligned in horizontal or vertical stacks, construction materials exhibit considerable variability. For instance, pieces of lumber often differ in di-

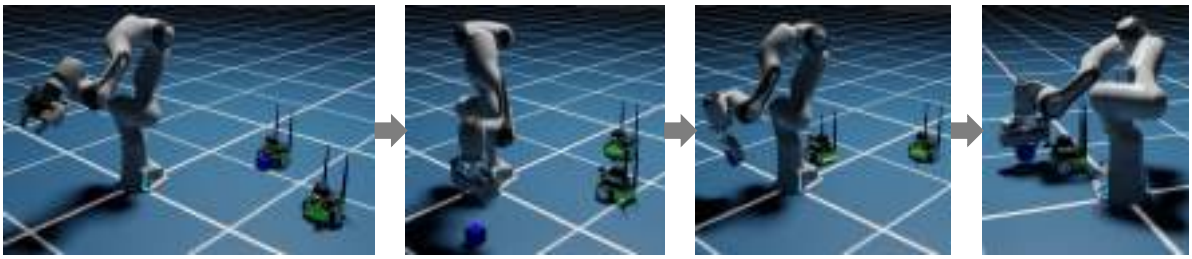


Figure 10. Robot motion planning for material delivery and handover coordination

mensions, while palletized materials may contain different items, such as tiles and shingles, each varying in weight. Moreover, some materials may need repositioning during unloading to facilitate subsequent tasks. For example, glass panels may require flipping to expose the correct side for façade installation. Proper planning must account for balanced weight distribution during transport to mitigate the risk of tipping caused by uneven loads affecting inertia and centrifugal forces, particularly for wheeled robots. An example is provided in Figure 11. This demonstration was adapted from a UR 10 Bin Stacking application containing a conveyor belt, a pallet where the bins should be stacked, and a UR 10 robot with a suction gripper. Bins arriving right-side up were flipped at a designated station to ensure proper stacking upside down. The placement alternated between sides of the pallet to balance weights. A supplementary video demonstrating the robot's motion planning is available at `https://youtu.be/yHbuoG7d05A`. Future work will integrate sensors for weight tracking and incorporate this data into the the manipulator's motion planning to maintain load balance while minimizing travel distance. In addition, sim-to-real discrepancies will be analyzed by examining how stiffness and damping parameters in the simulation setup correlate with the accuracy of joint properties in URDF robotic models and real-world variations in loads and terrain friction.



Figure 11. UR10 bin stacking and motion planning with weight-balanced placement

3.2 Velocity Replay

The velocity embeddings obtained after training are visualized in Figure 12, with dimensionality reduced using t-Distributed Stochastic Neighbor Embedding (t-SNE). These embeddings are stored as compressed vectors of the original velocity profiles. During retrieval, they are mapped back to velocity, as shown in Figure 13, where regression models reconstruct multiple possible original profiles.

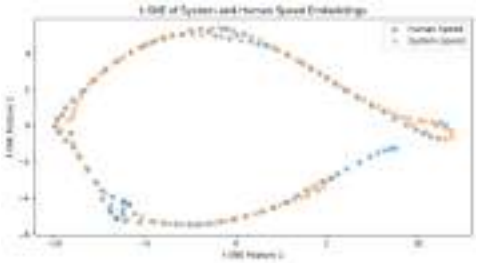


Figure 12. t-SNE visualization of velocity embeddings after contrastive Siamese Network learning

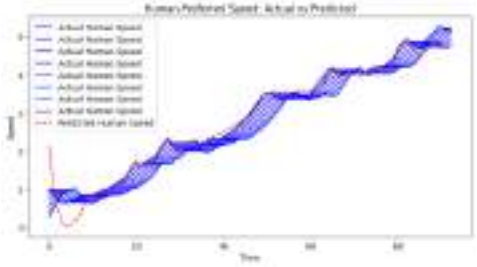


Figure 13. Illustration of mapping embeddings back to velocity using the regression model

The core objective here is to contrast our proposed method with currently dominant approaches—Reinforcement Learning with Human Feedback (RLHF) and Imitation Learning (IL). The learning mechanisms of the three methods and their performance comparisons are shown in Figure 14. Mostly importantly, IL functions similarly to behavioral cloning, allowing for a near-perfect replay of velocity profiles with high learning efficiency. However, as a supervised learning method, IL does not involve reasoning or prediction beyond what it has seen during

end-to-end training. Once the environment changes (e.g., unexpected obstacles, disruptions, or external forces), IL cannot adapt. RLHF uses human input to shape the reward function, guiding the robotic agent's actions. With extensive training episodes, the agent is likely to develop a well-represented internal world model that informs its decisions. However, the training process can be highly time-consuming. Given the semi-structured nature of construction sites, robots are expected to follow designated paths while remaining aware of unexpected proximity intrusions. In our approach, a dedicated module for velocity embeddings enables compact storage and efficient replay. This approach offers greater flexibility and better aligns with the ontology of human learning.

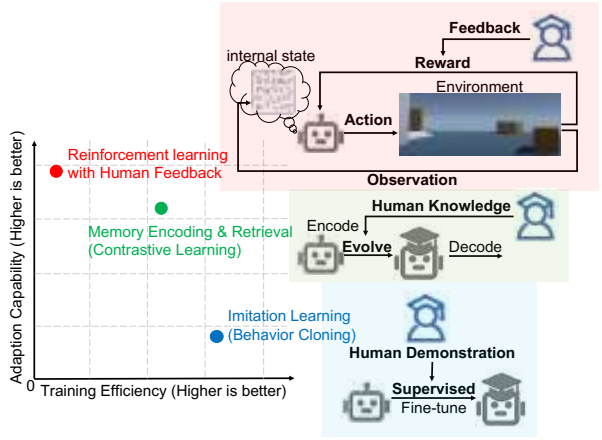


Figure 14. Comparison of RLHF, IL, and our contrastive learning approach using an encoder-decoder model

3.3 Shared Control Comparisons with EEG-Based Brainwave Input

We present a case study that simulates the possibility of incorrect human inputs in shared control. These inaccuracies can arise from two factors: (1) humans making incorrect decisions and (2) errors in detecting their intentions, such as missed detections or false alarms. In this simulation, we introduce potential detection errors by interpreting human inputs from brain signals measured via scalp electroencephalography (EEG). We trained a Bayesian inference framework to interpret EEG signals related to velocity adjustment intentions (acceleration or deceleration). Details on the EEG signal methodology can be found in the authors' other publications. Although the methodological framework accounts for uncertainty by using probabilistic reasoning, inherent noise in the signal spectrum and human inability to maintain consistent control can lead to fluctuations and inaccuracies in detected inputs.



Figure 15. Simulation platforms: (left) Gazebo in third-person perspective; (right) Unity in first-person perspective

To evaluate our approach, we tested in two scenarios (as shown in Figure 15). First, we set up a TurtleBot in Gazebo to follow another TurtleBot moving at a predefined velocity profile. Human users can intervene the following Turtlebot through three shared control paradigms: overriding, weighted pooling, and the proposed Bayesian updating method. EEG signal was streamed via WebSocket from Emotiv, processed into velocity commands in real time, and broadcast using the `cmd_vel` topic in ROS, which was subscribed to by the following TurtleBot. We measured how shared control can reconcile fluctuations in human inputs while achieving a balanced integration with robotic autonomy. As evidenced in Figure 16, our Bayesian updating method can smooth fluctuation in human inputs, always preserve the dual-agent decision-making rationale, and allows the robotic agent to follow preferred velocity patterns without requiring continuous human intervention.

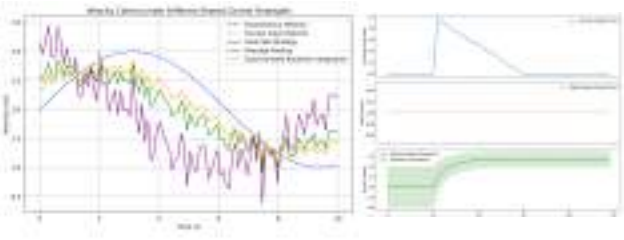


Figure 16. Shared control performance comparisons

Second, we tested how well a robot in Unity resumes motion after stopping due to autonomous obstacle avoidance detected by LiDAR monitoring. Although collisions were prevented, the key question was whether and how the robot could navigate past a narrow passage threshold and continue moving without getting stuck. In Unity, autonomous driving was enabled using the NavMeshAgent component, and the EEG data stream was subscribed to via the TCP protocol. Task completion times were measured across multiple trials for the three shared control modes and are presented in Table 1. In overriding mode, the human required significantly more time to complete the maneuver. In weighted (average) pooling mode, the hu-

man had to exert strong mental effort to insert control and influence the robot's motion. However, in our proposed Bayesian updating approach, a small human input acted as a minor disturbance but could help the robot escape a local deadlock. After a brief intervention, the human can hand over control to the robot to resume autonomous mode and complete the task efficiently.

Table 1. Task Completion Time

Shared Control Mode	Task Completion Time (s)
Overriding	> 30
Weighted Pooling	10–25
Bayesian Updating	< 10

4 Conclusion

Mobile robots are envisioned as learning agents within a shared control paradigm, where we define an interoceptive cognitive architecture that enables them to develop an internal drive for self-reflection based on feedback from collaborative decision-making human counterparts. The knowledge acquired through this shared control process is abstracted, structured, and encoded, allowing for more flexible retrieval and faster adaptation in future similar situations compared to conventional RLHF and IL frameworks. The proposed shared control framework also helps counterbalance potential errors in human inputs by leveraging dual-agent intelligence and empowers robotic autonomy to automatically follow human-preferred velocity patterns compared to coventional overriding and weighted pooling shared control. This proposed approach is expected to advance robotized construction automation by shfiting from rigid, pre-programmed behaviors to greater autonomy and adaptive intelligence, enabling robots to coordinate better with human coworkers and multi-robot systems. Furthermore, this research creates opportunities to handle more heterogeneous and complex tasks within the modular construction vision, thus contributing to safer and more sustainable, decarbonized construction practices.

5 Acknowledgement

The authors would like to acknowledge the financial support received for this research from the US National Science Foundation (NSF) Grant SCC-IRG 2124857. Any opinions in this paper are those of the authors and do not necessarily represent those of the NSF.

References

[1] Z. Zheng, F. Wang, G. Gong, H. Yang, and D. Han. Intelligent technologies for construction machinery using data-driven methods. *Automation in Construction*, 147:104711, 2023.

[2] S. K. Yevu, E. K. Owusu, A. P. C. Chan, S. M. E. Sepasgozar, and V. R. Kamat. Digital twin-enabled prefabrication supply chain for smart construction and carbon emissions evaluation in building projects. *Journal of Building Engineering*, 78:107598, 2023.

[3] C. Liu, J. Wu, X. Jiang, Y. Gu, L. Xie, and Z. Huang. Automatic assembly of prefabricated components based on vision-guided robot. *Automation in Construction*, 162:105385, 2024.

[4] K. You, L. Ding, C. Zhou, Q. Dou, X. Wang, and B. Hu. 5g-based earthwork monitoring system for an unmanned bulldozer. *Automation in Construction*, 131:103891, 2021.

[5] D. Kim, S. Lee, and V. R. Kamat. Proximity prediction of mobile objects to prevent contact-driven accidents in co-robotic construction. *Journal of Computing in Civil Engineering*, 34(4):04020022, 2020.

[6] C.-J. Liang, T.-H. Le, Y. Ham, B. R. K. Mantha, M. H. Cheng, and J. J. Lin. Ethics of artificial intelligence and robotics in the architecture, engineering, and construction industry. *Automation in Construction*, 162:105369, 2024.

[7] Giovanni Pezzulo, Francesco Rigoli, and Karl Friston. Active inference, homeostatic regulation and adaptive behavioural control. *Progress in Neurobiology*, 134:17–35, 2015. doi:10.1016/j.pneurobio.2015.09.001.

[8] B. Eun. The zone of proximal development as an overarching concept: A framework for synthesizing vygotsky's theories. *Educational Philosophy and Theory*, 51(1):18–30, 2019.

[9] K. Khetarpal, M. Riemer, I. Rish, and D. Precup. Towards continual reinforcement learning: A review and perspectives. *Journal of Artificial Intelligence Research*, 75:1401–1476, 2022.

[10] P. E. Hart, N. J. Nilsson, and B. Raphael. A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions on Systems Science and Cybernetics*, 4(2):100–107, 1968.

[11] C. Berge. *Hypergraphs: Combinatorics of Finite Sets*. Elsevier Science, 1984.

[12] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah. Signature verification using a "siamese" time delay neural network. In *Proceedings of the 7th International Conference on Neural Information Processing Systems*, pages 737–744, Denver, Colorado, 1993.

Evaluation of an Autonomous Driving Dumper using a Robot Evaluation Framework

Tim Bernhard[1], Yuan-Jen Huang[1], Anne Fischer[1], Cynthia Brosque[2] and Johannes Fottner[1]

[1]Chair of Materials Handling, Material Flow, Logistics, School of Engineering and Design, Technical University of Munich, Garching, Germany
[2]Department of Civil and Environmental Engineering, Stanford University, Stanford, USA
tim.bernhard@tum.de, yuan-jen.huang@tum.de, anne.fischer@tum.de, cbrosque@stanford.edu, johannes.fottner@tum.de

Abstract

This paper evaluates the applicability of an autonomous driving dumper for a construction project. Conducted as part of a joint lecture between Stanford University and the Technical University of Munich in 2024, this study provides students with hands-on experience in field assessments of construction robotics. The dumper, developed in partnership between the Technical University of Munich and the industry partner Wacker Neuson, is analyzed using the Robotics Evaluation Framework. Autonomous dumpers present a promising solution in the construction industry, offering improved safety, productivity, and the ability to address labor shortages, particularly on small and confined sites. The Robotics Evaluation Framework used is a systematic methodology for assessing feasibility and performance across safety, quality, schedule, and cost metrics. The results demonstrate that the autonomous dumper improves safety through better ergonomics, enables operators to focus on higher-value tasks, and enhances overall efficiency. While quality remains comparable to conventional dumpers, the autonomous system experiences 20% longer cycle times, presenting a scheduling challenge. Despite this, cost analysis reveals amortization after 1.43 years or 75 projects, highlighting its economic viability and potential for fleet scalability.

Keywords – Construction robotics; Autonomy; Mobile Robotics; Robotics Evaluation Framework (REF)

1 Introduction

The construction industry faces a transformative era of unprecedented demand and significant challenges [1, 2]. With global infrastructure needs surging and urbanization accelerating, the demand for construction work continues to grow. However, a critical labor shortage, driven by a shrinking pool of skilled workers, directly impacts productivity, making it increasingly difficult for the industry to keep pace with rising workloads [3, 4]. This shortage contributes to inefficiencies, as fewer trained professionals can complete projects on time and within budget. As a result, the construction sector continues to lag behind other industries in productivity, highlighting the urgent need for innovative solutions to bridge this gap. Construction projects overrun average cost and time relative to the original budget and schedule at 70 percent and 61 percent, respectively [4]. These delays due to labor shortages and inefficient workflows underscore the need for automation to bridge productivity gaps.

In this context, the market for autonomous construction equipment is emerging as a pivotal force. This market is projected to expand from USD 4.40 billion in 2024 to USD 9.77 billion by 2030 at a compound annual growth rate of 14.2% [5]. Despite its potential, the adoption of automation in construction faces significant hurdles. In Germany, challenges include a rigid labor training structure that mandates three years for operator qualification. A shortage of skilled operators further exacerbates the issue. Procurement practices prioritizing the cheapest bid over innovation pose further challenges. [6]. There is also limited familiarity with robotics or advanced automation, alongside resistance to autonomously driven equipment during operator training. A lack of widespread experience with autonomous systems in construction workflows compounds these issues.

This paper employs the Robotics Evaluation Framework (REF) from Brosque and Fischer [7] to systematically evaluate an autonomous driving dumper, developed by Zöls et al. [8] from the Chair of Materials Handling, Material Flow, Logistics at the Technical University of Munich (TUM) and the dumper manufacturer Wacker Neuson. The REF is a structured methodology designed to evaluate construction robots based on product, organization, process, safety, quality, schedule, and cost metrics.

This paper is structured as follows. Section 2 provides insights into the current development of automated construction equipment. The REF is introduced in section 3, and section 4 applies the REF to the autonomous dumper compared to conventional dumpers to evaluate similar processes and sizes to accurately assess their capabilities. This use case relies on direct testing under lab conditions on the outdoor facilities at TUM. Section 5 presents challenges and opportunities. Finally, section 6 is dedicated to the conclusion and outlook.

2 Literature review

This section explores the landscape of construction robotics, outlining the field's advantages, challenges, and current state of development. Construction robots offer significant potential to revolutionize the construction industry by enhancing efficiency, precision, and safety. These robotic systems, equipped with advanced sensors, actuators, and cognitive algorithms, outperform traditional methods by executing tasks with greater accuracy, yielding higher-quality results with fewer errors [9]. Their ability to undertake repetitive, dangerous, or physically demanding jobs significantly reduces site accidents [10]. By assuming such tasks, robots enable human workers to focus on higher-value activities, fostering safety and efficiency.

While construction robotics presents numerous benefits, the field is still evolving and tackling considerable challenges. Current research primarily emphasizes technological advancements to boost productivity and automate specific construction processes. However, a comprehensive evaluation framework for comparing these technologies, especially against traditional practices, remains absent, highlighting a critical research gap.

Despite the promising advantages, deploying robotics in the construction industry faces challenges due to the dynamic and unstructured nature of construction sites [9–11]. These challenges can be broadly categorized:

- **Safety Challenges:** Ensuring safe robotic operations on crowded construction sites is a critical concern [9]. Multiple activities – often involving workers, materials, and equipment – occur simultaneously, making task execution more complex [10].
- **Process Adaptability Challenges:** Construction processes lack standardization and frequent, unpredictable changes [9, 11]. This dynamic environment demands robotic systems capable of adapting to diverse circumstances and adjusting their actions based on real-time conditions [9, 11].
- **Environmental Challenges:** Uncontrollable weather conditions, such as rain, strong winds, and dust, significantly impact the consistent performance and reliability of robots on construction sites [10].

These factors can degrade sensor accuracy, limit mobility, and impact the overall effectiveness. Achieving full autonomy for construction robots is particularly challenging due to the obstacles posed by unstructured environments and the need for real-time multi-system coordination [11]. Nonetheless, several robots have been successfully deployed to perform specific repetitive tasks that are well-suited for automation.

Construction projects are typically divided into three main tasks: superstructure, substructure, and site preparation [11]. The subsequent sections will explore robotic applications within these categories and related areas.

Table 1 offers an overview of current robotic applications to provide context for the subsequent discussion. These are listed according to their relevance to an autonomous driving dumper. Relevance is judged by process, specific tasks, their on-/off-site usage and whether they are mobile or stationary. The overview also indicates the robot's automation level, based on degrees of freedom and operating conditions. In addition, it shows whether safety, quality, schedule, and cost aspects have been evaluated – aspects detailed further in Section 3.

Table 1. Evaluating relevant construction robotics

Reference	Tasks of robots	Process relevance to dumper	On-site/off-site	Mobile/stationary	Deg. of autonomy	Safety	Quality	Schedule	Cost
[12]	Prefabrication	○	Off	S	◐			X	
[13]	Transport in prefabrication	○	Off	M	◕	X		X	X
[14]	Structural assembly	◔	On	S	◐				
[15]	3D concrete printing	◔	On	S	◐		X		X
[16]	Beam assembly	◔	On	M	◔			X	
[17]	Ceiling board construction	◔	On	M	●	X		X	
[18]	Pile driving	◐	On	M	●		X		
[19]	Waste sorting	◐	Both	M	●		X		
[20]	Excavation	◕	On	M	●		X		
[21]	Material transportation	◕	On	M	●	X	X		
[22]	Material loading	◕	On	M	◐		X	X	
[23]	Material loading and transportation	●	On	M	◐		X		
[8]	Bulk material handling	●	On	M	◐			X	

Drawing from the successful implementation of industrial robot arms in manufacturing, off-site construction presents a viable area for robotics adoption. To address the challenge of non-standardized building components in construction, multi-functional robotic cells are being developed to automate labor-intensive and

time-consuming prefabrication processes [12]. Automated Guided Vehicles (AGVs), another mature technology widely utilized in logistics, are also highly relevant. AGVs offer high levels of automation, flexible mobility, and efficient space utilization. This makes them valuable for material handling in stationary prefabrication and module production systems [13].

Transitioning to on-site applications, cable-driven robots have emerged as a versatile solution. Their lightweight design facilitates easy transport and deployment on construction sites [14, 15]. Operated using elevated pulleys, these systems can support full-scale architectural projects. They also adapt to irregular site geometries. For instance, cable-driven robots have been validated for accurately placing granular materials in complex architectural-scale enclosures [14]. Moreover, a cable-driven parallel robot with a concrete extruder demonstrates precise 3D printing capabilities for complex concrete structures. This approach reduces material waste and labor costs while accommodating diverse construction environments [15].

Beyond cable-driven systems, specific robots are designed for repetitive tasks and productivity enhancement on-site. Robotic beam assembly systems automate bolting processes, allowing operators to control robotic transport mechanisms and bolting devices [16, 24]. Ceiling construction robots automate the positioning and installation of ceiling boards [17]. This minimizes the need for human work at elevated positions, improving safety.

In substructure processes, autonomous robots are designed for tasks such as transporting and driving interlocking sheet piles into the soil. This robot ensures precise positioning and interlocking by utilizing a vibratory hammer and its weight, coupled with alignment mechanisms. This enables the construction of continuous stabilization walls [18].

Robotics is also being applied to broader site management tasks. For example, a robotic system equipped with computer vision can autonomously patrol construction sites. It identifies and sorts waste materials and employs optimized gripping strategies for handling diverse waste types and forms [19].

Furthermore, mobile robots are increasingly being developed for earthmoving tasks in site preparation. The hydraulic excavator for an autonomous purpose (HEAP), a walking excavator, demonstrates precise localization, movement tracking, and accurate end-effector control. Its unique wheeled-legged design allows adaptability to uneven terrains, validated in applications like dry stone wall assembly and forestry work [20]. A retrofitted dump truck also showcases autonomous capabilities. It manages steering, acceleration, and braking. Path tracking performance has been evaluated, and an emergency stop feature is integrated to ensure safety [21].

In line with advancements in Artificial Intelligence (AI), autonomous navigation for wheel loaders in short-loading cycles is being explored using reinforcement learning. By decomposing loading tasks, AI agents manage to reverse, reach reversal points, navigate to dump trucks, and control steering, velocity, and bucket positioning [23]. Automated shovel trajectories have been developed and evaluated against manual operation for shoveling. Metrics include full bucket rate, operation time, and fuel consumption [22].

Finally, the autonomous driving dumper, as presented by Zöls et al. [8] with obstacle detection capabilities, exemplifies the potential of collaborative robots in construction environments.

While current robotic applications focus primarily on enhancing productivity, their evaluation often overlooks critical factors such as safety and cost-effectiveness. The REF offers a structured, uniform approach to assess the implementation of construction robots, addressing both practical performance and broader structural considerations [25, 26].

3 Methodology

This analysis focuses on utilizing the REF, a methodical tool used for uniform evaluations of the integration of robotics and autonomous systems into construction sites [7]. The REF was selected for this study due to its structured approach to assessing the integration of robotics into construction projects. The REF balances feasibility, performance, and risk considerations, offering a comprehensive evaluation addressing technical and operational factors critical to successful deployment. The REF consists of two key stages. The first stage evaluates the feasibility of deploying a robot on a project [7, 27]. This is done by analyzing three independent variables – product, organization, and process – referred to as the POP analysis:

1. **Product:** Evaluates robot features for the work environment.
2. **Organization:** Assesses worker and organizational requirements.
3. **Process:** Covers operational changes, quality assurance, and documentation.

After determining feasibility through the POP analysis, the REF then evaluates the overall performance and impact of robotic implementation compared to manual construction through the SQSC analysis, which focuses on safety, quality, schedule, and cost. This offers a comparison between robotic implementation and manual deployment [7, 27]:

4. **Safety**: Focuses on quantifying the impact accidents, ergonomic strain, and hazards.

5. **Quality**: Evaluates the accuracy of robotic construction work.
6. **Schedule**: Assesses task completion time with robots vs. traditional construction.
7. **Cost**: Compares initial investment to long-term returns.

4 Use Case

4.1 Overview

The REF methodology is being implemented on the autonomous driving dumper, developed by the Chair of Materials Handling, Material Flow, Logistics at the Technical University of Munich (TUM) together with the industry partner Wacker Neuson (see Figure 1). It is compared with its conventional counterpart, Wacker Neuson DT10e Dumper and is evaluated in terms of feasibility and advantages for the use on-site.



Figure 1. Autonomous dumper project, developed by TUM and Wacker Neuson

4.2 POP analysis

The first step is to perform an initial feasibility check for the robot, in this case, the autonomous dumper, using the POP analysis.

4.2.1 Product analysis

To perform product analysis, thorough information was gathered through dispatched observations based on pictures, videos, and interviews with researchers, industry partners, and instructors. The objective of the autonomous dumper is to transport bulk materials on construction sites. The dumper has a high degree of mobility, making it capable of remote control operation or even following autonomously. In the future, this equipment might interact with other dumpers or equipment at the site. The dumper has a steering interface, a remote control via WLAN, and sensors such as IMUs for the position of the bucket, the loading unit, and the main vehicle, GNSS for the global position, cameras for object detection, and ultrasonic sensors for safety reasons. Sensor data is processed locally by the autonomous dumper. Weighing 680 kg and featuring a ground clearance of 80 mm, the autonomous dumper can operate in harsh conditions. However, challenges may arise during heavy rain when certain sensors struggle to detect objects. The dumper is not commercially available, as it is part of a research project at the Chair of Materials Handling, Material Flow, Logistics at the TUM. Its potential applications extend to any construction site requiring the transportation of bulk materials in confined spaces.

4.2.2 Organization analysis

Introducing the autonomous dumper into the construction process offers some organizational advantages. In Germany, operators must undergo three years of training before operating heavy equipment, making the existing manual labor shortage more noticeable. Using the autonomous dumper requires less training, and the operator can simultaneously work on multiple pieces of equipment, allowing for a more efficient work outcome. Instead of a conventional dumper, one can use one or multiple autonomous dumpers and improve flexibility on the construction site. However, there is no experience with robots or higher levels of automation, and there is little acceptance of electric or automated equipment at the training grounds.

4.2.3 Process analysis

The processes were analyzed using the videos provided for the autonomous dumper, the results of Zöls et al. [8], and expert interviews at the Bayerische Bauakademie. There are some process changes when switching from the conventional to the autonomous dumper. The communication shifts from verbal to digital. While transporting bulk materials is not typically considered a value-adding task, basic quality control measures — such as ensuring load stability and minimizing spillage — are still relevant. In this use case, manual and autonomous dumpers maintain similar standards in these areas, leading to no significant difference in quality outcomes. There are also no progress reports on a process level, except for a construction load, which documents whether an area is cleared. The detailed workflows of manual and robotic construction are depicted in Figure 2.

The process for the conventional dumper begins with the operator starting the vehicle and getting on the equipment. The operator starts with the loading process, which takes 51s. Then, the operator maneuvers the load to its destination for 51s and unloads it within 11s. To end the process, the operator drives to the starting point for

23s while being empty.

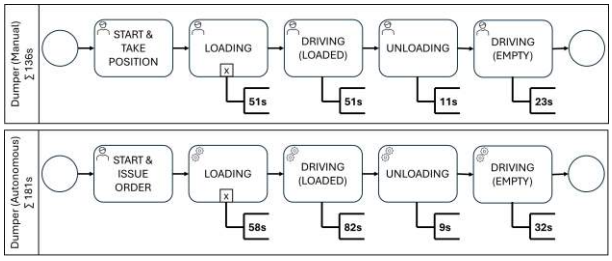


Figure 2. Comparison of the conventional and autonomous dumper

The workflow of robotic construction is slightly different. The process begins with the operator providing the autonomous dumper with its task. After this, no interaction with the operator is required, and the operator can focus on value-adding tasks. The process for the autonomous dumper follows the process of the conventional dumper. The loading takes about 58s, the driving with the load 82s, the unloading 9s, and the driving with no load 32s.

Since the conventional and autonomous dumper can perform similar tasks, the initial feasibility check for the process is positive. All three independent variables check positive for the initial feasibility. This allows for further investigation of this project according to the dependent variables to assess the impacts of deploying the robot.

4.3 SQSC analysis

The dependent variables are now evaluated following the SQSC analysis and the initial feasibility check. The Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) method was selected for this evaluation because it ranks alternatives based on multiple criteria. It effectively balances safety, quality, schedule, and cost factors. This allows for weighted prioritization — such as emphasizing safety over minor schedule delays. This approach ensures a comprehensive assessment of the autonomous dumper's performance relative to manual operations [28]. Safety, given its paramount importance, has a TOPSIS weight of 0.5. Quality, with a TOPSIS weight of 0.1 is less critical since the transportation of bulk materials does not have strict quality standards that either equipment could compromise. The schedule is assigned a weight of 0.2, reflecting the importance of clearing areas within a reasonable timeframe. Costs are also weighted at 0.2, highlighting the need to offer the robot solution competitively.

4.3.1 Safety Analysis

In manual operations, safety depends on several factors. The operator's level of experience plays a crucial role. Additionally, safety is influenced by how well workers adhere to safety rules, such as avoiding movement in blind spots. Manual operation exposes the operator to constant vibrations, impacting health and performance over time. The age of the dumper also impacts safety, as does the type of dumper, particularly in terms of its ergonomics connected to comfort variables. In robotic operations, safety is determined by different criteria. The effectiveness of sensors and software is paramount. The type of human interaction required also affects safety, as does the level of collaboration between humans and robots. The autonomous dumper demonstrated better safety outcomes in terms of ergonomics compared to manual operation, reducing repetitive tasks for operators by 10%. This lowers the risk of strain-related injuries due to vibrations since the operator does not have to stand on the dumper.

4.3.2 Quality Analysis

In manual operations, the quality depends on the operator's experience for accuracy and repeatability, and it is comparable to that of robots. The operator's experience influences the rework needed, but it tends to be slightly better than that of robots. Material waste is excluded since the use case involves transporting bulk material. In robotic operations, the quality depends on the effectiveness of sensors and software for accuracy and repeatability, and it is comparable to that of a conventional dumper. The sensors and software influence the needed rework. Material waste is also excluded from consideration since the use case involves transporting bulk material.

4.3.3 Schedule Analysis

The schedule was evaluated according to Figure 2. The cycle times for the conventional and the autonomous dumper of bulk material are calculated as the sum of the different subprocess times of loading, driving (loaded), unloading, and driving (empty). The overall cycle time of the conventional dumper accumulates to 136s. The cycle time of the autonomous amounts to 181s which is about 20% higher.

4.3.4 Cost Analysis

An autonomous dumper's purchase price is higher than a conventional dumper's, increasing initial capital expenditure (+52.42k euros). However, conventional dumpers lack sensors, limiting automation, while autonomous dumpers are equipped with advanced sensors that enhance efficiency and safety (+5k euros). Conventional dumpers demand more human labor (+22 hours or +1k euros using 44,50 euros/hour), often requiring both an operator and a helper. In contrast, one operator can oversee multiple autonomous dumpers, optimizing labor costs. In total, the costs for the

autonomous dumpers are only 2 % higher (+4k euros p.a.). Autonomous dumpers are cost-competitive for smaller projects with high labor needs, as they handle smaller bulk loads with minimal operator support, making them ideal for confined sites. By enabling operators to focus on value-adding tasks, these dumpers boost productivity. They could also function in collaborative fleets, improving efficiency and reducing project duration and costs [29]. According to the REF framework, autonomous dumpers are viable and cost-effective for small to medium-scale projects, with a payback period of approximately 1.43 years or 75 projects for a full 4-year utilization.

4.3.5 Sensitivity Analysis

In the assessed case, the autonomous dumper outperforms the conventional dumper in the REF. A sensitivity analysis was conducted to further evaluate the performance of the conventional and autonomous dumper. The sensitivity analysis explored variations in cost and cycle time, with deviations of ±25%, ±50%, and ±75%. Results showed that while increased cycle times negatively impacted schedule performance, the autonomous dumper consistently outperformed manual operations in safety and cost metrics. The dumper maintained economic viability even under conservative cost assumptions, with amortization still achievable within 1.6 years. These findings confirm the robustness of the autonomous solution across varying operational scenarios.

5 Challenges and Opportunities

Adopting autonomous construction equipment like the Wacker Neuson DT10e Dumper offers opportunities and challenges. Safety is paramount, requiring reliable systems and technologies like SLAM-based dynamic mapping to ensure seamless integration with workers and equipment, which must be tested across varied conditions to build trust. Limited acceptance of autonomy, particularly in training environments, and the need for effective collaboration with traditional equipment are notable hurdles. The initial dumper tests have demonstrated key safety enhancements in autonomous operation. Unlike conventional dumpers, the autonomous DT10e eliminates operator exposure to vibrations, significantly improving ergonomics. Additionally, the dumper has advanced human-recognition capabilities, allowing it to detect nearby workers and either follow or stop accordingly. These safety features enhance on-site interaction between autonomous and manual operations, reducing the risk of accidents and improving overall workflow efficiency. Scalability also demands advanced systems to manage interconnected fleets. Advanced systems might involve a combination of centralized fleet management software, real-time data-sharing platforms, and robust communication networks.

Despite these challenges, the potential benefits are substantial. The autonomous dumper's compact size and ability to handle lighter loads make it ideal for smaller projects, automating transport tasks so workers can focus on value-adding activities and boosting productivity. Scalable fleets of dumpers could reduce costs and downtime, even competing with wheel loaders. Dynamic mapping enhances collaboration, enabling integration with connected fleets or complementary products like vibratory plates. Contractors can improve efficiency by outsourcing non-value-adding tasks, while manufacturers like Wacker Neuson have the opportunity to expand their market by offering interconnected fleet solutions. The REF evaluation underscores the dumper's potential to address labor shortages, improve efficiency, and drive construction industry innovation.

6 Conclusion and Outlook

The REF evaluation of the autonomous Wacker Neuson DT10e Dumper demonstrates its potential to enhance construction workflows. While quality remains comparable to conventional dumpers, the autonomous system experiences 20% longer cycle times, presenting a scheduling challenge. Despite this, cost analysis reveals amortization after 1.43 years or 75 projects, highlighting its economic viability and potential for fleet scalability. Currently, the dumper's autonomy is limited to driving. Extending its capabilities to include automated loading and unloading would enable fully autonomous bulk material transport, reducing the need for human intervention and increasing operational efficiency. In doing so, it can complement larger equipment, allowing operators to focus on value-adding tasks. In the future, fleets of dumpers could improve efficiency, generate dynamic site maps, and adapt to smaller projects or confined spaces [29]. Future advancements in autonomous dumpers could be driven by integrating AI for task coordination, path planning, enhanced sensor systems for improved real-time obstacle detection, and robust communication networks to facilitate seamless data exchange between machines in a fleet. These technologies enable more complex, coordinated operations, improving overall site efficiency and safety. Future REF studies could explore the scalability of autonomous dumpers and their performance in complex earthmoving tasks, particularly by comparing autonomous fleets to larger traditional equipment.

This paper underscores the value of the REF as a systematic framework for evaluating construction robotics, supporting informed decision-making, and promoting the integration of innovative technologies. While challenges such as acceptance and scalability

remain, the findings underscore the promise of autonomous dumpers to boost productivity and drive innovation. Autonomous construction robotics, such as the Wacker Neuson DT10e Dumper, hold significant potential to address key industry challenges, including labor shortages, low productivity, and the growing demand for innovation and automation in construction workflows. Advancements in technology and broader adoption of REF will be crucial for realizing the full potential of construction robotics.

Acknowledgements

This research was funded by the TUM Global Incentive Fund. All support is gratefully acknowledged. Special thanks go to Prof. Fischer for the possibility to join the Construction Robotics class (CEE 327) at Stanford University, Wacker Neuson for the joint dumper development and Bayerische Bauakademie for the test opportunity. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the funders or industry partners.

References

[1] McKinsey & Company, The next normal in construction: *How disruption is reshaping the world's largest ecosystem.* Online: www.mckinsey.com, Accessed: 09/12/2024.

[2] Dolny O. R., Remmers C., and Michels T., The Future of Construction Machinery Manufacturers. Online: www.porsche-consulting.com, Accessed: 09/12/2024.

[3] PWC PricewaterhouseCoopers GmbH Wirtschaftsprüfungsgesellschaft, Bauindustrie unter Druck. Online: www.pwc.de, Accessed: 09/12/2024.

[4] McKinsey Global Institute, Reinventing Construction: A route to higher productivity. Online: www.mckinsey.com, Accessed: 24/12/2024.

[5] Wadhwani P. and Ambekar A., Autonomous Construction Equipment Market Size - By Product, Forecast 2024 - 2032. Online: www.gminsights.com, Accessed: 09/12/2024.

[6] Leiss F., Shortage of Skilled Workers on German Construction Sites Worsens. Online: https:// www.ifo.de/en/press-release/2021-10-06/shortage-skilled-workers-german-construction-sites-worsens, Accessed: 24/12/2024.

[7] Brosque, C. and Fischer, M., A robot evaluation framework comparing on-site robots with traditional construction methods. *Constr Robot*, 6, (2), pages 187–206, 2022.

[8] Zöls, K., Schock-Schmidtke, A., Rothmeyer, F., Huber, A., Vidal Soroa, D., Wohlschläger, L., Schöberl, M., and Fottner, J., Efficiency of an Automated Bulk Material Transport. in *10. Baumaschinenfachtagung, Dresden, 2024*.

[9] Liu, Y., A.H., A., Haron, N. A., N.A., B., and Wang, H., Robotics in the Construction Sector: Trends, Advances, and Challenges. *J Intell Robot Syst*, 110, (2), 2024.

[10] Attalla, A., Attalla, O., Moussa, A., Shafique, D., Raean, S. B., and Hegazy, T., Construction robotics: review of intelligent features. *Int J Intell Robot Appl*, 7, (3), pages 535–555, 2023.

[11] Melenbrink, N., Werfel, J., and Menges, A., On-site autonomous construction robots: Towards unsupervised building. *Automation in Construction*, 119, p. 103312, 2020.

[12] Yang, C.-H., Tsai, L.-T., Chen, Y., Hei Chui, Y., Liang, C.-J., and Alton, B., Design and Prototype of a Multi-Functional Robotic Cell for the Fabrication of Exterior Retrofit Panels. In *Proceedings of the 3rd Future of Construction Workshop at the International Conference on Robotics and Automation (ICRA 2024)*, Yokohama, Japan, 2024.

[13] Yang, Y. and Pan, W., Automated guided vehicles in modular integrated construction: potentials and future directions. *CI*, 21, (1), pages 85–104, 2021.

[14] Dierichs, K., Kyjánek, O., Loučka, M., and Menges, A., Construction robotics for designed granular materials: in situ construction with designed granular materials at full architectural scale using a cable-driven parallel robot. *Constr Robot*, 3, 1-4, pages 41–52, 2019.

[15] Tho, T. P. and Thinh, N. T., Using a Cable-Driven Parallel Robot with Applications in 3D Concrete Printing. *Applied Sciences*, 11, (2), p. 563, 2021.

[16] Chu, B., Jung, K., Lim, M.-T., and Hong, D., Robot-based construction automation: An application to steel beam assembly (Part I). *Automation in Construction*, 32, pages 46–61, 2013.

[17] Hachijo, T., Igarashi, S., Tani, T., and Indo, M., Development of Autonomous Robots for Construction. In *Proceedings of the 41st International Symposium on Automation and Robotics in Construction*, Lille, France, 2024.

[18] Melenbrink, N. and Werfel, J., Autonomous Sheet Pile Driving Robots for Soil Stabilization. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 339–345, Montreal, QC, Canada, 2019.

[19] Wang, Z., Li, H., and Yang, X., Vision-based robotic system for on-site construction and demolition waste sorting and recycling. *Journal of Building Engineering*, 32, p. 101769, 2020.
[20] Jud, D., Kerscher, S., Wermelinger, M., Jelavic, E., Egli, P., Leemann, P., Hottiger, G., and Hutter, M., HEAP - The autonomous walking excavator. *Automation in Construction*, 129, p. 103783, 2021.
[21] Komatsu, T., Konno, Y., Kiribayashi, S., Nagatani, K., Suzuki, T., Ohno, K., Suzuki, T., Miyamoto, N., Shibata, Y., and Asano, K., Autonomous Driving of Six-Wheeled Dump Truck with a Retrofitted Robot. in *Springer Proceedings in Advanced Robotics, Field and Service Robotics*, G. Ishigami and K. Yoshida, Eds., Singapore: Springer Singapore, 2021, pages 59–72.
[22] Chen, Y., Jiang, H., Shi, G., and Zheng, T., Research on the Trajectory and Operational Performance of Wheel Loader Automatic Shoveling. *Applied Sciences*, 12, (24), p. 12919, 2022.
[23] Borngrund, C., Bodin, U., Sandin, F., and Andreasson, H., Autonomous Navigation of Wheel Loaders using Task Decomposition and Reinforcement Learning. In *2023 IEEE 19th International Conference on Automation Science and Engineering (CASE)*, pages 1–8, Auckland, New Zealand, 2023.
[24] Jung, K., Chu, B., and Hong, D., Robot-based construction automation: An application to steel beam assembly (Part II). *Automation in Construction*, 32, pages 62–79, 2013.
[25] Skibniewski, M., Vanegas, J., and Russell, J., Construction Robotic Equipment Management System (CREMS). In *Proceedings of the 6th International Symposium on Automation and Robotics in Construction (ISARC)*, San Francisco, CA, USA, 1989.
[26] Brosque, C. and Fischer, M., Safety, quality, schedule, and cost impacts of ten construction robots. *Constr Robot*, 6, (2), pages 163–186, 2022.
[27] Brosque, C., Galbally, E., Khatib, O., and Fischer, M., Human-Robot Collaboration in Construction: Opportunities and Challenges. In *2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, pages 1–8, Ankara, Turkey, 2020.
[28] Hwang, C.-L. and Yoon, K., Methods for Multiple Attribute Decision Making. in *Lecture Notes in Economics and Mathematical Systems, Multiple Attribute Decision Making*, M. Beckmann, H. P. Künzi, C.-L. Hwang, and K. Yoon, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 1981, pages 58–191.
[29] Hiltunen, M. S. and Heikkilä, R. K., Advancements and Future Visions in Earthmoving Swarm Technology. In *Proceedings of the 41st International Symposium on Automation and Robotics in Construction*, Lille, France, 2024.

LLM-PCPP: A novel robot prioritized coverage path planning approach for complex environments

Zhaofeng Hu[1] **and Ci-Jyun Liang**[1]

[1]Department of Civil Engineering, Stony Brook University, USA

zhaofeng.hu@stonybrook.edu, ci-jyun.liang@stonybrook.edu

Abstract -

This study introduces a novel prioritized coverage path planning method that combines the efficiency of Large Language Models (LLMs) with traditional optimization techniques to enhance the identification and traversal of key points of interest (POIs) on semantic maps of complex environments. By leveraging LLMs, the proposed approach dynamically identifies POIs, addressing inaccuracies and inefficiencies associated with manual selection. The LLM-A* algorithm is then utilized to compute the shortest paths between POIs, significantly reducing operational and storage parameters and thereby minimizing computational overhead. Finally, the traversal problem is formulated as a Traveling Salesman Problem for optimal route planning. This method demonstrates substantial improvements in computational efficiency, particularly benefiting resource-constrained robotic systems, and offers a scalable solution for efficient path planning in complex environments.

Keywords -

Prioritized Coverage Path Planning ; Large Language Model; A* Algorithm;

1 Introduction

The use of autonomous robots and drones [1] in task-specific environments has gained significant attention due to their ability to optimize operations, reduce human error, and improve safety [2]. In outdoor construction sites, indoor warehouses, and other complex environments, efficient path planning for robotic systems is essential to achieve task goals effectively. In robot navigation, coverage path planning (CPP) focuses on ensuring complete area coverage [3]. However, these uniform approaches fail to account for the heterogeneous complexity of real-world environments, leading to redundant scans in simple regions and insufficient focus on critical areas.

To address these challenges, prioritized CPP has emerged as a promising solution for selectively covering critical points of interest (POIs) while minimizing resource consumption [3]. However, identifying these POIs is a challenging and error-prone process, often reliant on manual input or predefined heuristics, which limits scalability and introduces inaccuracies. Moreover, computational constraints in resource-limited devices, such as drones or small mobile robots, require more efficient algorithms for coverage path planning.

This study introduces a novel prioritized CPP method that leverages Large Language Models (LLMs) for identifying POIs on semantic maps, combined with the LLM-A* [4] algorithm for optimizing paths between these points. By utilizing the rich contextual information embedded in semantic maps [5], our approach enables more accurate and dynamic identification of key regions of interest, addressing inefficiencies and errors associated with manual selection. LLMs analyze the semantic structure of the environment, allowing the system to prioritize areas based on task-specific relevance. The LLM-A* algorithm further enhances efficiency by reducing operational and storage requirements, making it particularly suitable for resource-constrained robots. After connecting the identified POIs through LLM-A*, the problem is modeled as a Traveling Salesman Problem (TSP) [6] to determine the optimal path.

2 Related Work

2.1 Coverage Path Planning

Coverage Path Planning (CPP) is a fundamental problem in robotics, with applications ranging from environmental monitoring to robotic cleaning and agricultural surveying. Traditional CPP methods aim to ensure complete area coverage [3]. Although these approaches are straightforward and widely used, they often fail to account for semantic information about the environment, leading to inefficiencies in resource allocation and redundant coverage in simple regions. To address these limitations, non-uniform CPP strategies have been proposed. For instance, Kazemdehbashi [7] introduced an adaptive grid-based decomposition algorithm that optimizes UAV coverage paths by dynamically adjusting grid size based on environmental complexity, reducing computation and coverage time. In addition, Girdhar et al. [8] proposed a curiosity-driven path planning approach, leveraging information-theoretic metrics to prioritize regions with higher informational value. Similarly, Sadat et al. [9] introduced a coverage strategy

utilizing space-filling curves to achieve efficient exploration of regions with varying levels of interest, making it suitable for tasks requiring adaptive focus on high-priority areas. Manjanna et al. [10] proposed a reward-driven selective coverage path planning algorithm using a finite-horizon Markov Decision Process (MDP), which prioritizes high-value regions and generates time- and energy-efficient trajectories for multi-robot path planning.

While non-uniform CPP strategies demonstrate adaptability and regional prioritization, their reliance on predefined heuristics and manual metrics often limits their applicability in complex environments. Our approach addresses these limitations by utilizing LLMs for automated and dynamic identification of POIs from semantic maps. Additionally, the integration of the LLM-A* algorithm ensures computational efficiency and scalability, making it well-suited for diverse real-world scenarios.

2.2 Large Language Models in Path Planning

Large Language Models (LLMs), such as GPT [11], have been employed to generate task-specific goals, interpret natural language instructions, and provide context-aware decision-making for navigation tasks [12]. The integration of LLMs into robots has opened new avenues for enhancing path planning through their ability to process and interpret high-level semantic information. Latif [13] proposed a probabilistic path planning approach using GPT-3.5-turbo, showcasing its ability to optimize robot navigation in complex environments. The method outperformed traditional algorithms like A* and RRT in processing time, demonstrating the potential of LLM in terms of efficiency and cost savings. Meng et al. [4] introduced LLM-A*, a hybrid algorithm combining the global reasoning capabilities of large language models with the precise pathfinding of A*. The approach significantly reduces computational operations and memory usage while maintaining path validity, making it well-suited for large-scale path planning scenarios.

Furthermore, LLMs have been utilized in coverage path planning tasks. Kong et al. [14] developed a multi-layer framework that integrates LLMs into coverage path planning for mobile robots. Their approach enhances the spatial inference capabilities of LLMs and achieves efficient, reliable path planning by combining high-level reasoning with low-level control, demonstrating the potential for LLMs in embodied AI tasks. These studies demonstrate the potential of LLMs to go beyond traditional numerical optimization methods by incorporating contextual and semantic understanding into the path planning process. Despite these advancements, those methods are limited to path planning tasks between two points or simply covering path planning tasks without obstacles. Our study builds on these foundations by leveraging LLMs to identify POIs and integrating them into a prioritized coverage path planning framework. This framework can handle complex environments and effectively bridge the gap between semantic understanding and efficient navigation.

3 Methodology

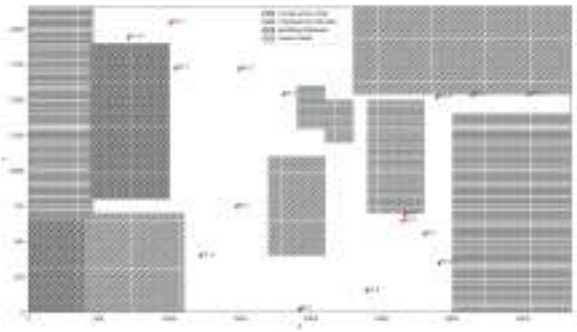
Our proposed methodology integrates semantic information extraction, point of interest (POI) identification, and optimized path planning. First, the environment is represented as a 2D semantic grid map, combining spatial and contextual information. Points of Interest (POIs) are then identified using a Large Language Model (LLM), which leverages semantic reasoning to prioritize task-relevant regions. The LLM-A* algorithm computes efficient paths between POIs while maintaining scalability and computational efficiency. Finally, the problem is formulated as a Traveling Salesman Problem (TSP), with the optimal route derived through heuristic and constraint-based optimization.

3.1 Semantic Information Grid Map Construction

To enable computational processing and algorithmic path planning, the first step is to convert the real-world environment into a 2D semantic information grid map. This discretization captures the spatial features of the environment while simplifying the problem for efficient computation. Data representing the real-world environment is collected through various sources, such as satellite imagery, LiDAR scans, or high-resolution drone footage (Figure 1a). This data provides essential spatial and structural details, including boundaries, obstacles, and open areas. Subsequently, data processing is performed to construct the grid map, ensuring that the number and size of the grids align with the resolution of the original image to maintain high-fidelity modeling, as shown in Figure 1b. Each cell in the grid represents a discrete area in the real-world environment. The processed data is converted into a grid-based matrix representation. Each cell in the matrix is assigned a binary value indicating its navigability. For example, cells representing obstacles are marked as non-navigable, while open spaces are marked as traversable. We convert the obstacles to rectangular shapes for computational efficiency, but they can be represented as complex geometries. This grid map forms the basis for subsequent steps, enabling spatial computations and path planning algorithms. In addition, this grid map is accompanied by semantic information, i.e., natural language descriptions indicating the location and type of obstacles, which will be provided to the LLM for reasoning.



(a) A real environment captured by a drone.



(b) Semantic grid map.

Figure 1. Real Environment and Semantic Grid Map.

3.2 Point of Interest (POI) Identification

The process begins with identifying Points of Interest (POIs) within the grid map. These POIs represent critical locations that must be prioritized during path planning and are determined using the semantic and contextual capabilities of a Large Language Model (LLM). By employing prompt engineering [15] to provide the LLM with a detailed description of the grid map and task-specific objectives, such as locating operational zones, high-risk areas, or strategic points, the model processes the input and leverages its contextual understanding to identify areas of significance based on task requirements.

The LLM then generates a list of potential POIs, which are subsequently validated against predefined criteria to ensure their relevance and alignment with the overall objectives. For example, POIs may include equipment storage areas, hazard zones, or task-specific targets. Once validated, the identified POIs are annotated on the 2D grid map, as illustrated in Figure 1b. This annotated visualization serves as a tool to verify the accuracy of the LLM's output and provides a clear depiction of the prioritized areas, laying the foundation for the subsequent path planning processes.

3.3 Distance Calculation Using LLM-A*

The next step is to compute the shortest paths between all pairs of POIs using the LLM-A* algorithm. This hybrid approach improves computational efficiency and reduces memory usage while maintaining path validity. The LLM-A* algorithm works by integrating LLM-generated waypoints as intermediate targets to guide the A* search process.

The POIs and the obstacles in the grid map are first defined as input states. The LLM-A* algorithm generates a sequence of waypoints using global insights from the LLM, providing a coarse roadmap through the environment. Waypoints divide the entire environment into distinct regions for partitioned processing, followed by path computation using the standard A* algorithm. This method of regional division significantly optimizes the computational cost of heuristic path planning algorithms [16]. Additionally, with the introduction of LLM for waypoint identification, the waypoints gain functional coverage capabilities. While the robot's path length may slightly increase, it enables the robot to cover more points on the map that are rich in semantic information.

The algorithm builds a distance matrix by calculating the shortest paths between all pairs of POIs. Each pairwise calculation leverages the LLM-A*'s ability to avoid unnecessary node expansions, significantly reducing the number of operations and storage requirements compared to traditional A*. By incorporating LLM-generated global insights with the deterministic precision of A*, LLM-A* enables the efficient and accurate computation of the distance matrix, laying the groundwork for the subsequent Traveling Salesman Problem (TSP) optimization. This hybrid approach addresses the limitations of standalone A* and LLM-based methods, ensuring both scalability and robustness in large-scale path planning scenarios.

3.4 Traveling Salesman Problem (TSP) Solution

To determine the optimal path visiting all identified POIs exactly once and returning to the starting point, the Traveling Salesman Problem (TSP) is formulated using the distance matrix from 3.3. Google OR-Tools is employed to solve the TSP, utilizing a hybrid approach that balances computational efficiency with solution quality. The initial solution is generated using the *PATH_CHEAPEST_ARC* heuristic, which iteratively selects the minimum-cost edges to construct a feasible route. This strategy provides a fast approximation to the TSP. Further optimization is achieved through constraint-based refinement, ensuring the path adheres to task-specific requirements, such as visiting all POIs exactly once and returning to the starting point. This approach leverages heuristic initialization and constraint optimization to

achieve near-optimal solutions while maintaining computational efficiency.

4 Expirement

4.1 Experimental Setup

To evaluate the effectiveness and efficiency of the proposed approach, we conducted experiments in simulated environments with varying levels of complexity, including one virtual indoor scene and two real outdoor construction site scenes. The experiments were designed to evaluate the performance of the LLM-A* algorithm and its integration into the overall system for solving the TSP.

Real outdoor construction site scenarios were represented using 2D grid maps generated with different resolutions and varying distributions of obstacles. The maps ranged from simple layouts with minimal obstacles to highly complex configurations with dense barriers. POIs were placed strategically across the maps to simulate real-world task requirements.

The simulation environment was configured to reflect realistic constraints, including image resolution, number of POIs (5 to 15 per map), and obstacle density (low 10%, medium 30%, and high 50%).

The experiments were implemented in Python, with the LLM-A* algorithm and TSP solver integrated into the workflow. The LLM used for global reasoning was GPT-3.5, accessed via API. For TSP optimization, a genetic algorithm was employed. The intermediate targets for LLM A* were determined in LLM and fed to A* Python code independently.

4.2 Evaluation Metrics

The proposed approach is evaluated using three key metrics: path length, operations, and storage.

Path Length: This metric measures the total distance of the computed path. While shorter paths are typically desirable in many scenarios, for prioritized coverage path planning (PCPP) tasks, a slightly longer path may be beneficial as it allows for the exploration of more semantically rich areas, aligning with the task's objectives.

Operations: The number of computational steps required to compute the path is recorded. This metric reflects the algorithm's efficiency in terms of processing effort and computational complexity.

Storage: The memory usage during the pathfinding process is measured to evaluate the algorithm's scalability and suitability for resource-constrained systems. Lower storage requirements indicate better optimization of the search process.

4.3 Experimental Details

4.3.1 Experiment 1: Complex Outdoor Construction Site Scenario

The first experiment evaluates the performance of the proposed approach in a complex outdoor construction site scenario derived from drone footage. The scene was converted into a high-resolution semantic grid map to simulate realistic navigation challenges. This real-world scenario and related grid map are shown in Figure 1. The resolution of the grid map is 3840×2160 with a high obstacle density, representing 50% of the grid cells. ChatGPT-4o served as the agent for generating semantically meaningful POIs and providing heuristic guidance for the LLM-A* algorithm. Table 1 summarizes the key parameters of this experimental setup.

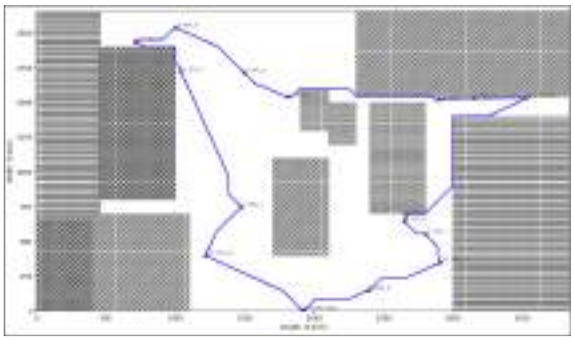
Table 1. Details of Experiment 1: Complex Outdoor Scenario

Parameter	Value
Grid Map Resolution	3840 × 2160
Obstacle Density	High (50%+)
Obstacle Types	Green Fields, Construction Sites, Equipment, Building Materials
Task Guidance	Avoid the obstacles, ensure coverage of most areas
LLM Agent for POIs	ChatGPT-4o
LLM Agent for LLM-A*	ChatGPT-4o
Number of POIs	14

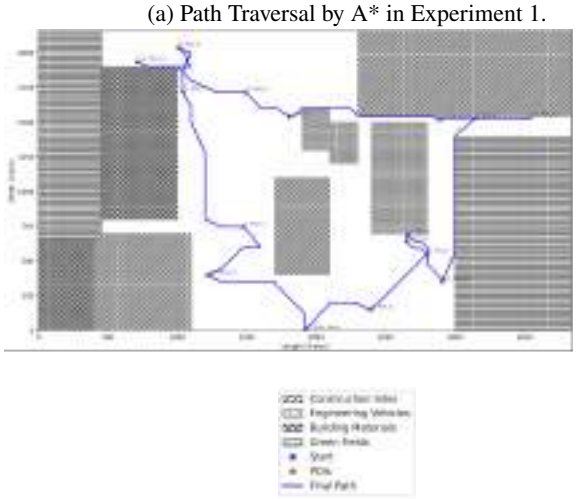
Figure 2 and Table 2 illustrate the comparative performance of A* and LLM-A* in Experiment 1. Figure 2 presents the specific paths taken by each algorithm. The paths highlight LLM-A*'s focus on semantically rich exploration over strict distance minimization. Table 2 provides quantitative insights into the results, demonstrating that while the total path length for LLM-A* increased by 13.6%, it significantly reduced operations by 41.1% and storage requirements by 36.9%. Therefore, Figure 2 and Table 2 highlight LLM-A*'s effective balance between computational efficiency and semantic exploration, making it an appropriate choice for prioritized coverage path planning tasks in complex environments.

Table 2. Comparison of A* and LLM-A* in Experiment 1

Metric	A*	LLM-A*
Total Path Length (TSP)	9270.45	10530.60 ↑ 13.6%
Total Operations	72343616	42636892 ↓ 41.1%
Total Storage	44987486	28383237 ↓ 36.9%



(a) Path Traversal by A* in Experiment 1.

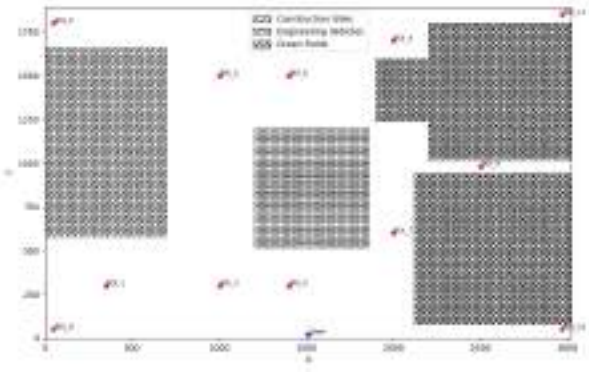


(b) Path Traversal by LLM-A* in Experiment 1.

Figure 2. Path Planning Comparison in Experiment 1: A* vs LLM-A*.



(a) Simple real environment in Experiment 2.



(b) Semantic grid map with POIs in Experiment 2.

Figure 3. Real Environment and Semantic Grid Map in Experiment 2.

4.3.2 Experiment 2: Simple Outdoor Construction Site Scenario

The second experiment evaluates the performance of the proposed approach in a simple outdoor construction site scenario derived from drone footage. The scene was transformed into a semantic grid map with a lower resolution compared to Experiment 1, representing a simpler environment for navigation. Points of Interest (POIs) were generated using ChatGPT-o1, while the LLM-A* algorithm employed ChatGPT-4o for heuristic guidance. The scenario of experiment 2 and the related grid map with POIs are in Figure 3.

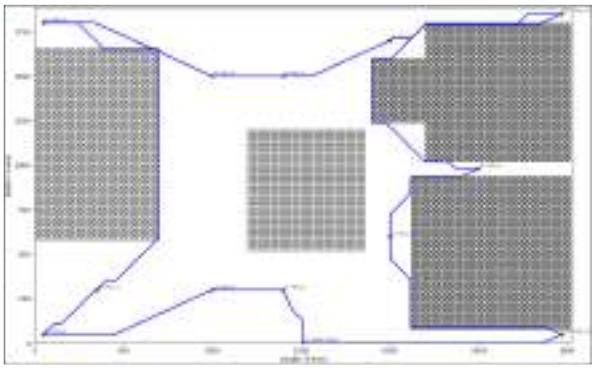
This experiment involved a grid map with a resolution of 3024×1890 and a medium obstacle density, representing 30% of the grid cells. The goal was to evaluate the algorithm's performance in less complex scenarios, focusing on efficiency and scalability. Table 3 summarizes the key parameters of this experimental setup.

Figure 4 and Table 4 illustrate the specific paths and performance metrics for A* and LLM-A* in Experiment 2. Figure 4 highlights the distinct paths taken by each algorithm, showing LLM-A*'s preference for semantically rich exploration, which resulted in a 23.9% increase in total path length as detailed in Table 4. At the same time, LLM-A* demonstrated superior computational efficiency with a 34.2% reduction in operations and a 34.0% reduction

Table 3. Details of Experiment 2: Simple Outdoor Scenario

Parameter	Value
Grid Map Resolution	3024 × 1890
Obstacle Density	Medium (30%)
Obstacle Types	Green Fields, Construction Sites, Equipment, Building Materials
Task Guidance	Avoid obstacles, ensure coverage of most areas
LLM Agent for POIs	ChatGPT-o1
LLM Agent for LLM-A*	ChatGPT-4o
Number of POIs	11

in storage requirements. These findings demonstrate that LLM-A* continues to balance computational efficiency and semantic exploration even in simpler environments.



(a) Path Traversal by A* in Experiment 2.



(b) Path Traversal by LLM-A* in Experiment 2.

Figure 4. Path Planning Comparison in Experiment 2: A* vs LLM-A*.

Table 4. Comparison of A* and LLM-A* in Experiment 2

Metric	A*	LLM-A*
Total Path Length	13562.50	16300.53 ↑ 23.9%
Total Operations	60556220	39873931 ↓ 34.2%
Total Storage	39494015	26063827 ↓ 34.0%

4.3.3 Experiment 3: Virtual Indoor Warehouse Scenario

The third experiment evaluates the performance of the proposed approach in a virtual indoor warehouse scenario. The scene was designed with a resolution of 3024×1890 and included medium obstacle density (30%). The environment featured four types of obstacles: shelves, automated sorting stations, cold storage areas, and hazardous areas. To evaluate the comprehension capabilities of LLMs and expand potential application scenarios, we conducted task guidance tests in this environment focusing not only on obstacle avoidance and maximizing map coverage but also on executing specific task instructions tailored to the setting, such as ensuring detailed exploration around the cold storage area for proper monitoring. LLMs were prompted to understand the task and generate additional Points of Interest (POIs) near the cold storage area. The virtual scenario of experiment 3 only has a designed grid map with POIs, which is in Figure 5.

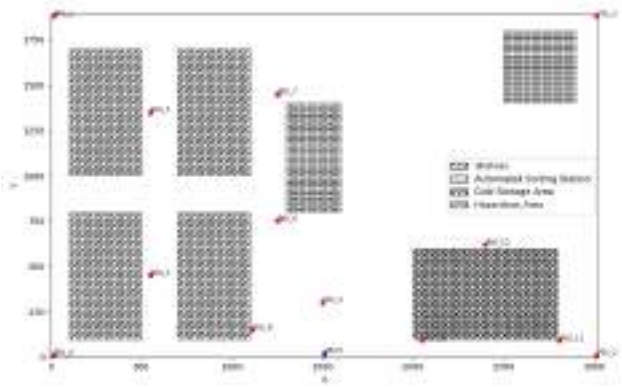


Figure 5. Semantic grid map with POIs in Experiment 3.

ChatGPT-o1 was used to generate semantically relevant POIs, and ChatGPT-4o provided heuristic guidance for LLM-A*. Table 5 summarizes the key parameters of this experimental setup.

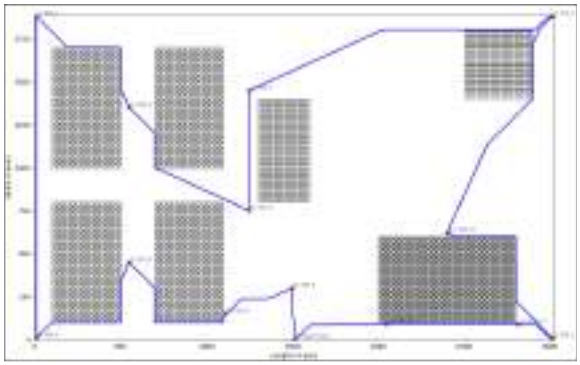
Table 5. Details of Experiment 3: Virtual Indoor Warehouse Scenario

Parameter	Value
Grid Map Resolution	3024 × 1890
Obstacle Density	Medium (30%)
Obstacle Types	Shelves, Sorting Stations, Cold Storage, Hazardous Areas
Task Guidance	Emphasis on cold storage
LLM Agent for POIs	ChatGPT-o1
LLM Agent for LLM-A*	ChatGPT-4o
Number of POIs	78

Figure 6 and Table 6 present the comparative performance of A* and LLM-A* in Experiment 3, conducted

within a virtual environment with specific task guidance. Figure 6 illustrates the distinct paths traversed by each algorithm, showcasing LLM-A*'s approach to semantically rich exploration, which aligns with the task's guided objectives. As shown in Table 6, LLM-A* increased the total path length (TSP) by 18.7% (from 13208.39 to 15682.17), reflecting its prioritization of meaningful interaction over strict path minimization.

At the same time, LLM-A* significantly reduced computational demands, with a 41.8% reduction in total operations and a 30.6% reduction in storage requirements. These improvements demonstrate the algorithm's ability to optimize resources while fulfilling specific guidance task requirements in a controlled virtual setting. Overall, Experiment 3 also highlights LLM-A*'s strength in balancing computational efficiency with task-specific semantic exploration.



(a) Path Traversal by A* in Experiment 2.



(b) Path Traversal by LLM-A* in Experiment 3.

Figure 6. Path Planning Comparison in Experiment 3: A* vs LLM-A*.

Table 6. Comparison of A* and LLM-A* in Experiment 3

Metric	A*	LLM-A*
Total Path Length	13208.39	15682.17 ↑ 18.7%
Total Operations	61625226	35904779 ↓ 41.8%
Total Storage	35970025	24992699 ↓ 30.6%

4.3.4 Comparative Analysis

The results provide a comprehensive evaluation of LLM-A*'s performance in varying environments and task setups. While LLM-A* consistently yields longer overall paths (ranging from 13.6% to 23.9% increases), it significantly reduces computational operations (by 34.2% to 41.8%) and lowers storage requirements (by 30.6% to 36.9%) compared to standard A*. These results underscore the advantage of incorporating semantic guidance into path planning: the algorithm selectively emphasizes task-relevant regions (e.g., general guidance for avoiding obstacles in outdoor settings and specific guidance for monitoring cold storage in warehouses), effectively accessing large segments of the search space.

Furthermore, the slight increase in total path length should not be considered a negative outcome. This increase is akin to an exploration phase, effectively expanding the overall coverage area of the map and fulfilling the requirements of PCPP tasks for semantically rich regions. Therefore, this additional path length enhances the algorithm's capability to cover complex environments. In terms of efficiency, LLM-A* consistently demonstrated significant reductions in both computational operations and storage requirements. These results highlight its strong adaptability and scalability, particularly in resource-constrained systems.

Moreover, the performance of LLM-A* remains consistent across all experiments, demonstrating robustness in various levels of environmental complexity. Notably, the performance slightly improved in Experiment 1, which involved complex environments, and in Experiment 3, which featured medium-density environments with specific task guidance. This suggests that the algorithm is not only adaptable to different environments but also excels when faced with customized task requirements.

5 Conclusion

This paper presented a novel Prioritized Coverage Path Planning (PCPP) approach—referred to as LLM-PCPP—that integrates large language models with conventional path planning techniques. By leveraging LLMs to identify context-rich Points of Interest (POIs) and employing LLM-A* to compute efficient paths between these points, the proposed method significantly reduces com-

putational overhead and memory usage. The experimental results across three distinct scenarios (complex outdoor, simple outdoor, and virtual indoor) demonstrated that, while path lengths increased moderately (13.6% to 23.9%), LLM-PCPP consistently reduced the number of operations (34.2% to 41.8%) and memory requirements (30.6% to 36.9%) compared to standard approaches. These findings underscore the advantage of incorporating semantic cues into coverage path planning by LLM.

Future work will focus on exploring collaborative multi-robot systems and developing complex task-specific guidance, such as determining the farthest points for advanced robots to explore in unknown environments. These efforts will further validate the robustness and scalability of LLM-PCPP, positioning it as a promising solution for advanced coverage applications in resource-constrained and semantically diverse settings. In addition, the POIs found by LLM will be validated in the future. This can be done by consulting with field experts. Finally, benchmarking with the state-of-the-art CPP methods will be conducted to evaluate our PCPP method.

References

[1] Ci-Jyun Liang, Thai-Hoa Le, Youngjib Ham, Bharadwaj RK Mantha, Marvin H. Cheng, and Jacob J. Lin. Ethics of artificial intelligence and robotics in the architecture, engineering, and construction industry. *Automation in Construction*, 162: 105369, 2024. doi:10.1016/j.autcon.2023.105369.

[2] Justin M Haney and Ci-Jyun Liang. A literature review on safety perception and trust during human–robot interaction with autonomous mobile robots that apply to industrial environments. *IISE Transactions on Occupational Ergonomics and Human Factors*, 12(1-2):6–27, 2024. doi:10.1080/24725838.2023.2283537.

[3] Anirudh Krishna Lakshmanan, Rajesh Elara Mohan, Balakrishnan Ramalingam, Anh Vu Le, Prabahar Veerajagadeshwar, Kamlesh Tiwari, and Muhammad Ilyas. Complete coverage path planning using reinforcement learning for tetromino based cleaning and maintenance robot. *Automation in Construction*, 112:103078, 2020. doi:10.1016/j.autcon.2020.103078.

[4] Silin Meng, Yiwei Wang, Cheng-Fu Yang, Nanyun Peng, and Kai-Wei Chang. LLM-A*: Large language model enhanced incremental heuristic search on path planning. *arXiv preprint*, arXiv:2407.02511, 2024. Available at https://arxiv.org/abs/2407.02511.

[5] Zhengyi Chen, Hao Wang, Keyu Chen, Changhao Song, Xiao Zhang, Boyu Wang, and Jack CP Cheng. Improved coverage path planning for indoor robots based on bim and robotic configurations. *Automation in Construction*, 158:105160, 2024. doi:10.1016/j.autcon.2023.105160.

[6] Michael Jünger, Gerhard Reinelt, and Giovanni Rinaldi. The traveling salesman problem. In *Handbooks in Operations Research and Management Science*, volume 7, pages 225–330. Elsevier, 1995.

[7] Sina Kazemdehbashi. Adaptive grid-based decomposition for uav-based coverage path planning in maritime search and rescue. *arXiv preprint*, arXiv:2412.00899, 2024. Available at https://arxiv.org/abs/2412.00899.

[8] Yogesh Girdhar, David Whitney, and Gregory Dudek. Curiosity based exploration for learning terrain models. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 578–584. IEEE, 2014. doi:10.1109/ICRA.2014.6906948.

[9] Seyed Abbas Sadat, Jens Wawerla, and Richard Vaughan. Fractal trajectories for online non-uniform aerial coverage. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2971–2976. IEEE, 2015. doi:10.1109/ICRA.2015.7139617.

[10] Sandeep Manjanna, Nikhil Kakodkar, and Gregory Dudek. Multi-robot path planning for selective coverage. In *ICRA 2016 Workshop on Fielded Multi-robot Systems Operating on Land, Sea, and Air*, Stockholm, Sweden, 2016.

[11] Josh Achiam et al. Gpt-4 technical report. *arXiv preprint*, arXiv:2303.08774, 2023. Available at https://arxiv.org/abs/2303.08774.

[12] Zirui Zhao, Wee Sun Lee, and David Hsu. Large language models as commonsense knowledge for large-scale task planning. *Advances in Neural Information Processing Systems*, 36, 2024.

[13] Ehsan Latif. 3p-llm: Probabilistic path planning using large language model for autonomous robot navigation. *arXiv preprint*, arXiv:2403.18778, 2024. Available at https://arxiv.org/abs/2403.18778.

[14] Xiangrui Kong et al. Embodied ai in mobile robots: Coverage path planning with large language models. *arXiv preprint*, arXiv:2407.02220, 2024. Available at https://arxiv.org/abs/2407.02220.

[15] Jules White et al. A prompt pattern catalog to enhance prompt engineering with chatgpt. *arXiv preprint*, arXiv:2302.11382, 2023. Available at https://arxiv.org/abs/2302.11382.

[16] Adi Botea, Martin Müller, and Jonathan Schaeffer. Near optimal hierarchical path-finding. *Journal of Game Development*, 1(1):1–30, 2004.

Remote Physical Control for Upgrading Heavy Construction Equipment

Meta Soy[1], Jiaming Fu[2], Yifu Wu[4], Jiansong Zhang[1], Dongming Gan[3], Jin Wei-Kocsis[4], and Byung-Cheol Min[4]

[1]School of Construction Management Technology, Purdue University, USA
[2]Department of Mechanical Engineering, University of West Florida, USA
[3]School of Engineering Technology, Purdue University, USA
[4]Department of Computer and Information Technology, Purdue University, USA

soym@purdue.edu, jfu@uwf.edu, wu1584@purdue.edu, zhan3062@purdue.edu, dgan@purdue.edu, kocsis0@purdue.edu, minb@purdue.edu

Abstract

The construction industry heavily uses many different categories of heavy equipment on a daily basis. Generally, such equipment is used extensively and lasts a long time with limited to no possibility of updating or enhancing. Especially, with the current advancements of technology, there is a need to modernize such mechanical equipment. With regard to equipment control, most equipment requires users to stay close to or directly control the equipment. In this paper, the authors propose an upgrading system that allows the remote operation of heavy equipment through push and pull mechanism. By utilizing a single-board computer such as Raspberry Pi with linear actuators, the system can allow the users to control the heavy equipment remotely. The proposed system is successfully demonstrated by upgrading an indoor tower crane in the D. Dorsey Moss Construction Lab at Purdue University. This type of upgrade could support other heavy equipment in the construction field through integrating various sensors and actuators for digital enhancement.

Keywords –

Heavy Equipment Upgrading; IoTs; Construction 4.0; Actuators; Construction Automation

1 Introduction

Construction 4.0 concepts lead to more adoption of various advanced digital technologies into the construction industry, including Building Information Modelling (BIM), Internet of Things (IoTs), Robotics, Automation, etc. [1]. Such integration spurs the need for more digitalization and automation, which results in increased productivity, safety and efficiency in the construction sector. With a focus on IoTs and Robotic Automation, the heavy equipment systems in construction need to be upgraded from traditional control (hands-on) to remote control, and eventually to fully autonomous systems.

While new construction equipment is built with certain smart capacities, there are still many problems with traditional heavy equipment in the sector. This equipment needs upgrades to fit the IoTs and Robotic application contexts of the future. In the work of Khan et al., a lot of emerging technologies have been equipped with heavy construction equipment such as Radio Frequency Identification (RFID), Geographic Information System (GIS), Onboard Instrumentation System (OBIS), Artificial Intelligence (AI), Digital Twin (DT), Internets of Things (IoTs), Cyber Physical Systems (CPSs), Computer Vision (CV), Telematics, and so on [2]. Their study showed how those add-on technologies improved the performance of construction equipment. For example, Hera et al. successfully experimented with fully autonomous forestry machine by equipping the AORO platform with GNSS, UEISIM RTW target and I/O computer, Jetson Xavier and Router, Stereo camera, and so on [3]. Zhang et al., with the purpose of guiding and improving the quality of dynamic compaction for geotechnical improvement process, developed a control system with mounting various supported devices such as AT311 antenna, Hydraulic sensors, Code the sensor, CB71108 car tablets and MC100 Beidou smart host on a mobile crane [4].

While many studies focus on the use of those technologies and their integrated benefit, there are not many that address the upgrade of existing physical systems. In case of an attached innovative physical system, Chen et al. developed a Proportional-Integral-Derivative (PID) based Rotation-control device which

was used to attach at the end of a construction crane's hoist and effectively control suspended item with precise rotation through wireless remote control [5]. Yusof et al. developed a new tele-operated electro-hydraulic actuator which was attached with mini excavator to achieve a full tele-operation of the machinery [6].

While equipment has control buttons for hands-on operations, the ability to control these buttons from a distance is one of the key features for upgrading the system to an IoT with remote control access. Therefore, this paper focuses on developing a plug and play system with a push and pull mechanism, which can be used to allow the remote and/or algorithmic operation of heavy equipment in the construction industry. The proposed system aims to illustrate the potential integration with remote control of a construction crane as a case study.

2 Literature Review

Seeing the significance of automation and robotics in construction, especially the need to upgrade machinery to align with the construction 4.0 concepts, enabling the remote operation of construction tools and equipment is the first step to get them connected. The development of tele-operated electro-hydraulic actuators by Yusof et al. [6] could be seen as a key example of the mechanical attachment to the existing machinery for remote operations. In their work, electro-hydraulic actuators are developed specifically to attach to the mini excavator's various driving shifts through a push-pull mechanism. While the design showcased the possible use of the electro-hydraulic actuator with mechanical levers, the system was overly powerful and too complicated for smaller control hardware such as button press or reduced force lever. For smart control actuators, they were seen the most in IoT applications for smart home control and monitoring [7,8]. In the home IoT applications development, control and monitoring devices' "IoT brains", connectivity protocol, user interface, sensors, actuators, and others are the key components to successfully develop an IoT system [9]. With the purpose of developing a low-cost smart automation system, Eleyan and Fallon [8] developed a system that used Phidget Module to control a Servo with actuator that was used for locking the door. The system was designed to communicate with android phone's application through Message Queuing Telemetry Transport (MQTT) communication protocol [8]. In the development of smart home security system by Islam [10], the microcontroller, Yun Arduino, was used while the MQTT was selected due to its light weight connecting protocol to connect with Android application (from phone) and Arduino board. Rubies et al., while emphasizing the non-contact and non-intrusive device which enable users to control elevator button remotely (from phone), proposed an IoT system using servo motor (actuators) for pushing the button, Esp32 microcontroller, web-based interface, and Wi-Fi as communication protocol [11]. Laluma et al. implemented the IoT in the water treatment plant with the purpose of creating a control automation process through Adafruit.io server. Laluma et al. proposed a control system which uses Raspberry Pi, MQTT protocol, and servo motor (actuator) to control the waterflow [12].

Through the literature review, MQTT protocol is identified as a lightweight message protocol commonly used for IoT applications. And for the brain of the IoT, there are many boards available with different capabilities. Raspberry Pi was seen as a capable Single Board Computer (SBC) which can be used with MQTT protocol and can be configured for various control systems.

3 Methodology

In many IoTs projects that include the remote-control operation, single board computer/micro control is becoming more and more affordable and has sufficient capability in supporting system control development. While the system was purposefully developed for controlling any push-pull mechanism such as button- or lever-based control remotely, our proposed plug and play system in this paper makes use of single board computer, electrically motorized actuators, and control algorithm with MQTT communication protocol. The smart control module/components can be easily customized and attached to the existing control components/buttons of an existing machinery. The potential of the proposed system is demonstrated through lab testing of a tower crane control system.

Figure 1 illustrates the proposed system architecture, which contains two main sub-systems: (1) the user interface which can be run on any computer, and (2) the control system.

3.1 User Interface (UI)

The User Interface (UI) is implemented as a runnable .exe file which is customized to fit the requirement of the control system. Figure 2 shows the interface which is used in the crane control case study in Section 4.

The UI algorithm is developed with TKinter python library with MQTT communication protocol. MQTT is a publish-subscribe and machine to machine (M2M) communication protocol which is widely used for IoTs connectivity [13]. The UI used MQTT as the communication backbone and acted as a client which connected to the MQTT Broker running on the SBC. The MQTT Publisher in the UI allows the user to publish the message from those control actions to the control system (Section 3.2) which will further act accordingly.

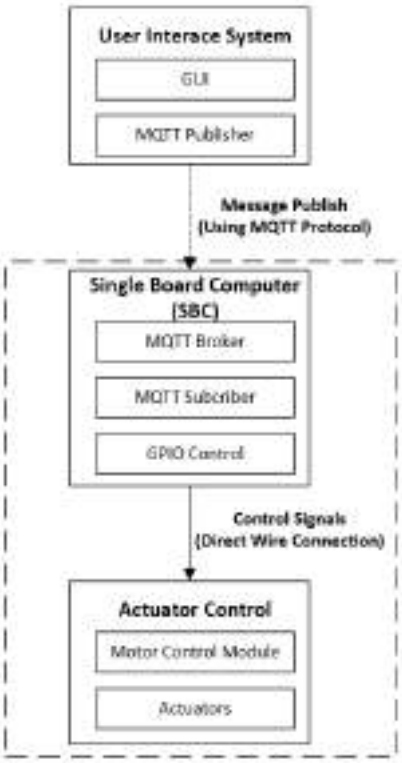


Figure 1. System architecture.



Figure 2. Custom user interface for crane control.

3.2 Control System

The control system (Figure 3) is a key hardware circuit which contains three main components: SBC, motor control driver, and actuator. The SBC, while running the developed control algorithm (Figure 1), triggered the MQTT broker (a central hub for all message exchange between the clients) with a specific IP address. At the same time, the SBC also acted as clients. From the UI, when IP address was provided, the system was connected. It means that through MQTT, the clients (UI) were connected to server. When the message was published from UI, the client on the SBC (MQTT subscriber through server) received the published message. The message through direct General-Purpose Input/Out (GPIO) pins were connected to the actuator control system. In the control algorithm, there is also an error handling part that triggers the emergency stop condition or reset the control when the SBC loses communication with the server.

The actuator control unit is the circuit which makes use of the motor control driver to run the motorized actuators. The selection of the motor control driver allows the control of different states of the motor in the actuator.

The circuit of the control system which uses Raspberry Pi 4B as SBC, L298N as motor driver controller, and Actuonix pq-12-30-12-P as actuator is shown in Figure 3.

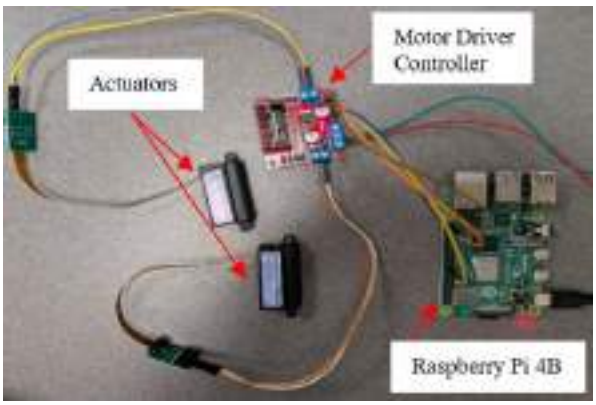


Figure 3. Control system circuit.

4 Application and Discussion

4.1 Hardware Setup

The system was customized as a plug and play instrument to control a crane controller for an indoor tower crane located in the D. Dorsey Moss Construction Lab in the KNOY building at Purdue University. The crane is operated by direct wire controller as shown in Figure 4.



Figure 4. Default crane control.

The default controller has 8 buttons with 8 different operational functions including START, STOP, UP, DOWN, LEFT (IN), RIGHT (OUT), FORWARD (clockwise), and REVERSE (counterclockwise). While the START and STOP are latching buttons, the others are momentary buttons. They are only activated when being pressed.

To attach the control system to the controller, a mount for the actuator was designed and 3D printed, as shown in Figure 5. The mount contains two parts: the upper part supports the actuator, and the lower part secures the entire structure on the crane's controller. The shape and size can be customized and modularized based on various types of crane controllers. This method can replace manual operation while achieving high speed and accuracy for remote button control.



Figure 5. 3-D printed customized mount for the actuator.

As shown in Figure 3, two actuators can be connected to a motor driver controller and connected to Raspberry Pi. To control the 8 buttons of the crane's controller, 8 actuators are connected to 4 motor driver controllers and connected to Raspberry Pi through 16 GPIO pins. For this case, GPIO in and out for STOP, START, UP, DOWN, IN, OUT, FORWARD, and REVERSE are (23;24), (25;8), (26;16), (20;21), (2;3), (17;27), (5;6), and (22;13) respectively. The configuration of the mount with actuators on to the crane's controller is shown in Figure 6. Figure 7 shows the complete set up and testing of the system on the lab's crane.



Figure 6. Crane remote control with mounts and actuators.



Figure 7. System setup for remote access crane.

The system setup for this crane controller needs two sources of power: (1) 12V power was needed for all the motor driver controllers which supplied to the actuators, (2) the Raspberry Pi ran on 3.3V power supply.

4.2 Crane Control User Interface

The development of the crane control UI (shown in Figure 2) resembles the crane controller. The interface required input from: (1) the IP address of the control system (Raspberry Pi as sever), and (2) status information whether the system is connected to the control system or not. As shown in Figure 8, the UI has 8 buttons as the controller which are divided into two different groups. Group 1 controls latching buttons for which a button is pushed to lock in place and the actuator can be used to push the button again to release it back. Figure 9 shows the algorithm which, when applied, allows the actuator to travel 0.3 seconds and push the latching button, then retreat to the original location.



Figure 8. Crane remote control GUI.

```
def extend_and_retract(pin_in, pin_out):
    GPIO.output(pin_in, GPIO.HIGH)
    GPIO.output(pin_out, GPIO.LOW)
    time.sleep(0.3)
    GPIO.output(pin_in, GPIO.LOW)
    GPIO.output(pin_out, GPIO.HIGH)
    time.sleep(0.3)
    GPIO.output(pin_in, GPIO.LOW)
    GPIO.output(pin_out, GPIO.LOW)
```

Figure 9. Control function algorithm for latching button.

The group 2 buttons are of the momentary type, for which continuous pressing is required to activate the

corresponding function. For the interface and activation, when pressing, the state of the button of group 2 (as shown in Figure 10) changes its state in the interface when informing the user that the button is being pressed. When the user clicks on the button one more time, the button state will change to normal, and the actuator retreats to its initial state. The governing function of extending the actuator while being pressed and retreating it when being re-clicked are shown in Figure 11.

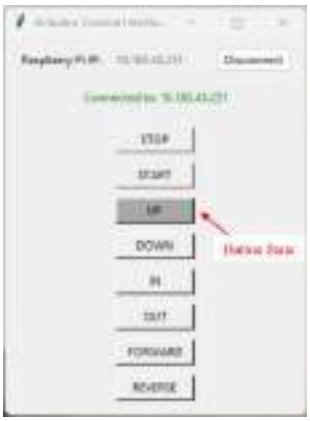


Figure 10. Button state change.

```
def extending(pin_in, pin_out):
    GPIO.output(pin_in, GPIO.HIGH)
    GPIO.output(pin_out, GPIO.LOW)
    time.sleep(0.3)
    GPIO.output(pin_in, GPIO.LOW)
    GPIO.output(pin_out, GPIO.LOW)

def retracting(pin_in, pin_out):
    GPIO.output(pin_in, GPIO.LOW)
    GPIO.output(pin_out, GPIO.HIGH)
    time.sleep(0.3)
    GPIO.output(pin_in, GPIO.LOW)
    GPIO.output(pin_out, GPIO.LOW)
```

Figure 11. Control function algorithm for extending or retracting the button.

As shown in Figures 9 and 11, the *time.sleep* function of the algorithm allows the actuator to travel (extend or retreat) for 0.3 seconds. It is a trial-and-error value which can make sure the length of travel distance is long enough to press the core of the button firmly. The duration can be varied based on the design of the mount.

4.3 Crane Control

After setting up the system as shown in Figure 7 and powering on the raspberry pi, the autorun of the system triggers the algorithm and allows the MQTT broker to operate. On a laptop computer, after launching the crane remote control application (Figure 2), with knowledge of IP address of the Raspberry Pi, the App was successfully connected to the PI. While remotely running the system, the proposed mounted platform allowed the user to remotely control all the functions of the crane's controller successfully. This demonstrates that the proposed system with push and pull mechanism can be integrated with various human-hand-friendly control mechanisms by simply altering the 3D printed mount design or mounting mechanism. For example, the system can be integrated to push or release buttons in a tower crane's cab or to shift operating levers and sticks inside an excavator or a forklift.

5 Conclusion

The construction industry relies heavily on the use of heavy equipment which needs an upgrade to align with the digitalization trend of the construction 4.0 concept. Remote control of those equipment such as excavator, crane, and other mechanical machinery which usually requires hand-on operation are the first steps toward a fully autonomous machine operation. However, to retrofit those components requires significant time and budgets. To address that, this paper illustrates a plug and play system with simple and effective solutions for upgrading existing equipment with button- or lever-based control. This is achieved through a push and pull mechanism to connect to the existing controller. With the MQTT communication protocol, the remote control of the system can be achieved. The completed remote-control system was tested with a construction tower crane in the lab environment. The result showed that it allows the user to successfully control the crane wirelessly. It is a proof of concept that can be used for all different devices with button- or lever-based control mechanism.

Despite the successful demonstration of the system, the following limitations of the system are acknowledged: (1) the 3D printed mounting mechanism is required to be modelled and designed based on the retrofitting purpose and existing control hardware, (2) the control system is mainly based on one type of single board computer (Raspberry Pi 4) and communication protocol (MQTT), and (3) only one selected type of actuator was used. Despite those limitations, the case study demonstrates promising capability and good potential of the system. Future research is recommended to investigate the full tele operation of the crane, linking the physical system with the digital system for digital twin applications, and the development of autonomous operations.

6 Acknowledgement

The authors would like to thank the National Science Foundation (NSF). This material is based on work supported by the NSF under Grant No. 2222838. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

The authors extend their gratitude to Fernando Mututhanthrige P for his valuable contributions to this work.

7 References

[1] H. Aslam and N. Abid, Towards digitizing the construction industry: State of the art of construction 4.0, *International Journal of Construction Research*, vol. 6, no. 1, 2018, doi: 10.14455/ISEC.RES.2019.184.

[2] A. U. Khan, L. Huang, E. Onstein, and Y. Liu, Overview of Emerging Technologies for Improving the Performance of Heavy-Duty Construction Machines, *IEEE Access*, vol. 10, pp. 103315–103336, 2022, doi: 10.1109/ACCESS.2022.3209818.

[3] P. La Hera, O. Mendoza-Trejo, O. Lindroos, H. Lideskog, T. Lindbäck, S. Latif, S. Li, and M. Karlberg, Exploring the feasibility of autonomous forestry operations: results from the first experimental unmanned machine, *Journal of Field Robotics*, vol. 41, pp. 942–965, 2024. doi: 10.1002/rob.22300.

[4] K. Chen, H. Zhou, and X. Shan, Rotation-control device for construction cranes with nested PID control, *Autom Constr*, vol. 146, p. 104704, Feb. 2023, doi: 10.1016/j.autcon.2022.104704.

[5] A. A. Yusof, M. N. A. Saadun, H. Sulaiman, and S. A. Sabaruddin, "The development of tele-operated electro-hydraulic actuator (T-EHA) for mini excavator tele-operation," in *2016 2nd IEEE International Symposium on Robotics and Manufacturing Automation (ROMA)*, IEEE, Sep. 2016, pp. 1–6. doi: 10.1109/ROMA.2016.7847800.

[6] L. Zhang, L. Wang, Z. Tian, J. Li, and B. Luo, Research on construction guide by dynamic compaction under complex foundation condition based on digital equipment, *in Third International Conference on Intelligent Traffic Systems and Smart City (ITSSC 2023)*, 12989, 217-222, 2024.

[7] G. R. Kanagachidambaresan, Actuators Used in Rapid Prototyping, in Role of Single Board Computers (SBCs) in Rapid IoT Prototyping, Internet of Things, Cham: Springer, 2021, pp. 1-10. doi: 10.1007/978-3-030-72957-8_4

[8] A. Eleyan and J. Fallon, IoT-based Home Automation Using Android Application, *2020 International Symposium on Networks, Computers and Communications (ISNCC)*, Montreal, QC, Canada, 2020, pp. 1-4, doi: 10.1109/ISNCC49221.2020.9297320.

[9] M. A. Khan, I. Ahmad, A. N. Nordin, A. E.-S. Ahmed, H. Mewada, Y. I. Daradkeh, S. Rasheed, E. T. Eldin, and M. Shafiq, Smart Android Based Home Automation System Using Internet of Things (IoT), *Sustainability*, vol. 14, no. 10717, 2022. doi: 10.3390/su141710717.

[10] A. Islam, Android Application Based Smart Home Automation System Using Internet of Things, *2018 3rd International Conference for Convergence in Technology (I2CT*), Pune, India, 2018, pp. 1-9, doi: 10.1109/I2CT.2018.8529752.

[11] E. Rubies, R. Bitriá, E. Clotet, and J. Palacín, Non-Contact and Non-Intrusive Add-on IoT Device for Wireless Remote Elevator Control, *Applied Sciences*, vol. 13, no. 3971, 2023. doi: 10.3390/app13063971.

[12] R. H. Laluma, R. Giantara, B. Sugiarto, Gunawan, C. A. Siregar and S. Risnanto, Automation System of Water Treatment Plant using Raspberry Pi.3 Model B+ Based on Internet of Things (IoT), *2019 IEEE 13th International Conference on Telecommunication Systems, Services, and Applications (TSSA)*, Bali, Indonesia, 2019, pp. 72-76, doi: 10.1109/TSSA48701.2019.8985516.

[13] W. Pipatsakulroj, V. Visoottiviseth and R. Takano, muMQ: A lightweight and scalable MQTT broker, *2017 IEEE International Symposium on Local and Metropolitan Area Networks (LANMAN)*, Osaka, Japan, 2017, pp. 1-6, doi: 10.1109/LANMAN.2017.7972165.

Robot-to-Human Construction Tool Handover Grasp Prediction for 6-DOF Robotic Arm with Parallel Gripper

Yu-Lun Kuo[1], Guan-Yong Xiong[2], Jacob J. Lin[1] *, Ci-Jyun Liang[2]

[1]Department of Civil Engineering, National Taiwan University, Taiwan
[2]Department of Civil Engineering, Stony Brook University, USA

R12521717@ntu.edu.tw, R12521614@ntu.edu.tw, jacoblin@ntu.edu.tw, Ci-Jyun.Liang@stonybrook.edu

Abstract -

Workers often spend significant time locating tools on construction sites, resulting in productivity losses. With the continuous advancement of robotic technologies, robot-to-human handover offers a potential solution to address this issue. However, existing robot handover methods face challenges in handling the irregular shapes and dangerous nature of tools, as well as the dynamic and complex environments of construction sites. To tackle these challenges, this research proposes a method that enables robots to effectively predict and grasp construction tools while ensuring safety. A grasp dataset consisting of four commonly used construction tools—pliers, screwdrivers, wrenches, and safety glasses—was generated in a simulation environment. During the annotation process, grasping areas for robots and humans are distinguished to reduce potential risks in handover processes. The dataset includes multi-angle RGB-D images and detailed annotated grasp poses specifically designed for robotic grasps. This proposed dataset provides a foundation for training and validating grasp prediction models, enabling robots to handle tools with irregular shapes in various scenarios. The proposed approach integrates data annotation, scene generation, grasp pose prediction, and motion planning to achieve safe and efficient handovers. Validation experiments conducted in a simulation environment using a UR5 robotic arm equipped with a parallel gripper demonstrate a 92.5% success rate. This research provides a foundation for the application of robot-to-human handovers on construction sites and highlights the potential of using virtual datasets to address real-world challenges.

Keywords -

Robot-to-human Handover; 6-DOF Grasp Dataset; Robotic arm; Grasp pose prediction

1 Introduction

The construction industry is facing labor shortage [1], and workers on construction sites often spend significant time searching for tools, further reducing productivity [2]. With the continuous advancement of robotic technologies, more construction sites are adopting automation and robotic systems to improve efficiency [3]. For example, an increasing number of construction robots have been deployed on construction sites [4]. In this context, robot-to-human handover has become a critical approach to addressing this issue.

Robot-to-human handover consists of three main steps: object searching, localization, grasping, and delivering tools. The entire process must ensure both efficiency and safety [5]. However, construction sites are highly dynamic and complex environments. Most current construction robots perform pre-programmed tasks, making them less adaptable to the variability of construction sites [6]. Compared to the operational environments of existing robot-to-human handover methods, construction sites present greater challenges in terms of environmental complexity, requiring robots to adapt to diverse scenarios and respond flexibly [7]. In addition, construction tools vary widely in size and shape, often exhibiting irregular geometries and potential hazards. Ensuring safety during the handover process is therefore critical.

To address the issues mentioned above, this paper proposes a grasp prediction method for hazardous construction tools. We develop a 6-DOF grasp pose dataset and introduce a method to distinguish between robotic and human grasping regions. The proposed method integrates a grasp dataset from GraspNet [8], grasp pose prediction, and motion planning, enabling robots to effectively grasp handheld tools while ensuring safety. The rest of this paper is organized as follows. Section 2 reviews related work on grasp pose prediction and robot-to-human object handover, identifying research gaps. Section 3 describes the proposed model architecture. Section 4 outlines the experimental setup. Section 5 presents the results. Section 6 concludes the paper and discusses future work.

2 Related Work

This study addresses the challenges of safe and efficient grasping and handover in human-robot interaction, particularly on construction tools, which involve higher safety risks. Consequently, two main research areas are reviewed: (1) Grasp Pose Prediction and (2) Robot-to-Human Han-

dover. We focus on these domains because the quality of a robot's initial grasp directly influences the risk of subsequent handovers—an issue of critical importance in high-risk environments such as construction sites. By examining previous methods, we establish a foundation for our proposed approach to ensure efficient tool handovers through accurate and context-aware grasp prediction.

2.1 Grasp pose prediction

Grasp pose prediction is a critical component of robotic manipulation, especially in tasks that require safe and stable object handovers. Several methods have been proposed to tackle this problem in recent years. Traditional methods often rely on geometric heuristics or predefined rules [9]. However, these methods are limited in their ability to generalize to objects with complex geometries. To address these limitations, some researchers have leveraged CNNs and deep learning techniques, enabling the model to adapt to various shapes [10][11][12]. While deep learning has shown promising results, these methods typically lack context-awareness. Recent research such as Co-Grasp [13] incorporates human preferences in the grasping process, enhancing collaboration between robots and humans.

2.2 Robot-to-Human Handover

Robot-to-human handover refers to a collaborative interaction where one party, the giver, transfers an object to another, the receiver. This process requires precise coordination of both cognitive and physical actions to ensure a smooth exchange [5]. Therefore, researchers have developed a range of approaches to address the difficulties involved in robot-to-human handovers.

Some studies rely on heuristic rules, such as positioning the robotic arm's grasping point on the opposite side of the object from the human hand pose [14]. However, such approaches are not suitable for all types of objects. Other methods remove potential collision-prone grasp points from the human affordance map [15]. While these methods are effective for objects with regular shapes, the inherent risks associated with construction tools place a greater emphasis on safety considerations during the handover process. Additionally, some researchers have proposed using tactile and auditory cues to ensure safety during handovers [16]. However, their work does not address the prediction and selection of optimal grasping points.

In this paper, we propose a method for robotic arms to predict grasping positions specifically for construction tools. Our approach prioritizes safety and provides a foundation for secure and efficient handovers after the grasping phase.

3 Methodology

While traditional grasping methods often rely on simple geometric heuristics or rectangular grasp annotations, they tend to fall short in handling objects with irregular geometries and in providing the necessary context-awareness for dynamic environments. In contrast, GraspNet leverages 6-DOF grasp pose annotations, which allow for more precise and adaptable grasp predictions. This capability is particularly crucial for our application on construction tools, which frequently exhibit complex shapes and potential safety hazards. Therefore, despite the existence of simpler strategies, we chose GraspNet as our baseline due to its demonstrated robustness and superior performance in generating accurate grasp poses for irregular objects.

Based on the limitations in existing studies, we developed a comprehensive workflow to achieve safe construction tool handovers. The proposed methodology is composed of four main components: data annotation, assembly scene generation, grasp prediction, and robot motion planning. These components are designed to work in synergy, enabling safe robot-to-human tool handovers. The proposed framework is shown in Figure 1.

For tool selection, we chose tools commonly found on construction sites. Additionally, the shape of the tools was an important consideration; the selected tools exhibit significant diversity and irregularity in their shapes. In the annotation process, we differentiated between human and robotic grasping regions. This distinction ensures that the predicted grasp poses are focused on designated locations. After annotating the data, we generated the scenes using a Realsense D435 camera to capture images in a simulation environment. Each scene includes RGB-D images, object 6D poses, and masks. For grasp prediction, we employed the GraspNet-baseline [8] model as our prediction framework. We fine-tuned it using the generated dataset and conducted experiments in a simulation environment to verify the reliability of the predicted grasp poses. This approach ensures that the workflow effectively addresses safety in robot-to-human handover.

3.1 Data Annotation

We selected four tools from the Occupational Safety and Health Administration (OSHA) Hand and Power Tools Manual classified as hazardous during operation: adjustable wrench, screwdriver, pliers, and safety glasses. This highlights the importance of ensuring safety during the handover process. The selected tools are illustrated in Figure 2. The screwdriver and adjustable wrench are sourced from the YCB dataset [17], while the pliers and safety glasses are newly introduced objects in this dataset.

Among the selected tools, wrenches, pliers, and screwdrivers are widely used handheld tools on construction

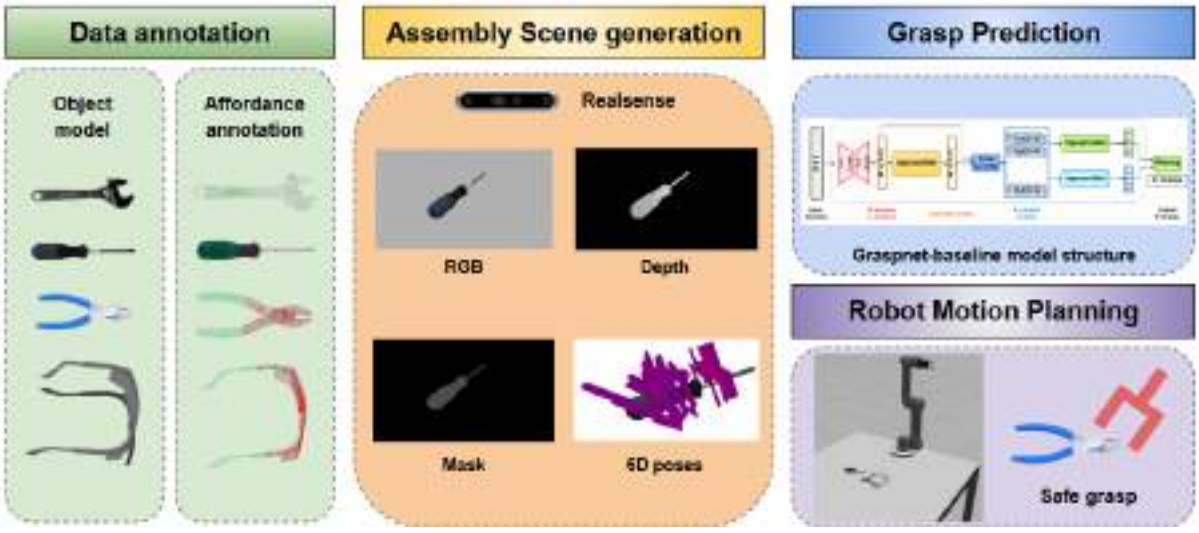
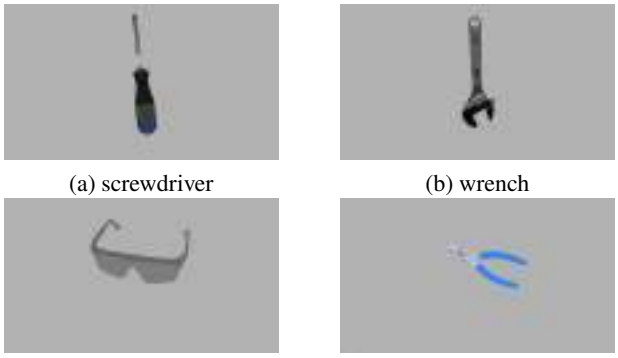


Figure 1. The overview architecture of the proposed handover strategy



(a) screwdriver (b) wrench

(c) safety glasses (d) plier

Figure 2. Commonly used construction tools on construction sites

sites. These tools were specifically chosen for their diverse shapes and sizes, providing a comprehensive evaluation of the versatility of the proposed handover method. Additionally, safety glasses, as an essential protective tool on construction sites, are included to assess their grasping and handover performance in the construction site scenario.

The annotation process consists of two steps. First, based on the tool's usage scenario, we manually annotate the grasping areas for both human and robot interactions. The annotated data is then saved for grasp label generation. For example, in the case of a wrench, humans typically hold the handle for operation, while the other end may pose safety risks due to its sharpness. Therefore, the robot grasping area is annotated on the sharp end of the wrench, while the human grasping area is annotated on the handle.

The second step involves annotating the grasp labels. Various grasp datasets employ different annotation methods based on their intended usage. Examples of well-known datasets, such as Jacquard [18] and Cornell grasping dataset [19], are shown in Table 1. These datasets use rectangle annotations for grasp labels, which are less flexible compared to GraspNet. GraspNet employs annotations in the form of 6-DOF poses, which are particularly advantageous in complex construction environments. The additional precision provided by 6-DOF annotations enhances the robotic arm's ability to execute accurate and reliable movements in such challenging settings. As a result, we chose GraspNet as the annotation format for our grasp labels.

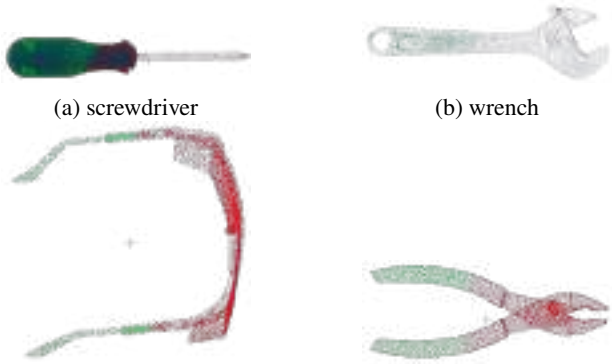
The annotated data are stored as point clouds, with green areas representing human grasping regions and red areas representing robot grasping regions. This separation not only ensures that the robot and human grasping positions are distinguished during the grasp prediction process but also facilitates safe handling by enabling humans, as receivers, to take the object securely from the handle. This dual focus enhances safety in robot-to-human handover scenarios. The annotation for each tool is shown in Figure 3.

3.2 Assembly Scene Generation

In order to construct a dataset for model training, we placed the tools flat on a table and selected the Realsense D435 camera for data collection. Camera calibration was performed prior to data collection to ensure accurate camera poses. The camera captured RGB-D images from 33 different poses, allowing us to gather more information about the tools from various angles. Corresponding masks were manually annotated, and both the camera and object

Table 1. Properties of public grasp datasets

Dataset	Grasp label	6D pose	Total images	Modality	Data source
Cornell [19]	Rect.	No	1035	RGB-D	1 Camera
Jacquard [18]	Rect.	No	54K	RGB-D	Sim
GraspNet [8]	6-DOF	Yes	97K	RGB-D	1 Camera
Ours	6-DOF	Yes	132	RGB-D	Sim



(a) screwdriver (b) wrench

(c) safety glasses (d) plier

Figure 3. Visualization of annotated 3D models of every tool

6D poses were recorded. The dataset included a total of 132 images, along with their corresponding camera and object 6D poses.

3.3 Grasp Prediction

After preparing the training data, we used GraspNet-baseline as the model to predict grasp poses. The structure of the model is shown in Figure 4. During training, the initial learning rate was set to 0.01, and the batch size was 4. The learning rate was reduced to half its value every 5 epochs.

Once the training was completed, we first visualized the predicted grasp poses to verify their validity. Next, the RGB-D images from the simulation environment were used as input to the model, which generated the predicted grasp poses. In the following step, motion planning was employed to test the accuracy of these predicted grasp poses.

3.4 Robot Motion Planning

Once the dataset and scene setup were complete, we employed a fine-tuned model that takes RGB-D images as input and predicts multiple candidate grasp poses in 6-DOF, each accompanied by a confidence score. Among these predicted grasp poses, we select the one with the highest score as the optimal grasping position. The UR5 robotic arm then executes the grasp at this selected pose. This ensures stable and precise grasping by leveraging the model's scoring mechanism to identify the most reliable pose for each tool.

4 Experiment

To evaluate our approach, we compare its performance against the original GraspNet model, which serves as a benchmark. Specifically, the benchmark uses the official GraspNet implementation with the original pretrained weights, without being fine-tuned on our newly collected dataset. In contrast, our model starts from the same pretrained weights but is then fine-tuned on a custom construction-tool dataset containing domain-specific annotations. By juxtaposing the benchmark and our fine-tuned model, we can measure the impact on grasp prediction performance.

We first conduct robotic experiments to demonstrate that our predicted grasp poses can align well with simulation-world grasping. The success rate is calculated as the ratio of successful grasps to total attempts. A grasp is considered successful if the robot securely holds the tool without slippage, avoids collisions during both grasping and returning to the home position, and remains in stable contact with the tool throughout the entire motion. Failures are categorized into three primary types: (1) slippage during the grasping process, (2) collisions with the tool or environment, and (3) the tool dropping while the robot returns to the start pose.

4.1 Experimental Setup

The experiments were conducted in the Gazebo simulation environment using a UR5 robotic arm equipped with a parallel gripper. A RealSense D435 camera was employed for visual input, providing RGB-D data. The detailed experimental setup is depicted in Figure 5. All tools were placed flat on a table to simulate realistic conditions.

4.2 Evaluation Procedure

We tested both the benchmark and our fine-tuned model by predicting the top ten grasp poses (based on each model's grasp scores) for each tool, then executing each

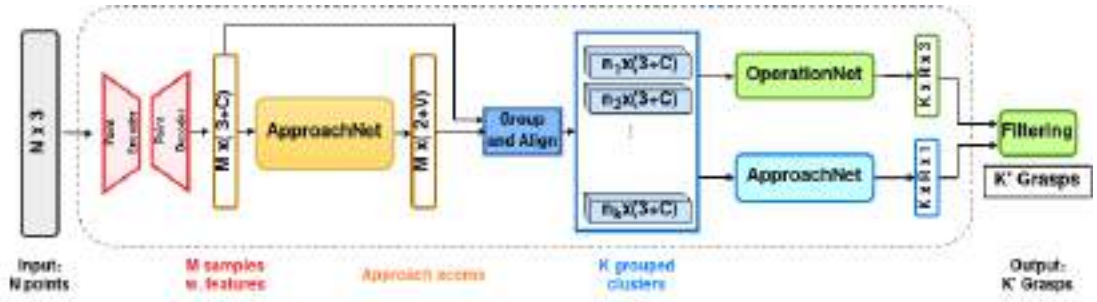


Figure 4. The overview of Graspnet-baseline [8] network

pose in simulation. To determine whether a grasp was successful, we checked:

1. Whether the predicted grasp pose lay within the designated "robot grasp region" of the annotated tool model.
2. Whether the grasp was maintained without slippage throughout the approach and retraction phases.
3. Whether no collision occurred between the robot and the tool (or the environment) at any point.

Any violation of these conditions such as grasping outside the annotated region, colliding with the tool, or dropping the tool during motion was recorded as a failure. In addition to evaluating the top 10 highest-scoring grasp poses, we also tested the highest-scoring pose specifically to check if the model's top-ranked grasp was valid. This procedure provides insight into each model's robustness and flexibility for a range of grasp candidates.



Figure 5. The experiment scene in Gazebo

5 Results

5.1 Benchmarking with GraspNet-baseline

Tables 2 summarizes the results of the benchmark model. The benchmark model achieve a success rate of 30% . For certain objects such as the screwdriver and plier, it exhibits a high collision failure rate. Specifically, with the screwdriver, half of the top 10 grasps failed due to collisions. In particular, when using the original GraspNet model, many predicted grasps lie outside the designated robot-grasp region and often result in collisions with the object, which reduces the overall success rate.

Table 2. Benchmark results: Success rates for the original GraspNet-baseline (top 10 grasp scores)

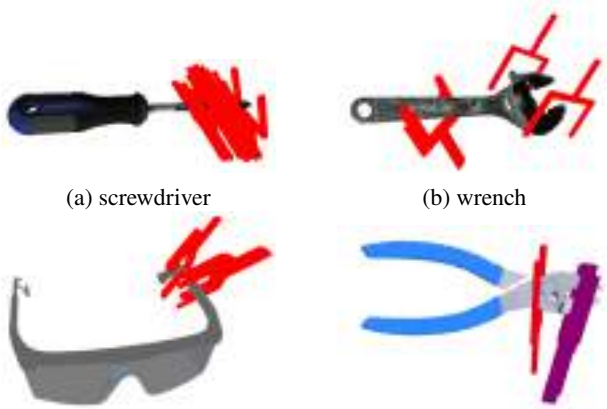
Object	Success rate
Screwdriver	20%
Adjustable wrench	40%
Plier	20%
Safety glasses	40%

5.2 Performance of Our Fine-Tuned Model

The predicted grasp poses were visualized and are shown in Figure 6, demonstrating the fine-tuned model's ability to generate feasible and collision-free grasp predictions. The model consistently produced grasp poses within the annotated regions for all tools, ensuring that the predictions were both accurate and aligned with safety requirements. The next section presents the results of the experiments, evaluating the predicted grasp poses in successfully grasping the objects.

The predicted grasp scores for each tool are summarized in Table 3. The success rates of the two experiments are detailed in Tables 4 and 5. For the highest-scoring grasp poses, all tools achieved a success rate of 100%, indicating that the proposed dataset and the trained model can generate highly reliable grasp poses. For the top 10 scoring grasp poses, the success rates were 100%, 80%, 90%, and 100% for the screwdriver, adjustable wrench, pliers, and safety glasses, respectively, with an overall average success rate of 92.5%. These results demonstrate the effectiveness of the proposed dataset in training the model to predict accurate grasp poses.

Despite the overall high success rates, some failure cases



(a) screwdriver (b) wrench

(c) safety glasses (d) plier

Figure 6. Qualitative results of the predicted grasp poses (top 10 highest-scoring grasps)

Table 3. Predicted grasp scores from fine-tuned model

Object	Highest	Top 10 average
Screwdriver	0.446	0.241
Adjustable wrench	0.680	0.467
Plier	0.385	0.376
Safety glasses	0.665	0.550

Table 4. Success rate of grasps with the highest grasp scores (fine-tuned model)

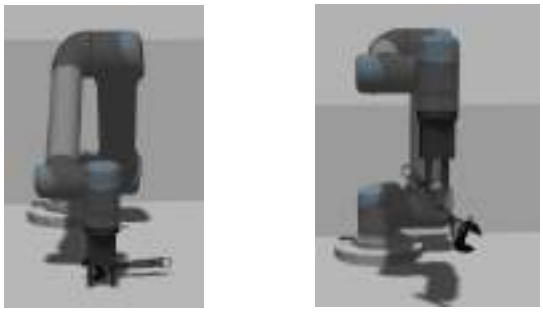
Object	Success rate
Screwdriver	100%
Adjustable wrench	100%
Plier	100%
Safety glasses	100%

Table 5. Success rate of grasps with top 10 grasp scores (fine-tuned model)

Object	Success rate
Screwdriver	100%
Adjustable wrench	80%
Plier	90%
Safety glasses	100%

were observed, particularly with the adjustable wrench, which achieved a success rate of 80% in the top 10 grasp experiments. Two failure cases occurred: one involved a collision between the robotic arm and the wrench during the grasping process, while the other involved slippage from the gripper. Figure 7 illustrates these two failure cases, with the left image showing a collision and the right image showing slippage.

The primary cause of these failures is the wrench's low height (1.4 cm) and flat geometry, which makes accurate grasp planning challenging. When the robotic arm executes the motion planning for grasping, the reduced thickness of the wrench can lead to misalignment, resulting in slippage or collisions. Similarly, the pliers slipped when lower-scoring grasp poses were tested. To further mitigate collisions and slippage, we plan to adopt a more rigorous collision detection strategy that discards grasp poses likely to cause interference. During the handover process, we will integrate reinforcement learning to optimize grasp stability and minimize the human operator's joint torque and displacement. By enhancing the robot's ability to adaptively maintain a secure grip while simultaneously reducing strain on the human arm, we aim to improve both safety and convenience in the tool handover process.



(a) collision (b) slippage

Figure 7. Failure cases of robotic arm grasping wrench

The computational efficiency of the approach is a critical factor for practical applications. The inference process requires an average of 1.6 seconds, while the motion planning process, including moving to the object, executing a stable grasp, and returning the arm to its home position takes 10.4 seconds. Although the current motion planning process takes longer to test the grasp stability. Integrating a reinforcement learning strategy during the handover phase will help reduce the overall planning time and enhance the system's for real-world deployments.

Overall, the experimental results confirm that the proposed dataset enables the GraspNet-Baseline model to predict reliable grasp poses. Despite occasional failures

with specific tools like the adjustable wrench, the dataset demonstrated its ability to adapt to diverse tool shapes and sizes. These results validate the applicability of the dataset and model for safe and accurate robotic grasping in construction tool handover scenarios.

6 Conclusion and Future Work

In this paper, we proposed a method for robot-to-human hand tool handover grasp prediction and developed a dataset that includes four types of handheld tools commonly used in the construction industry. The dataset consists of multi-angle RGB-D images and detailed annotations. During the annotation process, we introduced a method to distinguish between human and robotic grasping regions, ensuring that the robotic arm predicts grasp poses concentrated on the potentially dangerous end of the tool. This approach minimizes safety risks during human handling by keeping the safer end of the tool accessible. Furthermore, we trained the dataset using the GraspNet-baseline model and validated the predicted grasp poses in a simulation environment. The experimental results demonstrated that our dataset aligns well with grasping requirements.

Despite the promising results, there are several limitations. First, the study is based on simulation experiments with a relatively limited dataset. The dataset doesn't capture the diversity of tools in real-world construction sites (e.g., variations in covers, handles, and materials), which may affect the model's generalizability. We plan to expand the dataset by incorporating a wider range of tool types and shapes, and to conduct real-world experiments to validate the transferability.

Finally, while the study focuses solely on grasp prediction, we acknowledge that human-robot interaction factors such as grip force control and handover timing are critical for ensuring safe tool handover in dynamic construction environments. In our future work, we plan to integrate sensor-based feedback (e.g., tactile sensors for grip force and vision-based human motion tracking) and adaptive control algorithms to dynamically adjust the handover process. This approach will not only enhance the grasp prediction module but will also optimize the overall handover process, providing a more robust and practical solution for real-world applications.

Overall, by addressing these limitations, our work aims to advance the state-of-the-art in robotic tool handover and bring the approach closer to practical deployment in real-world human-robot collaborative tasks.

References

[1] Mani Golparvar-Fard, Feniosky Peña-Mora, and Silvio Savarese. Automated progress monitoring using unordered daily construction photographs and ifc-based building information models. *Journal of Computing in Civil Engineering*, 29(1):04014025, 2015. doi:10.1061/(ASCE)CP.1943-5487.0000205.

[2] Karianne Berntsen. Did you know construction workers spend 38 hours annually looking for their tools? On-line: https://www.abax.com, June 2020. Accessed: 01/12/2025.

[3] Nathan Melenbrink, Justin Werfel, and Achim Menges. On-site autonomous construction robots: Towards unsupervised building. *Automation in Construction*, 119:103312, 2020. doi:https://doi.org/10.1016/j.autcon.2020.103312.

[4] Bo Xiao, Chen Chen, and Xianfei Yin. Recent advancements of robotics in construction. *Automation in Construction*, 144:104591, 2022. ISSN 0926-5805. doi:https://doi.org/10.1016/j.autcon.2022.104591.

[5] Valerio Ortenzi, Akansel Cosgun, Tommaso Pardi, Wesley P. Chan, Elizabeth Croft, and Dana Kulić. Object handovers: A review for robotics. *IEEE Transactions on Robotics*, 37(6):1855–1873, 2021. doi:10.1109/TRO.2021.3075365.

[6] Ming Zhang, Rui Xu, Haitao Wu, Jia Pan, and Xiaowei Luo. Human–robot collaboration for on-site construction. *Automation in Construction*, 150:104812, 2023. ISSN 0926-5805. doi:https://doi.org/10.1016/j.autcon.2023.104812.

[7] Ci-Jyun Liang, Xi Wang, Vineet R. Kamat, and Carol C. Menassa. Human–robot collaboration in construction: Classification and research trends. *Journal of Construction Engineering and Management*, 147(10):03121006, 2021. doi:10.1061/(ASCE)CO.1943-7862.0002154.

[8] Hao-Shu Fang, Chenxi Wang, Minghao Gou, and Cewu Lu. Graspnet-1billion: A large-scale benchmark for general object grasping. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11441–11450, 2020. doi:10.1109/CVPR42600.2020.01146.

[9] A.T. Miller, S. Knoop, H.I. Christensen, and P.K. Allen. Automatic grasp planning using shape primitives. In *2003 IEEE International Conference on Robotics and Automation (Cat. No.03CH37422)*, volume 2, pages 1824–1829 vol.2, 2003. doi:10.1109/ROBOT.2003.1241860.

[10] Joseph Redmon and Anelia Angelova. Real-time grasp detection using convolutional neural networks.

In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1316–1322, 2015. doi:10.1109/ICRA.2015.7139361.

[11] Sulabh Kumra and Christopher Kanan. Robotic grasp detection using deep convolutional neural networks. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 769–776, 2017. doi:10.1109/IROS.2017.8202237.

[12] Edward Johns, Stefan Leutenegger, and Andrew J. Davison. Deep learning a grasp function for grasping under gripper pose uncertainty. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4461–4468, 2016. doi:10.1109/IROS.2016.7759657.

[13] Abhinav K. Keshari, Hanwen Ren, and Ahmed H. Qureshi. Cograsp: 6-dof grasp generation for human-robot collaboration. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 9829–9836, 2023. doi:10.1109/ICRA48891.2023.10160623.

[14] Chongxi Meng, Tianwei Zhang, and Tin lun Lam. Fast and comfortable interactive robot-to-human object handover. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3701–3706, 2022. doi:10.1109/IROS47612.2022.9981484.

[15] Zixi Wang, Zeyi Liu, Nicolas Ouporov, and Shuran Song. Contacthandover: Contact-guided robot-to-human object handover. On-line: `https://arxiv.org/abs/2404.01402`, Accessed: 04/2024.

[16] Hongrui Yu, Vineet R. Kamat, Carol C. Menassa, Wes McGee, Yijie Guo, and Honglak Lee. Mutual physical state-aware object handover in full-contact collaborative human-robot construction work. *Automation in Construction*, 150:104829, 2023. ISSN 0926-5805. doi:https://doi.org/10.1016/j.autcon.2023.104829.

[17] Berk Calli, Arjun Singh, James Bruce, Aaron Walsman, Kurt Konolige, Siddhartha Srinivasa, Pieter Abbeel, and Aaron M Dollar. Yale-cmu-berkeley dataset for robotic manipulation research. *The International Journal of Robotics Research*, 36(3):261–268, 2017. doi:10.1177/0278364917700714.

[18] Amaury Depierre, Emmanuel Dellandréa, and Liming Chen. Jacquard: A large scale dataset for robotic grasp detection. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3511–3516, 2018. doi:10.1109/IROS.2018.8593950.

[19] Yun Jiang, Stephen Moseson, and Ashutosh Saxena. Efficient grasping from rgbd images: Learning using a new rectangle representation. In *2011 IEEE International Conference on Robotics and Automation*, pages 3304–3311, 2011. doi:10.1109/ICRA.2011.5980145.

Automated Decision-Making Tool for Optimal Long-Term Scheduling of MRR Strategies: A Case Study on Bridges

Mohammed Alsharqawi[1], Saleh Abu Dabous[2], and Tarek Zayed[3]

[1]Department of Construction Management, California State University Sacramento, United States of America
[2]Department of Civil and Environmental Engineering, University of Sharjah, United Arab Emirates
[3]Department of Building and Real Estate, The Hong Kong Polytechnic University, Hong Kong
alsharqawi@csus.edu, sabudabous@sharjah.ac.ae, tarek.zayed@polyu.edu.hk

Abstract – Maintenance, repair, and replacement (MRR) strategies for bridges are critical requirements for infrastructure management, demanding effective decision-making to balance cost and performance objectives. This research presents an automated decision-making tool designed to optimize long-term scheduling of MRR strategies, addressing the complexities of multi-objective problems in bridge management. The tool integrates performance modeling with a modified genetic algorithm to generate optimal MRR plans, ensuring efficient resource allocation while maintaining structural integrity. A case study demonstrates the tool's application, predicting bridge deterioration and proposing cost-effective MRR actions under performance-based contracting frameworks. Results highlight significant cost savings and enhanced bridge serviceability through timely interventions. While the tool currently focuses on bridge decks, it establishes a foundation for future enhancements to include additional bridge components and broader infrastructure networks. This study advances automated MRR scheduling by offering a robust solution for sustainable bridge management.

Keywords – Automation; Decision-Making; Maintenance, Repair, and Replacement (MRR); Optimal MRR Strategies; Long-Term Scheduling; Modified Genetic Algorithm (GA); Sustainable Bridge Asset Management; Infrastructure Management

1 Introduction

The optimization of bridge maintenance, repair, and replacement (MRR) strategies has gained substantial attention in infrastructure management, especially with the introduction of automated decision-making tools that enhance scheduling efficiency. These approaches utilize various mathematical models, algorithms, and data-driven techniques to assess the optimal timing and scope of MRR interventions for bridges. Predictive algorithms further enhance decision-making by identifying optimal MRR strategies based on projected deterioration. The study by Bolukbasi et al. [1] presents an optimization model that employs deterioration curves to forecast the degradation of bridge components, allowing for more timely and effective scheduling of rehabilitation activities. In a similar vein, Alsharqawi et al. [2] introduced an integrated condition rating and forecasting methodology to provide accurate condition assessments for bridge decks. This dual-method approach allows for a refined understanding of bridge deck health, contributing to effective long-term scheduling and resource allocation for MRR strategies. Furthermore, Neves et al. [3] proposed a probabilistic optimization model that combines preventive and essential maintenance actions, underscoring the value of a hybrid approach in optimizing bridge maintenance costs while maintaining safety standards.

Implementing performance-based contracting (PBC) has emerged as a valuable approach to bridge maintenance, focusing on setting specific performance metrics and outcomes rather than prescribing maintenance actions. This approach incentivizes contractors to meet defined service levels, often resulting in more efficient use of resources and improved bridge performance over time. Alsharqawi et al. [4] highlight the advantages of PBC in maintaining transportation assets, particularly bridges, by setting outcome-focused targets that contractors must meet to fulfill their obligations. Similarly, performance-based approaches in bridge maintenance have been shown to support innovative solutions, including the use of advanced materials and real-time monitoring systems, which help contractors maintain the infrastructure within specified parameters [5]. The use of PBC aligns with the goals of long-term scheduling in bridge maintenance as it provides structured, measurable goals that guide the maintenance schedule, allowing for optimized resource allocation and enhanced reliability of bridge infrastructure over its lifecycle [6].

2 Literature Review

Automated decision-making tools and frameworks play a crucial role in optimizing the long-term scheduling of MRR for infrastructure, especially for bridge networks. These tools address complex, multi-objective problems involving cost, reliability, traffic flow disruption, and network performance, among others. Multi-objective models in bridge maintenance scheduling often aim to minimize both maintenance costs and traffic delays. For instance, Mao et al. proposed a bi-level programming model that minimizes traffic delays while maximizing the number of bridges maintained within budget constraints. By using a simulated annealing algorithm, the model has shown effectiveness in reducing maintenance costs and traffic disruptions [7]. Another approach, developed by Bocchini and Frangopol, introduces a probabilistic framework utilizing multi-objective genetic algorithms for preventive maintenance scheduling across a highway bridge network. Their model optimizes between cost and network performance, offering bridge managers a Pareto front of solutions to balance engineering and economic considerations [8]. Montazeri and Touran propose a reliability-based decision-making framework, emphasizing the importance of life-cycle performance and maintenance costs. Their model aids bridge managers in evaluating the best repair and replacement strategies across a bridge's lifespan, considering structural specifications and budget limitations [9]. Computational models help optimize bridge maintenance schedules under uncertainty. Liu and Frangopol developed a probabilistic maintenance planning model using Monte Carlo simulations to address uncertainties in bridge deterioration, which improves prediction accuracy and provides confidence-based maintenance solutions [10]. For network-level optimization, Orcesi and Cremona introduced a genetic algorithm-based model that considers the bridge's position within a transportation network, which ensures optimal resource distribution and maintains the level of service across the network [11].

An important aspect of automated MRR tools involves balancing cost and reliability. Frangopol et al. demonstrated how integrating reliability-oriented models with cost analysis could enhance bridge maintenance planning by identifying economically viable MRR schedules that ensure safety over the structure's lifetime [12]. This approach aligns with the work of Liang et al., who developed decision criteria based on cost-benefit theory to maintain bridge deck pavement, underscoring the importance of a balanced approach in prioritizing both safety and budget considerations [13]. Lee et al. developed a preference-based maintenance scheduling method tailored for bridges in Korea, balancing life-cycle costs with bridge performance. Their multi-objective optimization model allows for stakeholder preferences, helping decision-makers choose cost-effective maintenance strategies [14]. The integration of innovative decision support model under performance-based contract settings has advanced MRR scheduling significantly. Alsharqawi et al. developed a model capable for budget optimizing of MRR decisions under PBC settings. Implementing the model enables contractors to execute various MRR plans under different budget and performance requirements, where each resulting plan is optimal along the contract period [15], [16].

In recent years, the complexity of bridge MRR scheduling has underscored the need for advanced automated tools that leverage algorithmic optimization to streamline decision-making. Traditional approaches to MRR scheduling are often constrained by manual assessments and simplistic cost-benefit analyses, which may fall short in capturing the dynamic and multi-faceted nature of bridge deterioration and maintenance requirements. An automated, algorithm-based tool can enhance this process by rapidly processing large datasets, predicting deterioration with high accuracy, and generating optimized schedules that minimize maintenance costs while maximizing bridge reliability and performance. This data-driven approach to MRR scheduling is essential for infrastructure managers seeking to allocate resources effectively, reduce traffic disruptions, and ensure the long-term safety and functionality of bridge networks.

3 Development of Automated Tool

The automated decision-making support tool inputs, tools, techniques, and outputs are illustrated in Figure 1 and discussed in the following sections. The proposed tool in this paper considers bridge deck in particular because it deteriorates faster than other components due to its direct exposure to traffic in addition to the external environmental factors (e.g., freeze-thaw cycling) [17], [18].

The development of the automated tool involves two key components: first, modeling the performance of the bridge deck, and second, optimizing the selection of MRR strategies.

3.1 Performance Modeling

3.1.1 Performance Evaluation

To evaluate current performance, assessing the condition of the bridge deck is essential. This assessment provides insights into the physical state of the infrastructure and an estimate of its remaining useful life

or maximum lifespan. Regular detailed inspections are conducted to identify significant defects and evaluate the extent of bridge deterioration. In North America, visual inspection and close observation of bridge elements are commonly used due to their low cost and minimal expertise requirements. If severe damage is identified during visual inspection, a more comprehensive condition survey is conducted using both destructive and non-destructive testing methods [2].

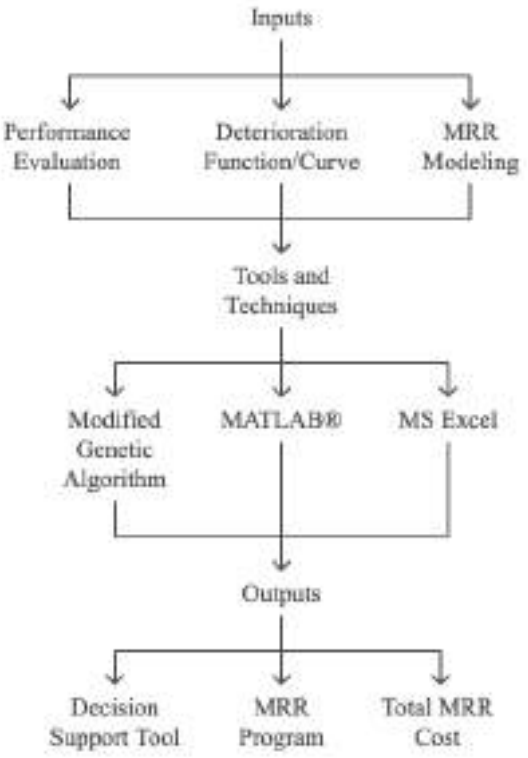


Figure 1. Development of the decision-making support tool inputs, tools, techniques, and outputs

3.1.2 Deterioration Function

To estimate element degradation over time, probabilistic or stochastic models are used, given that bridge deterioration follows a stochastic, rather than deterministic, process. Among these models, the Weibull distribution function (WDF) is frequently employed due to its advantages including: (i) it has been proven its effectiveness in representing concrete deterioration, and (ii) its parameters are straightforward to calculate [2]. Future performance rates are predicted based on the most recent inspection or condition assessment, providing a realistic representation of deterioration over time using Equation 1:

$$P(t) = 1 \times e^{ln(P_i)(t/t_i)^3} \quad (1)$$

where: P = component performance at time t; P_i = performance at time t_i; and t_i = inspection time.

3.1.3 MRR Modeling

MRR modeling involves identifying appropriate MRR strategies, estimating the resulting performance improvements, and calculating the associated costs. These MRR strategies are categorized into three types based on the bridge deck's performance (Table 1). For bridge decks in good condition, routine maintenance can be performed. Bridge decks with poor performance may require repair actions, such as increasing deck thickness through localized concrete repairs. For decks with critical performance levels, full deck replacement is recommended.

Modeling performance improvement and predicting the deterioration behavior of an asset after a MRR action is essential. The improvement values associated with each type of MRR relied on a study that was performed by Seo [19], as shown in Table 1. According to the literature review, performance is estimated to improve by 20% when repair actions are implemented to the bridge deck. In the case of bridge deck replacement, performance increases to approximately 90%. For maintenance actions, no performance improvement occurs, but this action slows the deterioration rate of the bridge deck, resulting in a gradual decline over time.

Calculating MRR costs is typically performed either as a fixed percentage of the initial construction or replacement cost, or as a unit cost. For instance, Seo [19] proposed predefined percentage-based estimates for different types of MRR, while Abu Dabous and Alkass [20] developed unit cost models for each MRR type. Each MRR action's unit cost includes direct costs, indirect costs, and a markup to cover contractor profit and contingency. Table 1 summarizes MRR action costs as a percentage of initial construction costs [19] and as fixed unit costs [20] , with unit costs adjusted to 2024 Canadian dollars for inflation.

Table 1. MRR strategies, improvements, and costs per strategy

MRR Strategy	Improvement	MRR Cost	
		% of initial construction	unit cost ($/m²)
Maintenance	0%	28.5%	$58.39
Repair	20%	65.0%	$1,098.89
Replacement	90%	100.0%	$1,438.72

3.2 Optimizing MRR Strategies

3.2.1 Long-Term Scheduling

As discussed earlier, integrating innovative PBC frameworks within a decision-making support methodology has significantly advanced MRR scheduling. There is a strong shift towards PBCs over traditional contracts for infrastructure management, including bridges, primarily due to cost savings. Additional advantages noted in the literature [21] include but not limited to more integrated services, improved asset management, and economies of scale. This approach incentivizes contractors to achieve specified service levels, often leading to more efficient resource use and enhanced bridge performance over time. Furthermore, the use of PBCs aligns with long-term bridge maintenance goals by providing structured, measurable objectives that guide the maintenance schedule, optimize resource allocation, and improve bridge reliability over its lifecycle [6]. Consequently, the proposed automated tool incorporates PBC settings.

3.2.2 Modified Genetic Algorithm

Since genetic algorithms (GAs) can provide a comparable level of accuracy in the area of global optimization while being more time efficient than conventional optimization techniques [22], they have been extensively used by many researchers in maintenance optimization problems for bridge management. Recently, Ghodoosi et al. [23], Xie et al. [24], Allah Bukhsh et al. [25], and Alsharqawi et al. [15] proposed models to find optimal maintenance and rehabilitation strategies using genetic algorithms. As a modification to the well-known GAs, several works started to design customized variants of the GA. These customized variants use multiple chromosomes with real numbers that represent the different parts of the solution for each individual rather than one chromosome. In addition, the GA operators are specially designed to fit the problem context. The modified GA starts with feasible individuals satisfying all constraints, which makes it easier to search for the optimal solution compared to the case of the original GA, where the feasibility of individuals might not be satisfied. It is worth noting that in this type of modified GA, a certain number of operators is defined and applied in all iterations; thus, the probabilistic nature of changing the individuals is removed, which reduces the number of parameters needed to be tuned. In this research, a modified GA is used due to its quick convergence and excellent performance, as in this modified GA, for each individual, a chromosome defines the potential MRR actions over the planning years. The chromosomes are generated in a way to satisfy the predefined performance thresholds. The population is replaced with a new random one in case of no improvement over several generations. The modified GA code is developed using MATLAB® software language and optimization toolbox.

In order to solve the MRR optimization problem, this algorithm starts by generating V individuals, each individual ($v \in V$) contains a random MRR strategy for each n year ($n \in N$). It sets the global total MRR cost to $+\infty$ and the global average performance to $-\infty$. In each iteration δ ($\delta \in \varphi$), the algorithm performs the following steps:

1. It calculates the yearly performance for each individual (p_v^n).
2. Then, it checks in each year if the yearly performance is within the limits and modifies the selected MRR strategy randomly as an attempt to fix any performance violation.
3. It calculates the average performance for each individual (P_v).
4. It calculates for each individual v the total MRR cost.
5. It identifies the local minimum total MRR cost and the local maximum average performance and uses them to update the global total MRR cost to $+\infty$ and the global average performance.

It is worth noting that the modified GA differs than the original GA in the following aspects: i) the generated population is feasible prior applying any operator; ii) the entire population is discarded if the optimal solution did not improve for iteration and a new population is generated randomly to allow for more diversity in the searching domain; and iii) the probability of crossover and mutation are eliminated. Crossover and mutation operators are applied in a different manner. In each iteration, the population is divided into subsets. Each subset has five individuals. The five individuals in the subset are ranked based on their fitness value. The best individual in the subset is selected and the five operators are applied on it. It should be noted that the first operator is do nothing which keeps the individual unchanged. Therefore, each subset of five individuals is replaced with a new subset containing the best individual in the subset and four new individuals.

3.2.3 Optimized MRR Program

As previously mentioned, bridge MRR actions are typically categorized into maintenance {2}, repair {3}, replacement {4} and no action {1} which is a do-nothing strategy. Specific MRR actions can be found under each category. With these four major categories, several combinations of actions can be generated over the analysis period producing large number of strategies or possible solutions. The longer the analysis period, the larger the search space is because the number of possible solutions equals to the number of MRR actions raised to

the power of number of years (i.e., analysis or contract period). For example, if the analysis period is ten years and there are four different MRR actions, the search space will be 1,048,576 (4^{10}). Accordingly, the modified GA-based optimization technique, as a non-linear mathematical modeling, is used to obtain an optimal or near-optimal solution from this large list of solution while meeting the needs and constraints related to the PBC settings. This approach can consider all possibilities and produce enhanced and robust solutions compared to the existing tools and simplified matrices typically used in practice.

The output from the automated tool is an optimized MRR program in which the MRR actions should be executed are determined, the performance prediction in each year is forecasted, and the calculations of the net present value (NPV) during the specified years of MRR plan is estimated. Furthermore, the deterioration curve forecasts the performance for the same duration. The performance of the bridge deck before and after MRR is computed, and performance values are determined based on the Weibull distribution function discussed earlier. The performance improvement is defined based on the type of the applied MRR action.

4 Application of the Tool

4.1.1 Description and Assumptions

A case study is discussed here as a proof of concept and illustration of the presented decision-making support model. The model is applied to a bridge located in Québec, Canada. Names and exact locations of the bridge are not given for confidentiality purposes as per agreements with data providers. An optimization scenario was carried out over the initial remaining life-cycle time of the bridge (till year 2040). The implementation of PBCs over periods of 10 years or more has proven to be an effective strategy for the sustained preservation of transportation networks. Additionally, the Australian experience with the 10-year contract in New South Wales (NSW) demonstrates that realistic PBC criteria and targets are both achievable and effective on a network-wide scale. For the MRR plan, a budget equals to $225,000 is set allowing for a possibility to cover the cost of replacing the bridge deck.

To run the tool, users must input the condition rating, construction year, year of the last inspection from the latest inspection report, and current year. The tool automatically calculates the current performance, utilizing its embedded deterioration function. Additional input variables are required to execute the decision-making optimization process. These variables include the duration of the MRR plan, budget constraints, interest rate (i=4% is used in this study), and performance thresholds. Moreover, MRR costs are incorporated based on the MRR modeling formulation, as showed in Table 1. Table 2 summarizes the case study data attributes. The subsequent section will present a set of trade-off list detailing MRR actions applied.

Table 2. Case study data attributes

Bridge Attribute	Input
Construction Year	1965
Age	58
Construction Cost	N/A
Area (square meter)	152
Inspection Year	2021
Performance at Inspection Year	63.99%

4.1.2 Results and Discussion

The 2021 inspection assessed the bridge deck's physical condition at 63.99%, indicating poor performance during that year. Using performance modeling, deterioration curves were mapped, starting with the ideal deterioration curve (representing ideal performance), knowing that the bridge was constructed back in 1965. To create the deterioration curve (performance without intervention), data from the 2021 inspection survey was used, and the curve was plotted using Eq. 1, with t_i corresponding to 2021. Implementing the automated tool on the case study allowed for the reconstruction of the updated deterioration curve (performance with MRR intervention).

The implementation of the modified GA-based optimization technique presented earlier provides a variation of MRR strategies applied at various years throughout the time horizon of the described scenario. Figure 2 shows the bridge deck performance without and with MRR interventions after applying the developed tool. The present worth of the optimized scenario for the case study is also calculated. The total MRR actions costs equal to $1,686,557.11. Table 3 presents the optimal MRR program developed from applying this tool where the MRR action and its corresponding cost are shown. In this case, it is suggested that the contracting agency tendering the contract can allocate a budget of no more than the total MRR actions costs. The budget assignment is based on the developed decision-making support tool which minimizing the total cost while maximizing the performance.

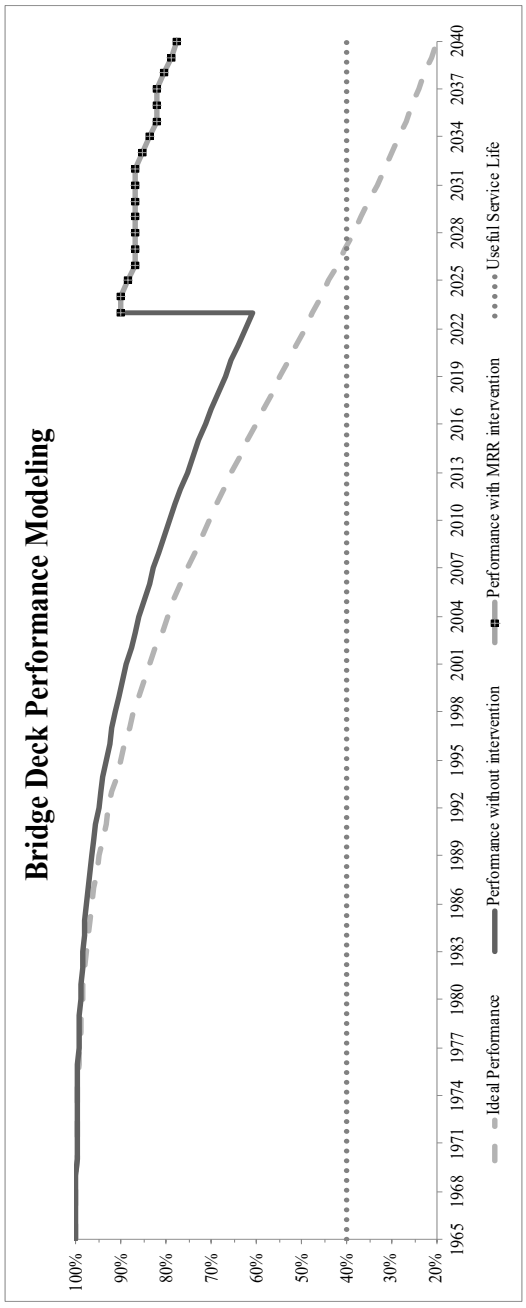


Figure 2. Case study performance modeling, including curves without and with MRR interventions

Based on the performance monitoring data, actual performance may vary compared to the predicted from the deterioration curve. Using real time data after performing inspections, the MRR program may be updated by running the tool again. The significance of this is that re-optimization of the MRR program using real time data and continuous validation of deterioration curve may result in cost saving and higher profit to the contractor in case of overestimating the deterioration rate and consequently scheduling the MRR action earlier than needed.

Table 3. Case study optimal MRR program

Year	MRR Action	MRR Cost	Total MRR Cost
0			$ 1,686,557.11
1	{4}	$ 210,274.46	
2	{2}	$ 4,591.20	
3	{3}	$ 148,490.20	
4	{3}	$ 142,779.04	
5	{4}	$ 179,743.49	
6	{3}	$ 132,007.25	
7	{3}	$ 126,930.05	
8	{3}	$ 122,048.12	
9	{3}	$ 117,353.96	
10	{3}	$ 112,840.35	
11	{1}	$ -	
12	{3}	$ 104,327.24	
13	{3}	$ 100,314.66	
14	{3}	$ 96,456.40	
15	{1}	$ -	
16	{2}	$ 2,651.30	
17	{3}	$ 85,749.39	
18	{1}	$ -	

5 Conclusion

The majority of existing decision-making methodologies for bridge management aim to balance the dual objectives of maintaining or extending bridge serviceability while minimizing total costs. These inherently conflicting goals significantly complicate the decision-making process. To address this bi-objective optimization challenge, this research has developed an automated decision-making tool for the long-term scheduling of MRR strategies. The convergence of probabilistic models and multi-objective optimization approaches in this study underscores the advancements in MRR planning, enhancing both efficiency and effectiveness. Further, research findings concluded that integrating the PBC approach into the MRR decision-making process offers better results in long-term plans. The developed tool not only achieves substantial cost savings but also prolongs the functional lifespan of bridges through timely and strategic maintenance interventions. By enabling bridge managers to tailor MRR strategies to various conditions and constraints, this tool promotes sustainable and efficient resource utilization while ensuring the serviceability and structural integrity of bridge infrastructure over time.

However, the research does have certain limitations: (1) the study primarily focused on bridge decks, without extending the tool's application to other bridge components or related assets; (2) the effects of MRR actions on bridge performance were restricted to maintenance, repair, and replacement strategies; (3) user and external costs (e.g., user delays) associated with MRR actions were excluded from the calculation of MRR unit costs, though these costs could be incorporated in future studies; and (4) the 10-year planning horizon was chosen to align with common performance-based contracting practices, yet it limits the direct comparison of predicted MRR plans against the actual long-term degradation of bridge decks.

To address uncertainties inherent in long-term forecasts, the tool supports periodic re-optimization with updated inspection data, and future work will incorporate extended monitoring and sensitivity analyses to further validate model predictions. Moreover, future research could expand the scope of the tool to encompass additional bridge elements and broader infrastructure networks, further enhancing its applicability and impact. Additionally, future work could integrate real-time sensor data from structural health monitoring systems to dynamically update the predefined deterioration models.

References

[1] M. M. Bolukbasi, D. Arditi, and J. Mohammadi, "Optimum rehabilitation schedule of steel highway bridges," *Structure and Infrastructure Engineering*, vol. 2, no. 2, pp. 141–152, Jun. 2006, doi: 10.1080/15732470500253297.

[2] M. Alsharqawi, T. Zayed, and S. Abu Dabous, "Integrated condition rating and forecasting method for bridge decks using Visual Inspection and Ground Penetrating Radar," *Automation in Construction*, vol. 89, pp. 135–145, May 2018, doi: 10.1016/j.autcon.2018.01.016.

[3] L. A. C. Neves, D. M. Frangopol, and A. Petcherdchoo, "Probabilistic Lifetime-Oriented Multiobjective Optimization of Bridge Maintenance: Combination of Maintenance Types," *J. Struct. Eng.*, vol. 132, no. 11, pp. 1821–1834, Nov. 2006, doi: 10.1061/(ASCE)0733-9445(2006)132:11(1821).

[4] M. Alsharqawi, S. Abu Dabous, and T. Zayed, "Performance-based contracting for maintaining transportation assets with emphasis on bridges," *MATEC Web Conf.*, vol. 120, p. 08012, 2017, doi: 10.1051/matecconf/201712008012.

[5] J. K. Lindly, and D. E. Hancher, "Performance-based maintenance contracting," *Transportation Research Record: Journal of the Transportation Research Board*, pp. 219–225, 2003.

[6] S. T. Ng and Y. M. Wong, "Performance-based payment scheme for public road maintenance," *Journal of Construction Engineering and Management*, vol. 132, no. 4, pp. 364–373, 2006.

[7] X. Mao, X. Jiang, C. Yuan, and J. Zhou, "Modeling the Optimal Maintenance Scheduling Strategy for Bridge Networks," *Applied Sciences*, vol. 10, no. 2, p. 498, Jan. 2020, doi: 10.3390/app10020498.

[8] P. Bocchini and D. M. Frangopol, "A probabilistic computational framework for bridge network optimal maintenance scheduling," *Reliability Engineering & System Safety*, vol. 96, no. 2, pp. 332–349, Feb. 2011, doi: 10.1016/j.ress.2010.09.001.

[9] N. Montazeri and A. Touran, "Applied Decision-Making Framework for Maintenance Scheduling in Bridge Management," in *Proceedings of the Creative Construction Conference 2019*, Budapest University of Technology and Economics, 2019, pp. 547–555. doi: 10.3311/CCC2019-075.

[10] M. Liu and D. M. Frangopol, "Optimal bridge maintenance planning based on probabilistic performance prediction," *Engineering Structures*, vol. 26, no. 7, pp. 991–1002, Jun. 2004, doi: 10.1016/j.engstruct.2004.03.003.

[11] A. D. Orcesi and C. F. Cremona, "A bridge network maintenance framework for Pareto optimization of stakeholders/users costs," *Reliability Engineering & System Safety*, vol. 95, no. 11, pp. 1230–1243, Nov. 2010, doi: 10.1016/j.ress.2010.06.013.

[12] D. M. Frangopol, M. Miyake, J. S. Kong, and E. S. Gharaibeh, "Reliability- and Cost-Oriented Optimal Bridge Maintenance Planning," in *Advanced Technology in Structural Engineering*, Philadelphia, Pennsylvania, United States: American Society of Civil Engineers, Apr. 2000, pp. 1–8. doi: 10.1061/40492(2000)21.

[13] X. Liang, J. Shi, and Z. Li, "Decision Method of Bridge Deck Pavement Maintenance Based on the Decision Maker's Objective," *Advances in Civil Engineering*, vol. 2021, no. 1, p. 5593722, Jan. 2021, doi: 10.1155/2021/5593722.

[14] S.-Y. Lee, W. Park, S.-Y. Ok, and H.-M. Koh, "Preference-based maintenance planning for deteriorating bridges under multi-objective optimisation framework," *Structure and Infrastructure Engineering*, vol. 7, no. 7–8, pp. 633–644, Jul. 2011, doi: 10.1080/15732479.2010.501565.

[15] M. Alsharqawi, S. Abu Dabous, T. Zayed, and S. Hamdan, "Budget Optimization of Concrete

Bridge Decks under Performance-Based Contract Settings," *J. Constr. Eng. Manage.*, vol. 147, no. 6, p. 04021040, Jun. 2021, doi: 10.1061/(ASCE)CO.1943-7862.0002043.

[16] M. Alsharqawi, S. Abu Dabous, and T. Zayed, "Designing a fuzzy-based framework for implementing performance-based contracts in bridge asset management," *Innov. Infrastruct. Solut.*, vol. 9, no. 4, p. 126, Apr. 2024, doi: 10.1007/s41062-024-01437-1.

[17] N. Gucunski *et al.*, *Nondestructive Testing to Identify Concrete Bridge Deck Deterioration*. Washington, D.C.: Transportation Research Board, 2012, p. 22771. doi: 10.17226/22771.

[18] N. Diamanti, A. P. Annan, and J. D. Redman, "Concrete Bridge Deck Deterioration Assessment Using Ground Penetrating Radar (GPR)," *JEEG*, vol. 22, no. 2, pp. 121–132, Jun. 2017, doi: 10.2113/JEEG22.2.121.

[19] J. H. Seo, "A Lagrangean Relaxation and Network Approach to Large-Scale Optimization for Bridge Management System," Doctor of Philosophy, Texas A&M University, Texas, United States, 1994.

[20] S. Abu Dabous and S. Alkass, "Managing bridge infrastructure under budget constraints: a decision support methodology," *Can. J. Civ. Eng.*, vol. 38, no. 11, pp. 1227–1237, Nov. 2011, doi: 10.1139/l11-082.

[21] W. A. Hyman, *Performance-Based Contracting for Maintenance*. Washington, D.C.: Transportation Research Board, 2009, p. 14266. doi: 10.17226/14266.

[22] G. Morcous and Z. Lounis, "Maintenance optimization of infrastructure networks using genetic algorithms," *Automation in Construction*, vol. 14, no. 1, pp. 129–142, Jan. 2005, doi: 10.1016/j.autcon.2004.08.014.

[23] F. Ghodoosi, S. Abu-Samra, M. Zeynalian, and T. Zayed, "Maintenance Cost Optimization for Bridge Structures Using System Reliability Analysis and Genetic Algorithms," *J. Constr. Eng. Manage.*, vol. 144, no. 2, p. 04017116, Feb. 2018, doi: 10.1061/(ASCE)CO.1943-7862.0001435.

[24] H.-B. Xie, W.-J. Wu, and Y.-F. Wang, "Life-time reliability based optimization of bridge maintenance strategy considering LCA and LCC," *Journal of Cleaner Production*, vol. 176, pp. 36–45, Mar. 2018, doi: 10.1016/j.jclepro.2017.12.123.

[25] Z. Allah Bukhsh, I. Stipanovic, and A. G. Doree, "Multi-year maintenance planning framework using multi-attribute utility theory and genetic algorithms," *Eur. Transp. Res. Rev.*, vol. 12, no. 1, p. 3, Dec. 2020, doi: 10.1186/s12544-019-0388-y.

Integration of a real-time orientation measurement system for a real-time evaluator (RTE) to measure the position and orientation of crane-lifted components

Balaji Selvakumar[1], Nolan W. Hayes[1], Yifang Liu[1], Bryan P. Maldonado[1], Mengjia Tang[1], and Diana Hun[1]

[1]Buildings and Transportation Science Division, Oak Ridge National Laboratory, United States of America *

selvakumarb@ornl.gov, hayesnw@ornl.gov, liuy5@ornl.gov, maldonadopbp@ornl.gov, tangm@ornl.gov, hunde@ornl.gov

Abstract -

Prefabrication of building components holds the potential to revolutionize the construction industry. Prefabrication consists of manufacturing building components, modules, and other elements in a factory to be shipped and installed on a construction site. Prefabricated components have been produced for various applications including precast concrete panels for new construction and exterior wall retrofits. The manufacturing process has seen much innovation in recent years; however, the installation process has seen minimal advancements. A real-time evaluator (RTE) was developed to reduce the installation cost of prefabricated components by reducing installation time, decreasing rework, and improving accuracy. The RTE uses off-the-shelf hardware and novel algorithms to assist erectors with component installation by measuring the real-time positions of connections and prefabricated components, providing installation guidance through a graphical user interface, and monitoring the accumulated installation errors. An overview of the RTE and the proposed workflow is presented. Previous on-site demonstrations provided valuable feedback from users on the potential areas for improvement of the system. One common request was real-time measurement of component orientation during lifting, a process that previously required that the component remain stationary while the laser tracker cycled through target prisms. This paper will present the incorporation and testing of a real-time orientation measurement system as it was implemented into the RTE, allowing for measurement of component orientation during movement.

Keywords -

prefabrication, installation, real-time, automation, accuracy, time

1 Introduction

Prefabrication of building components holds the potential to revolutionize the construction industry [1]. Prefabrication consists of manufacturing building components, modules, and other elements in a factory to be shipped and installed on a construction site [2]. Manufacturers can maximize production efficiency and better control quality by constructing elements within a controlled environment [3].

Modern surveying instruments, such as robotic total stations, expedite, automate, and improve the accuracy of building and land surveying. Modern instruments also include reality capture devices such as 3D laser scanners that can accurately measure the surrounding area. However, these devices are more commonly used to survey the as-built conditions after construction is completed. Recent research has developed algorithms to automatically generate as-built 3D models using laser scanning [4, 5, 6]. Methods like these are often utilized immediately after the components are placed and construction is completed. A more proactive use of the technology would be the active monitoring of the quality of construction compared to the design as it is assembled. By addressing errors as they occur in the construction process, the total construction time and cost can be reduced.

An installation tool that provides corrective guidance in real-time to reduce installation time, decrease rework, and reduce cost would revolutionize the construction industry. In other industries such as manufacturing, laser trackers are used to precisely assess and guide the assembly process. For example, in the aviation industry, three laser trackers measuring targets attached to a plane wing have been used to generate real-time guidance during attachment of the wing to the aircraft body [7]. Other machine vision techniques are used during the robotic assembly of cars [8]. Adaptation of these methods in construction applications could enable real-time evaluation of components during placement, providing guidance for users such that errors are addressed in a timely fashion.

To innovate the installation process with modern technologies in the construction industry, a real-time evaluator (RTE) was developed to reduce the installation cost of prefabricated components by reducing installation time, decreasing rework, and improving accuracy [9]. The function of the RTE is illustrated in Figure 1. The RTE uses developed software solutions with off-the-shelf hardware

*Notice: This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (https://www.energy.gov/doe-public-access-plan).

to assist erectors in completing a prefabricated component installation by measuring the real-time positions of connections and prefabricated components, providing installation guidance through a graphical user interface, and monitoring the accumulated installation errors. Previous on-site demonstrations provided valuable feedback from users on the potential areas for improvement of the system. One common request was real-time measurement of component orientation during lifting, a process that previously required that the component remain stationary while the laser tracker cycled through target prisms [10].

This paper will present the incorporation and testing of a real-time orientation measurement system as it was implemented into the RTE, allowing for measurement of component orientation during movement. The paper makes the following primary contributions to the body of knowledge:

- Laboratory experiments show that some off-the-shelf inertial measurement units (IMU) have a high-enough degree of orientation measurement accuracy to provide real-time orientation estimates to guide the crane-positioning of large-scale building components while still meeting typical tolerances. A laser tracker system is still required to meet positional accuracy requirements.
- Comparisons of measured changes in orientation of some off-the-shelf IMUs with advanced laser 6-degree-of-freedom measurement systems show good agreement even during highly dynamic movements.
- Laboratory experiments proved that a system integrating a laser tracker measuring a retroreflector prism position and an IMU measuring orientation can successfully track the position and pose of a rigid-body during typical construction crane-installation processes with an accuracy meeting the typical requirements for tolerances.

1.1 Related Work

This work aims to supplement the prior research work that only utilized a single laser tracker to (1) track a single target during movement with all motion assumed to be rigid body translation and (2) measure three retroreflector targets while stationary within a total time of 10 seconds to determine pose [10]. This paper presents an integrated system that uses (1) a laser tracker to measure the real-time position of a component based on the measured location of a retroreflector prism and (2) an inertial measurement unit (IMU) to measure real-time orientation of the component. The addition of the IMU enables real-time estimation of the pose of the measured component without having to stop motion and wait for approximately 10 seconds.

Prior literature review of work on crane lifting operations has determined that little research is conducted to track the position and orientation of the lifted object specifically for high-precision installation purposes [11]. Prior work has utilized Radio Frequency Identification (RFID), Global Positioning System (GPS), laser scanning, sensors, photogrammetry, and more to estimate the location and orientation of a crane payload. Most applications were for the purposes of path planning, blind lifting, and safety.

For example, previous work utilized GPS and RFID to monitor the position of the crane for safety purposes [12]. However, these systems were not accurate enough for the typical tolerances required for precise installation. Other work consisted of using vision-based systems to determine the 3D position and pose of a component from images [13]. However, the accuracy of the vision-based system depends on the distance per pixel which is affected by camera specifications and distance to target. For large construction applications, sight distances will likely be large (more than 10 meters), and therefore, the accuracy of the vision-based system to determine precise position and pose of the component during lifting may not be adequate without significant investment of multiple camera systems.

For RTE, the system uses an off-the-shelf robotic laser tracker, which many construction companies performing large-scale crane installation of components already have access to. The use of a laser tracker also ensures that positional accuracy of the component location can be measured within 3 mm which is the smallest typical tolerance for the installed position of a component. The addition of the IMU enables real-time estimates of component orientation during movement; however, it is not intended to completely replace the highly-accurate method of measuring component orientation through measurement of three retroreflector targets attached to the component determining pose of the rigid body.

2 Methodology

2.1 System Overview

The existing RTE system utilizes a robotic multi-station controlled by novel software including communication protocols to measure the locations of three reflectors attached at known distances from the corners of the object to be measured. To measure the orientation of the object, the laser tracker sequentially cycles between the three prisms, measuring their positions in 3D space. The time to cycle between all prisms takes roughly 10 seconds. This positional data is then used to establish a frame of reference, allowing for the calculation of the panel's position and orientation [10]. The measured orientation is compared to the target rotation matrix, which is derived from the digital twin model of the expected construction configuration. This difference determines the necessary translations and rotations required for the panel to align with its goal pose.



Figure 1. A real-time evaluator to expedite prefabricated component installation. It provides real-time guidance on installing prefabricated components. [9]

In the current setup for initialization, it is assumed that the panel remains stationary during the orientation measurement cycle, which takes approximately 10 seconds to complete because the total station must turn to, search for, and measure each of the three prisms. This first cycle provides the real-time position and orientation at that time frame. Once the first cycle is completed, the tracker locks on to one of the isolated prisms and updates the real-time translation values from the centroid of the panel assuming all movement is rigid body translation. While this system is highly accurate, it has a significant limitation: it can only update the panel's translation in real time when the panel is moving in 3D space. To calculate the rotation, the panel must be held stationary, and the laser tracker must complete a full cycle through the three prisms again, causing latency. This results in a slower update rate for rotation data and presents a bottleneck in dynamic scenarios where faster orientation updates are necessary. Because of this limitation in measurement speed, all motion is assumed to be rigid body translation while the panel is moving. Additionally, the assumption that the panel remains stationary during cycling through prisms is critical. In reality, if any movement occurs during measurement of the three targets, the calculated component orientation will not be accurate. To address these limitations, there is a need to integrate an Inertial Measurement Unit (IMU) into the system, which can provide approximate orientation estimates in real-time during the panel movement, improving the overall responsiveness of the system.

2.2 Proposed System

2.2.1 IMU Integration

In the proposed system, the IMU is integrated as a complementary sensor to address the limitation of the existing methodology, which requires the object to remain stationary for orientation calculation. The IMU provides approximate orientation updates while the object is in motion, significantly enhancing the system's responsiveness. Our system has two separate coordinate frames (described as RTE and IMU). In order to integrate the IMU's rotation matrix which is determined about its local body frame to the RTE, it is first aligned with the panel to establish the necessary transformation tree (identity matrix in our case). This alignment allows both systems to compute the rotation of the panel with respect to the system RTE frame. The goal pose for the panel is derived from the digital twin model. Once the complete transformation tree is established, the IMU is initialized in the RTE frame using orientation data obtained from the laser tracker and prisms. The algorithm calculates the panel's orientation and reports the difference between its current pose and the goal pose (from the digital twin). This initialization step ensures the IMU's measurements are accurately calibrated

and aligned with the reference frame of the system.

The IMU's orientation estimates are fused with the laser tracker's translation data to create a hybrid system. While the panel is in motion, the laser tracker system provides real-time translation measurements but cannot update rotation data due to its reliance on target cycling while the component is stationary. The IMU addresses this gap by providing real-time approximate orientation updates, which are less precise but sufficiently accurate to ensure continuous monitoring of the panel's orientation. The laser tracker, on the other hand, provides precise but intermittent orientation updates when the panel is static. By synchronizing and fusing data from these two sensors, the system benefits from the high accuracy of the laser tracker for precise orientation correction and the IMU's ability to offer continuous updates during motion. This hybrid approach ensures both precision and responsiveness, enabling dynamic tracking and adjustment of the panel's pose in real-time scenarios. In order to fuse the readings from both devices, it has to be calculated on the same reference frame. IMU and RTE's measurements are synchronized using the following transformation tree:

R_{imu} = rotation_imu
$R_{\text{imu_panel}} = I$
$R_{\text{RTE_m}} = R_{\text{mat_RTE}}$
$R_{\text{goal_m}} = I$
$R_{\text{final_imu_m}} = R_{\text{imu}} @ R_{\text{imu_panel}} @ R_{\text{RTE_m}} @ R_{\text{goal_m}}$

2.2.2 Communication System Enhancements

In the previous system, RTE used Bluetooth to communicate between a laptop computer running the software and the MultiStation. However, due to range limitations and constraints of concurrent active connections with Bluetooth, it was decided to switch to a Wireless Local Area Network (WLAN). The switch to WLAN enables seamless, fast, long range communication. Attempts to fuse data collected from the Multistation and IMU simultaneously via Bluetooth encountered communication issues. The switch to WLAN connection facilitates the synchronization and fusion of time-stamped sensor data from both systems thereby enhancing the overall performance of the system for field trials at actual construction sites where the range of Bluetooth could have been a limiting factor.

2.3 Lab Scale Experiments

For sensor selection, tests were conducted to compare the performance of multiple IMUs against the output of a Leica Absolute Tracker AT960 paired with T-probe. However, only the performance of the selected IMU is reported in this paper. The AT960 paired with a T-probe gives highly precise measurements of change in orientation (±15 microns) and translation of any object to which the T-probe is rigidly attached. For a final integration test, the performance of the IMU was evaluated using the Leica Nova MS60 Multi-Station in a lab-scale mockup panel installation setup. The algorithms were written in Python and interfaced with IMU and the robotic total station through WLAN connection automated using Leica GeoCOM commands. The following sections describe the scenarios in which the IMU was tested.

2.3.1 Experimental Setup - Sensor selection

The experimental system consists of a laser tracker and an IMU sensor (x-IMU3) integrated to achieve faster orientation estimation within the semi-loop cycle. The laser tracker is used for precise 3D position and orientation measurements by cycling between three trackable targets. While this method ensures high accuracy, it suffers from an average latency of 10 seconds per update, necessitating the integration of the IMU for quicker approximate orientation estimates.

Table 1. Static accuracy RMS error for a 360° rotation around each axis from the datasheet

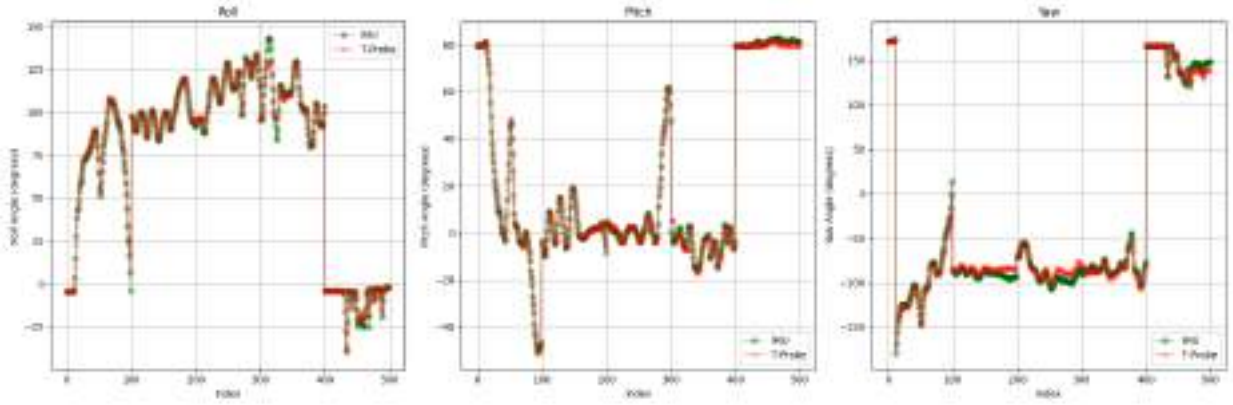
Characteristic	Static Accuracy
Inclination	0.5° RMS
Heading	1° RMS

Table 2. Dynamic error metrics for Roll, Pitch, and Yaw from our lab scale test

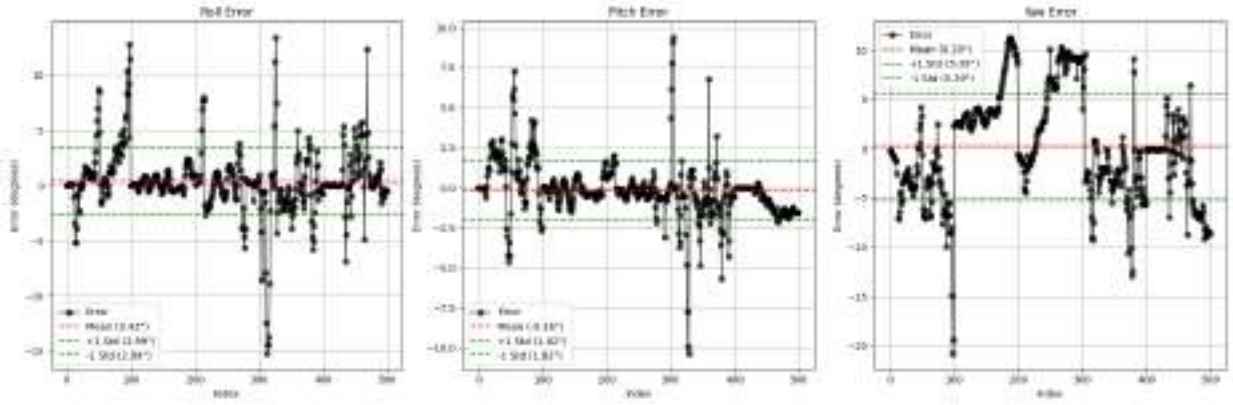
Metric	Roll (°)	Pitch (°)	Yaw (°)
Mean Error	0.43	-0.16	0.20
Standard Deviation	2.99	1.82	5.39

The IMU sensor was chosen based on the following criteria:

- **Accuracy and drift testing:** The IMU demonstrated high static accuracy as seen in Table 1, ensuring reliable orientation measurements during stationary or near-stationary conditions. To evaluate the dynamic accuracy, the IMU was mounted on T-probe and tracked using Leica's Absolute Tracker AT960, a 6DoF laser tracker technology. Data from AT960 and IMU were timestamped and synchronized. A comparison lab scale test between the AT960 and IMU was conducted over a duration of 5 minutes and repeated 3 times. From the results of this test, we concluded that the IMU is reliable enough to provide approximate orientation in highly dynamic scenarios. As shown in Figure 2b the maximum error was observed on the yaw axis, with a standard deviation of



(a) Ground truth vs IMU Data



(b) Error, mean and Standard Deviation

Figure 2. Dynamic accuracy tested using AT960 Laser Tracker as ground truth

5.39. Given that the IMU can provide data at a frequency of 10 Hz or better, it is not expected that the high standard deviation will be an issue since the result will simply be averaged over time. Additionally, the dynamic movements during the test were executed at a significantly higher severity than expected during crane lifting of prefabricated construction elements.

- **Dynamic testing (translation only):** In a translation only dynamic test, the IMU was mounted on a gantry robot system and translated in a square path with zero rotation. The highest average drift was observed in the yaw axis, with an average drift of 2.3 degrees over a period of 10 minutes. For crane-lifted installation of prefabricated components, a drift in yaw axis rotation of over 2.3 degrees in 10 minutes is not insignificant. However, the IMU can be recalibrated to eliminate drift each time the orientation is measured using the object-stationary target-cycling method.

- **Low latent wireless communication and software support:** The x-IMU3 sensor's customizable wireless data stream and its compatibility with Python API simplified its integration into the existing system, allowing seamless data acquisition and processing in rotation matrices, and other formats. Initial tests revealed magnetic drift near ferrous metals, which affected the accuracy of the IMU's rotation values. This issue was subsequently resolved by disabling the magnetometer readings.

2.3.2 IMU integration with RTE:

From the validation tests, we finalized x-imu3 as our choice of sensor for semi-loop integrated real-time orientation approximation. These tests were conducted using a lab-scale setup constructed for previous testing of the RTE system [9]. Figure 3 shows a representation of the lab-scale setup that depicts a 1/3rd scale overclad panel retrofit. Six panels can be assembled on a metal frame that supports a layer of oriented strand board, which serves as a working surface. Twelve connections can be attached to the working surface to serve as hanging points for the six panels. A simulated crane is located on top of the metal frame and can be operated behind the frame to move panels up

and down and left and right. While panels are moved with the crane, a set of three retroreflector prisms is attached to the panel. A multi-station is set up at least 5 meters away from the test frame to measure prisms. For the purpose of these experiments, the IMU is rigidly attached directly to the panel to be measured.

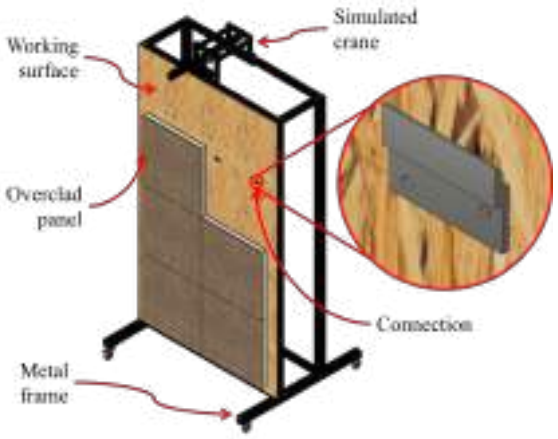


Figure 3. Representation of the lab experiment setup [9]

For the final integration of the IMU into the semi-loop system, a total of five tests were conducted following this procedure:

1. To simplify the test procedure, the panel is attached to a mock crane setup and the experiment begins with the panel located in the goal pose (i.e. the location where the panel should be installed).
2. The coordinates of the three prisms are measured using the multi-station. These measurements provide the position and orientation of the panel required to reach the goal pose based on the digital twin. This initial cycle takes 10 seconds on average.
3. The IMU, which reports orientation in the global frame, is initialized using the previous measurements from the RTE. A coordinate transformation tree is established to align the IMU and RTE's data with the digital twin's reference frame.
4. After initialization, the panel is moved away from its initial pose and RTE tracks a single prism, providing real-time translation updates, while the IMU delivers approximate orientation updates needed to achieve the goal pose.
5. After movement, the panel is brought to a stationary state, and the RTE algorithm commands the multi-station to cycle and measure the coordinates of the three prisms.
6. Upon completing the cycle, the RTE calculates and reports the precise rotation and translation needed to reach the goal pose. Simultaneously, the IMU's readings are timestamped and recorded and the offset value is observed.

The accuracy of the algorithm was evaluated by the difference between the rotation angles calculated from the IMU and those calculated from the RTE algorithm. The measured error in rotation angle was determined by subtracting the angle measured by the RTE in Step 2 from the angle measured by IMU in Step 6. Table 3 below summarizes the resulting Euler Angles observed from our tests, comparing the IMU's performance against precise measurements from the RTE system.

2.4 Results

The results of the experiments highlight the differences between the systems with and without IMU integration. The IMU integration significantly enhances the system's capability by eliminating the need for precise measurements from the RTE mid-cycle while the panel was in motion. Previously, the system did not provide orientation updates when the panel was in dynamic movement, as it required the panel to remain stationary to provide orientation updates. With the IMU integrated, real-time orientation updates became possible even when the panel was in motion, providing a significant improvement in responsiveness.

While the IMU does not reduce the cycle time for precise measurements conducted by the RTE, it complements the system by enabling approximate dynamic orientation updates during motion. This capability ensures that orientation data is continuously available, improving the system's utility in dynamic scenarios. The trade-off between accuracy and speed is analyzed, showing that the IMU effectively supports the laser tracker by maintaining acceptable accuracy levels for dynamic orientation tracking. The integration approach shows promise for scaling the system to larger applications. Future implementations could involve integrating multiple sensors to handle more complex scenarios and adapting the system for deployment in real-world environments.

However, the integration of the IMU introduces challenges such as drift, calibration, and maintaining accuracy under varying conditions. The lab-scale tests were conducted for durations of less than 10 minutes, suggesting that extended testing may help mitigate drift over longer periods. Nonetheless, the accuracy of the IMU is not sufficient enough to replace precise measurements from the RTE system, reinforcing its role as a complementary tool rather than a standalone solution.

Table 3. RTE vs IMU lab scale test results

Test	Angles (°)								
	RTE			IMU			Error		
	Roll	Pitch	Yaw	Roll	Pitch	Yaw	Roll	Pitch	Yaw
Initialization	0.00	0.00	0.00	0.00	0.01	0.01	0.00	-0.01	-0.01
Test 1	-26.05	-0.60	-0.30	-26.77	-1.08	-0.03	0.72	0.48	-0.27
Test 2	-0.09	-15.19	-0.11	0.19	-15.24	-0.56	-0.28	0.05	0.45
Test 3	0.01	25.16	0.06	-0.43	24.98	0.81	0.44	0.18	-0.75
Test 4	1.20	1.03	-21.93	-2.19	2.67	-21.61	3.39	-1.64	-0.32
Test 5	0.33	1.27	-17.78	1.13	1.03	-17.77	-0.80	0.24	-0.01

3 Conclusions

An Inertial Measurement Unit (IMU) was successfully integrated into a Real-Time Evaluator (RTE) system to measure the approximate orientation of crane-lifted components during movement. Previous versions of the RTE required that the component be held stationary to measure the location of three targets attached to the component. After IMU integration, approximate orientation can be measured in real-time without stopping the movement of the component. However, at the current levels of tested accuracy, we do not expect the IMU to completely replace the practice of precisely measuring stationary pose of the component by measuring three targets attached to it's rigid body to determine orientation. Lab experiments were conducted to evaluate IMU performance compared to a highly-accurate six-degree-of-freedom measurement system. Both static and dynamic laboratory experiments showed errors in roll- pitch, and yaw less than 0.5 deg compared to highly-accurate 6DoF laser tracking devices. The IMU was then integrated with the RTE, and an experiment was conducted to verify the integration of the IMU within the workflow of measuring a crane-lifted component as it is installed, providing real-time guidance to installers.

The results of the research prove that IMUs can be used as supplementary devices to measure an approximation of the roll, pitch, and yaw of crane-lifted components. However, for accurate assessments of plumbness and levelness, the IMU measurements alone are not accurate enough to replace conventional methods. Based on these results, the authors see a clear use case for the IMU to approximate orientation while the component is being moved. However, near the final installation point, more accurate methods of orientation measurement such as the previous method of cycled measurements of targets attached to the stationary component are required.

3.1 Future Research

Full-scale demonstrations, both in the laboratory and real-world environments, will be performed to assess the RTE's potential for improving installation efficiency. Future work will include testing the system at an actual construction site to quantify time savings and accuracy improvements on a full scale. Efforts will also focus on refining IMU accuracy over extended period of time, and exploring additional sensor integration to enhance the overall system capability. Additionally, advanced methods using artificial intelligence or machine learning may be explored to improve the accuracy and reliability of the IMU output, which may enable users to install only a single retroreflector on the lifted components and rely entirely on the IMU for orientation measurements, thereby reducing costs.

4 Acknowledgements

This research was supported by the DOE Office of Energy Efficiency and Renewable Energy (EERE), Building Technologies Office, under the guidance of Sven Mumme, and used resources at the Building Technologies Research and Integration Center (BTRIC), a DOE-EERE User Facility at Oak Ridge National Laboratory.

References

[1] Filipe Barbosa, Jonathan Woetzel, Jan Mischke, Maria Joao Ribeirinho, Mukund Sridhar, Matthew Parsons, Nick Bertram, and Stephanie Brown. Reinventing construction through a productivity revolution. Technical report, McKinsey Global Institute, 2017.

[2] Chris Goodier and Alistair Gibb. Future opportunities for offsite in the uk. *Construction Management and Economics*, 25:585–595, 06 2007.

[3] Yuan Chang, Xiaodong Li, Eric Masanet, Lixiao Zhang, Zhiye Huang, and Robert Ries. Unlocking the green opportunity for prefabricated buildings and construction in china. *Resources, Conservation and Recycling*, 139:259–261, 2018.

[4] Pingbo Tang, Daniel Huber, Burcu Akinci, Robert Lipman, and Alan Lytle. Automatic reconstruction of as-built building information models from

laser-scanned point clouds: A review of related techniques. *Automation in Construction*, 19(7):829–843, 2010.

[5] Qian Wang and Min-Koo Kim. Applications of 3d point cloud data in the construction industry: A fifteen-year review from 2004 to 2018. *Advanced engineering informatics*, 39:306–319, 2019.

[6] Aravinda S. Rao, Marko Radanovic, Yuguang Liu, Songbo Hu, Yihai Fang, Kourosh Khoshelham, Marimuthu Palaniswami, and Tuan Ngo. Real-time monitoring of construction sites: Sensors, methods, and applications. *Automation in Construction*, 136: 104099, 2022.

[7] New River Kinematics. Using Real-Time, 6D Object Tracking to Assemble Large Aerospace Components. Online: `https://www.kinematics.com/imagesweb/DeltaIV-2.pdf`. Accessed: 11/03/2024.

[8] Alexandra Papadaki and Maria Pateraki. 6d object localization in car-assembly industrial environment. *Journal of Imaging*, 9(3), 2023.

[9] Nolan W Hayes, Bryan P Maldonado, Mengjia Tang, Peter Wang, and Diana Hun. Optimization of prefabricated component installation using a real-time evaluator (rte) connection locating system. In Vicente Gonzalez-Moret, Jiansong Zhang, Borja García de Soto, and Ioannis Brilakis, editors, *Proceedings of the 41st International Symposium on Automation and Robotics in Construction*, pages 65–72, Lille, France, June 2024. International Association for Automation and Robotics in Construction (IAARC). ISBN 978-0-6458322-1-1. doi:10.22260/ISARC2024/0010.

[10] Mengjia Tang, Nolan W. Hayes, Bryan P. Maldonado, and Diana Hun. Component pose reconstruction using a single robotic total station for panelized building envelopes. In Vicente Gonzalez-Moret, Jiansong Zhang, Borja García de Soto, and Ioannis Brilakis, editors, *Proceedings of the 41st International Symposium on Automation and Robotics in Construction*, pages 105–112, Lille, France, June 2024. International Association for Automation and Robotics in Construction (IAARC). ISBN 978-0-6458322-1-1. doi:10.22260/ISARC2024/0015.

[11] Mohamed Hussein and Tarek Zayed. Crane operations and planning in modular integrated construction: Mixed review of literature. *Automation in Construction*, 122:103466, 2021. ISSN 0926-5805. doi:https://doi.org/10.1016/j.autcon.2020.103466. URL `https://www.sciencedirect.com/science/article/pii/S0926580520310463`.

[12] Heng Li, Greg Chan, and Martin Skitmore. Integrating real time positioning systems to improve blind lifting and loading crane operations. *Construction Management and Economics*, 31(6):596–605, 2013.

[13] Yihai Fang, Jingdao Chen, Yong K Cho, Kinam Kim, Sijie Zhang, and Esau Perez. Vision-based load sway monitoring to improve crane safety in blind lifts. *Journal of Structural Integrity and Maintenance*, 3 (4):233–242, 2018.

Robot-Assisted 3D Scene Reconstruction Using Fixed FOV LiDAR-Based Depth Camera

Vincent J.L. Gan[1,2]

[1] Department of the Built Environment, National University of Singapore, Singapore
[2] Centre of 5G Digital Building Technology, National University of Singapore, Singapore
vincent.gan@nus.edu.sg (correspondence)

Abstract –

This study introduces an innovative robot-assisted 3D scene reconstruction approach featuring a fixed field-of-view (FOV) LiDAR-based depth camera. The approach integrates a quadruped robot with LiDAR-camera for depth measurement and for capturing RGB colour and texture. The RGB-D information is then used in SLAM to enable detailed 3D scene reconstruction and representation of as-built conditions. This hybrid system reduces the costs and operational limitations of traditional terrestrial laser scanning (TLS) while maintaining decent accuracy for downstream tasks such as condition monitoring and scan-to-BIM. By enhancing scanning productivity, this approach can be further integrated with automated point cloud semantic instance segmentation techniques for a streamlined workflow in high-fidelity geometric modelling, advancing robotic 3D BIM reconstruction and contributing to efficient and accurate digital construction applications.

Keywords –

Ground Robot; LiDAR; Image-based Sensing; 3D Scene Reconstruction; Scanning

1 Introduction

There is a growing momentum to harness automated, intelligent technologies to capture and reconstruct 3D digital representations of the built environment, translating the information into virtual platforms. The digital representations (such as Building Information Models) are useful in the architecture, engineering and construction (AEC) industry due to their capability to enhance communication and information exchange between project stakeholders, facilitating lifecycle management of built assets [2]. Traditionally, this process relies on accurate scanning to capture 3D point cloud data reflecting the as-built condition of the structure. This is followed by extensive manual processing of the scan point cloud data to generate the 3D scene or geometric model [3]. This workflow is not only labour-intensive and time-consuming, but also prone to human errors [4].

Several studies have explored the use of Light Detection and Ranging (LiDAR) integrated with robotic platforms, such as ground quadruped robots and wheeled robots, for 3D scene reconstruction of built environments [5-9]. For example, ground robots equipped with 3D LiDAR sensors can autonomously plan routes, navigate to specific targets, and collect point cloud data. The integration of BIM and IndoorGML were presented to optimise quadruped robot navigation for 3D scanning, achieving 70–90% coverage with high accuracy (0.006–0.021 mm), reducing human effort, and outperforming heuristic algorithms in efficiency, coverage, and scan planning [10, 11]. Likewise, an automatic 3D scaffold reconstruction using robot-dog-acquired point clouds was also investigated, achieving 90.84% F1 in semantic segmentation and generating 3D CAD models through deep learning for improved safety management [12]. For building façades [13], ground robot was also used for automated 3D scene reconstruction and visual inspection, demonstrating its growing potential in addressing unique challenges in complex environments. Researchers are actively exploring Simultaneous Localisation and Mapping (SLAM) as a viable option to 3D scene reconstruction. SLAM enables real-time mapping and localisation by allowing a robot or sensor system to build a spatial map of its environment while simultaneously determining its position within that map [1]. In one of the studies, an autonomous mobile robot using 2D SLAM and real-time 3D point cloud mapping was introduced [14] to navigate construction sites, optimise scan positions, and capture RGB-mapped point clouds, addressing occlusions and reducing time for geometric reconstruction.

Nowadays, unmanned aerial vehicles (UAVs) have been utilised for tasks such as surveying outdoor areas,

which collect data more efficiently than ground robots. Collaborations between UAVs and ground robots have shown potential to overcome some limitations. Kandath et al. [15] proposed a system where UAVs enhance ground robot navigation by providing real-time obstacle location data in indoor environments. Furthermore, Kim et al. [16] introduced UAV-based 3D terrain mapping and generation. Peterson et al. [17] also developed an integrated robotic system with ground robots with drones, significantly enhancing the navigation and exploration for outdoor settings.

An example of multi-agent robotic system (MARS) in previous research is shown in Figure 1. The UAV agent is a lightweight drone equipped with an onboard camera, Wi-Fi module, and open-source controller. The UAV can capture 5MP images or 720p live video streams, transmitting data to a processing unit for tasks like path planning. It has a flight time of 8 minutes and 30 seconds, making it ideal for short inspection missions. The ground robot features a wheeled chassis with an intelligent controller and modules for human detection and avoidance. Designed for heavy payloads, the ground robot supports advanced sensing devices and extended operation times.

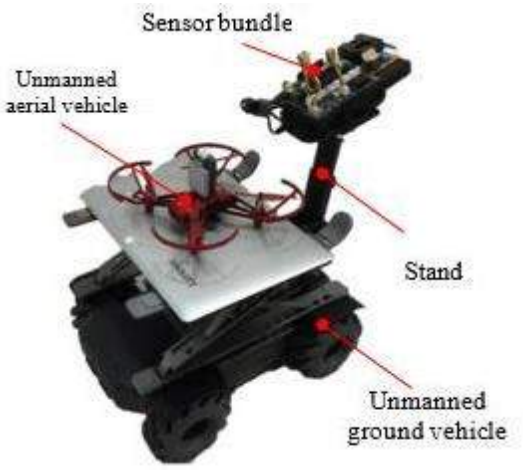


Figure 1. Ground robot equipped with UAV (configuration adapted from Reference [1])

In collaborative operation, the ground robot inspects surroundings and streams video data. As Figure 2 shows, the ground robot carries the UAV and follows a predefined path, capturing facility data and transmitting it to the mediating agent. The mediating agent directs the UAV when obstacles block the ground robot. The mediating agent calculates the optimal UAV trajectories for inaccessible areas and controls the UAV in navigating and scanning the areas that are inaccessible for the ground robot. Results validate the system's feasibility for indoor inspections in cluttered environments. With potential semantic navigation capabilities, ground robots and UAVs can further enhance their ability to interpret the built environment and navigate indoors. This advancement enables efficient and robust 3D scanning by intelligently interpreting and adapting to complex surroundings, ensuring comprehensive spatial data acquisition in challenging environment [18]. Previous research has highlighted the evolving role of ground robots and UAVs in advancing automated data capture in AEC projects.

Nevertheless, research in this field remains in progress with challenges persisting in achieving high-quality scanning outcomes [19]. In particular, significant challenges exist in achieving optimal scanning quality and cost efficiency. Traditional methods such as terrestrial laser scanning (TLS), while highly accurate, are often expensive and cumbersome, especially for large scenes where complex feature mapping requires manual intervention. Current robot-assisted systems, integrated with SLAM algorithms, show good potential for navigating complex environments while effectively capturing data. However, they often struggle to produce dense, high-quality 3D measurements. Furthermore, the use of cost-effective sensors to capture both depth and RGB information in a dense format remains underexplored, particularly without compromising depth accuracy.

This paper investigates a robot-assisted 3D scene reconstruction approach featuring a fixed field-of-view (FOV) LiDAR-based depth camera. By integrating a quadruped robot with this sensor, the approach leverages LiDAR for depth measurement and an integrated camera for capturing RGB, delivering detailed 3D representations of as-built conditions. This approach addresses the cost barriers of traditional TLS by employing an affordable sensor while preserving measurement accuracy. The quadruped robot's manoeuvrability enables it to efficiently scan diverse and complex structural areas, even in cluttered environments. By combining cost-efficient sensors with robotic mobility, this approach not only advances the field of robot-assisted 3D reconstruction but also opens new possibilities for applications such as smart city modelling, continuous monitoring, surveillance, and facilities management, providing a scalable and practical solution for the built environment.



Figure 2. Ground robot equipped with UAV for automated robotic inspection (configuration adapted from Reference [1])

2 Methodology

To achieve complete and accurate point cloud reconstruction that reflects as-built conditions, this study emphasises the integration of a ground quadruped robot with a fixed FOV LiDAR-based depth camera. This integration ensures robust spatial data acquisition while addressing scanning complex environments. To validate the proposed approach, an experiment is conducted on the NUS campus. The scan data are further analysed using two metrics: the completeness and accuracy of the point clouds.

The results provide insights into the feasibility and effectiveness of using quadruped robot-based scanning for detailed 3D scene reconstruction. The methodology and experimental setup are detailed in the subsequent sections.

2.1 Quadruped Robot-assisted Scanning

Figure 3 illustrates the setup of the quadruped robot, featuring an Intel RealSense L515 LiDAR-based depth camera mounted on top. The quadruped robot incorporates high-torque-density electric motors at each leg's knee, thigh, and hip joints, enabling precise control over its movements. This advanced actuation mechanism allows the robot to perform various dynamic actions,

such as walking, jumping, and executing pitch, roll, and yaw motions. These capabilities make it well-suited for operating efficiently in confined and cluttered environments for 3D scanning.

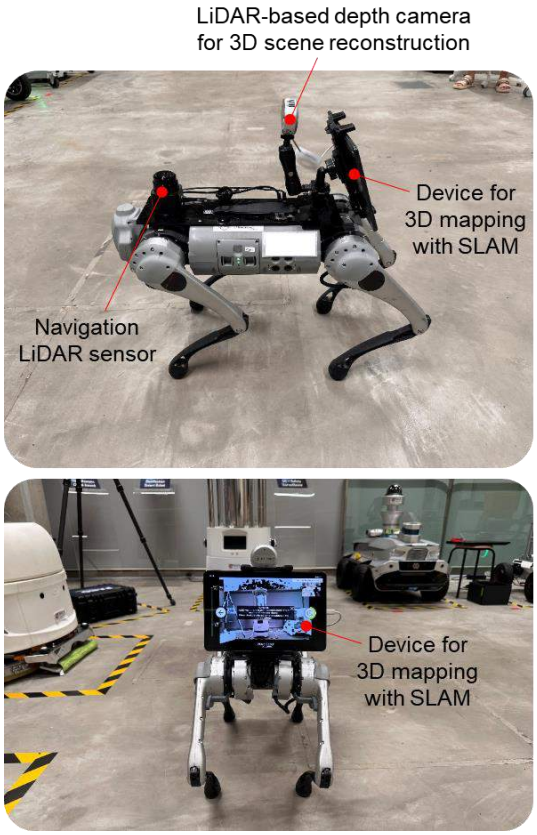


Figure 3. Ground quadruped robot equipped with fixed FOV LiDAR-based depth camera

Inspired by Reference [9], the LiDAR-based depth camera is directly connected to a device mounted on the quadruped robot for real-time SLAM-based 3D mapping. The sensor captures detailed depth maps and RGB colour and texture data, which are processed by the device for 3D scene reconstruction.

By integrating the LiDAR-based depth camera with the quadruped robot, the system utilises the robot's exceptional manoeuvrability to compensate for the fixed FOV of the LiDAR-based depth camera. As the robot moves in confined and cluttered spaces, this configuration enables efficient scanning of structural components while addressing obstacles and occlusions commonly encountered in such environments. The resulting data acquisition ensures decent 3D representations of as-built conditions, with depth measurements and rich spatial data supporting comprehensive mapping and scanning.

2.2 3D Scene Reconstruction

In this study, SLAM algorithm is utilised to map the constructed environment. SLAM operates by generating point clouds from individual scans and matching them within a global coordinate system. To overcome challenges such as obstacles and occlusions, the viewpoints for the robots are strategically identified, ensuring comprehensive scan completeness of the 3D scene. The ground robot then moves along the optimised path while scanning and capturing the RGB-D information for SLAM algorithm to generate the point clouds. This process results in the as-built 3D scenes with further refinement through loop detection.

The Intel RealSense L515 LiDAR-based depth camera is employed. This portable device is mounted on the ground robot's microprocessor and operated using Dot3D Scan software. The placement height is carefully adjusted to maximise the effective scanning range. Key specifications of the sensor include:

- **Dimensions**: 61 mm (diameter) x 26 mm (thickness)
- **Depth Range**: 0.25 to 9 meters
- **Accuracy**: ±14 mm
- **Horizontal Field of View (FOV)**: ~70°
- **Vertical Field of View (FOV)**: ~43°

LiDAR captures depth maps, while the image sensor captures RGB colour, and both are fused into SLAM for 3D scene reconstruction. This setup enables ground robot-mounted scanning to efficiently generate as-built 3D point clouds of the environment.

3 Experiment Results and Discussion

3.1 Testing Site

The effectiveness of quadruped robot-based scanning is demonstrated through experiments conducted on the NUS campus, focusing on 3D scanning and mapping within a selected laboratory in an educational building. Figure 4 illustrates the process of 3D scene reconstruction. In Figure 4(a), the robot moves to the first predefined scanning waypoints. In Figure 4(b), the robot adjusts its motion dynamically (e.g., tilting its head/sensor upward and downward) to scan both upper and lower parts of the structures. The captured RGB-D data are processed in real time for point cloud generation and 3D scene reconstruction (Figure 4(c)). In Figure 4(c), the depth sensor's FOV has a pyramid shape with a square projection at its bottom. This is a result of the sensor's line of sight on structural elements such as walls, meaning that when scanning a planar surface, the projected area should appear square-shaped rather than spherical.

L515 LiDAR sensor is fixed on the quadruped robot, meaning that as the robot moves and tilts its head upward or downward, or reorients in different directions, the sensor's viewpoint shifts accordingly. This movement allows the sensor to capture different regions of the indoor environment to enhance the completeness of 3D scene reconstruction. Using SLAM, the robot achieves dense, accurate scene mapping while travelling to subsequent viewpoints, including confined areas (Figure 4 (d, e)). The final output, shown in Figure 4(f), is a detailed, integrated 3D map of the scanned environment, seamlessly combining data from multiple scanned frames.

3.2 Visual Inspection of 3D Scenes

Figure 5 (a, b) showcases the reconstructed 3D scenes of the testing site, highlighting various structural elements. The reconstruction combines depth and RGB data captured by the LiDAR-based depth camera and processes it in real time using SLAM. The results are dense, visually enriched point clouds that represent the spatial and visual characteristics of the structure and the environment, including the orientation, shape, and

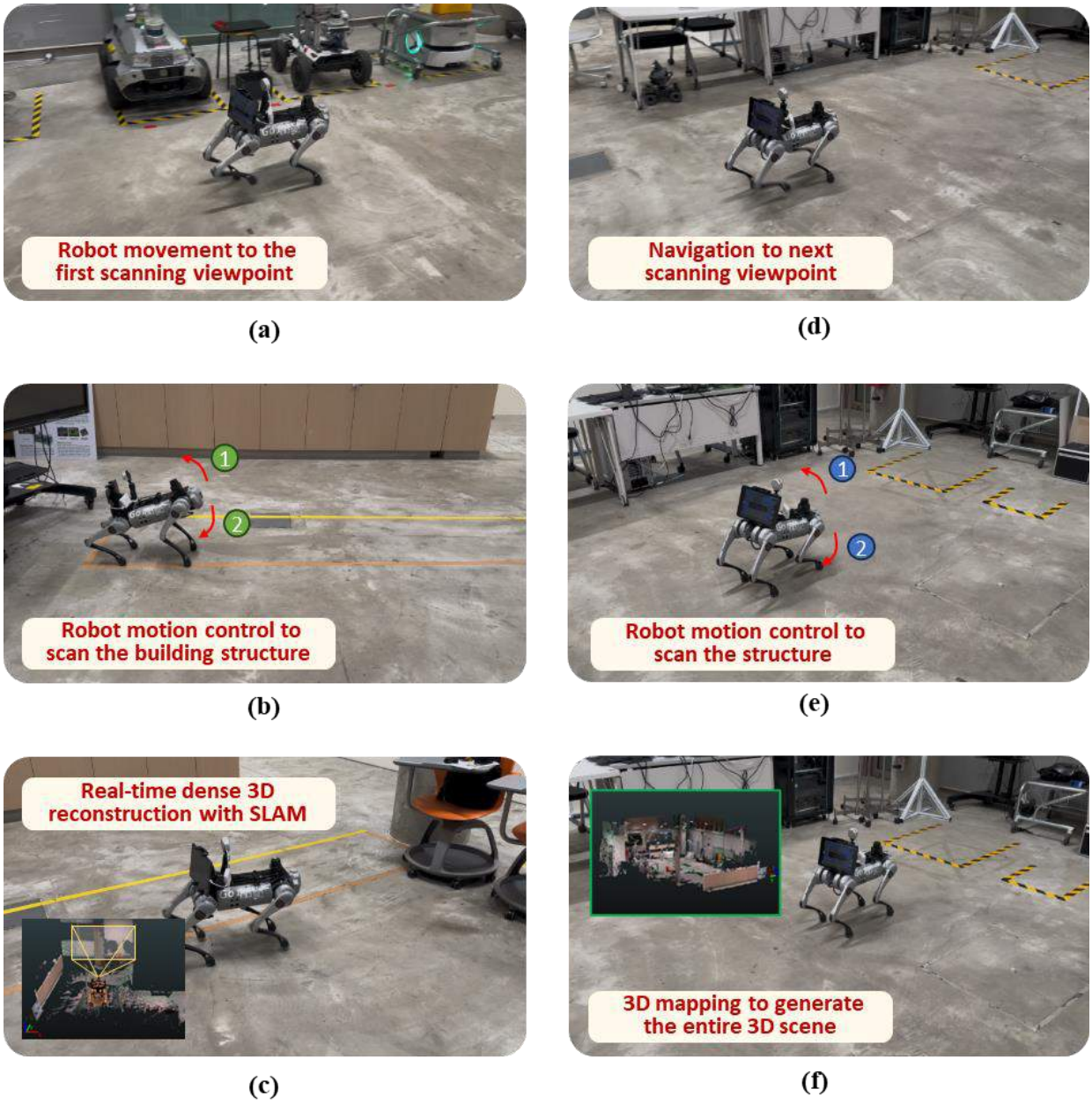


Figure 4. Automated 3D scanning using quadruped robot and fixed FOV LiDAR-based depth camera

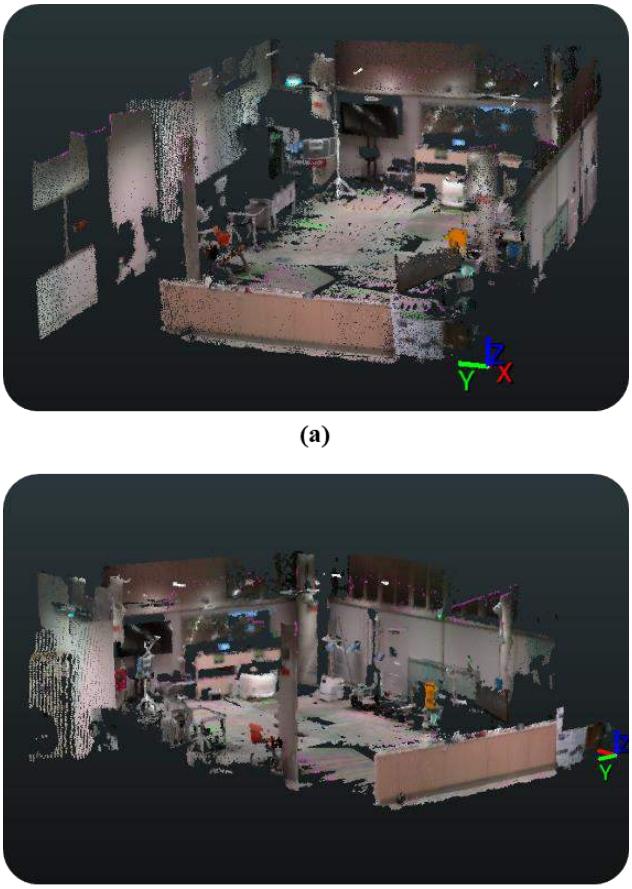


Figure 5. Reconstructed 3D scenes for the testing site

dimensions of walls, floors, furniture, and other architectural components. The addition of colour data further enhances the 3D scenes by reflecting surface textures and material properties. This rich geometric and texture information is invaluable for applications like training AI models for semantic segmentation, enabling effective differentiation and classification of structural and non-structural elements.

The limitation of robot-assisted scanning with a fixed FOV LiDAR-based depth camera is also apparent. Due to its limited FOV, the dimensions of the theoretical measurement area are constrained. The length and width of the measurement area can be calculated using Equations (1) and (2), as follows:

$$L = 2 \cdot \tan\left(\frac{\text{FOV}_{\text{horizontal}}}{2}\right) \cdot \text{Distance} \quad (1)$$

$$W = 2 \cdot \tan\left(\frac{\text{FOV}_{\text{vertical}}}{2}\right) \cdot \text{Distance} \quad (2)$$

wherein L refers to the length of the measurement area, W represents the width of the measurement area, $\text{FOV}_{\text{horizontal}}$ and $\text{FOV}_{\text{vertical}}$ stand for the horizontal and vertical FOV respectively (in degrees), and Distance is the distance from the sensor to the measurement area. Even with the quadruped robot's ability to adjust its sensor orientation for scanning ceilings and floor slabs, the ceiling and upper sections of the structure, including mechanical and electrical (M&E) systems, remain difficult to capture comprehensively due to the sensor's limited measurement distance and area. Such limitations may prevent the reconstructed 3D scenes from meeting the requirements of various applications, such as digital twin creation for M&E systems.

3.3 Quantitative Analysis of Data Quality

The completeness and accuracy of the scan data are analysed using 2D projections of the XZ and YZ planes from the 3D scenes, as shown in Figures 6 and 7. Unscanned areas, highlighted in purple in Figure 6, indicate that approximately 65% of the selected planes were successfully covered. Key structural elements, including walls, windows, and equipment, are clearly identifiable, though corner areas remain partially unscanned due to occlusions and the limited FOV of the LiDAR-based depth camera. A notable fraction of the

purple areas corresponds to windows, and the absence of depth data in these regions should not be considered an error but rather a natural limitation of the LiDAR sensor, as L515 struggles with reflective or transparent surfaces, making it unable to capture reliable depth data. However, this missing data can be leveraged to approximate or identify window boundaries from wall point clouds. The missing data in the upper regions of the wall is not solely due to transparent glass surfaces but may also result from sensor occlusion or range limitation. Future research could further analyse the correlation between these factors on 3D scene reconstruction.

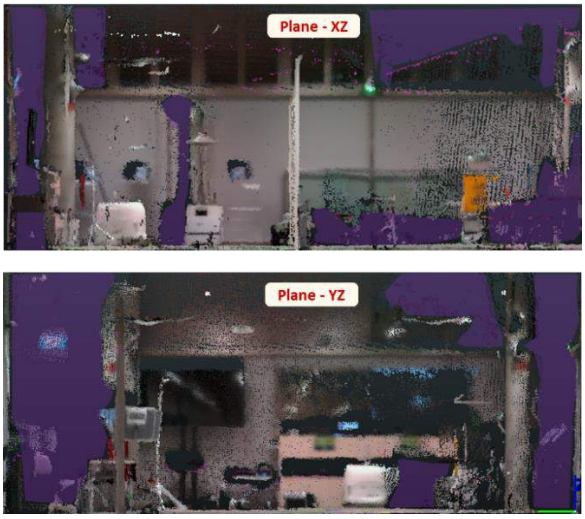


Figure 6. Completeness of scan data

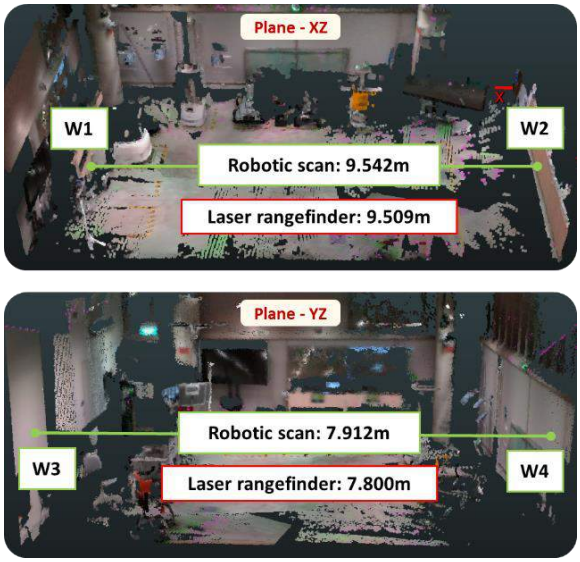


Figure 7. Accuracy of scan data

Figure 7 compares the distance measurement from robotic scanning against conventional laser rangefinders. For the XZ plane, the robotic scan records 9.542m from W1 to W2, while the laser rangefinder measures 9.509m. Similarly, for the YZ plane, the robotic scan reports 7.912m from W3 to W4, compared to the laser rangefinder of 7.800m. Figure 8 illustrates the discrepancies between the measurements obtained from robotic scan and laser rangefinder. The differences are minimal, with a deviation of only 1.4% in the XZ plane and 0.3% in the YZ plane. These differences may result from vibration during the robotic dog's movement, affecting SLAM-based 3D mapping, and the lower accuracy of L515 LiDAR-based depth camera used in the robotic scan.

Despite these constraints, the results demonstrate that robot-assisted scanning using a fixed FOV LiDAR camera is a viable approach for producing decent as-built 3D measurements. This highlights the potential of integrating robotics with LiDAR to facilitate spatial data collection, enabling downstream applications such as scan-to-BIM and facility inspection.

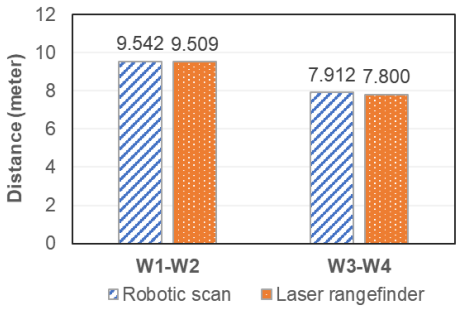


Figure 8. Comparison of distance measurement by robotic scan and laser rangefinder

4 Conclusions

This study demonstrates the potential of integrating a quadruped ground robot with a fixed FOV LiDAR-based depth camera for 3D point cloud reconstruction. The proposed system addresses challenges in capturing spatial data by leveraging the advanced manoeuvrability of a quadruped robot and the precision of a LiDAR-based depth camera. This combination enhances the robot's ability to efficiently move in confined spaces and cluttered environments, while enabling detailed 3D scanning and mapping. The research involves the validation through experimental studies conducted on NUS campus. The field tests focus on scanning an educational building to assess the system's capabilities in capturing as-built conditions. The quadruped robot, equipped with the Intel RealSense L515 LiDAR camera, demonstrates its ability to perform detailed 3D measurements. The results underscore the feasibility and effectiveness of the proposed quadruped robot-based scanning for detailed environmental mapping.

Integrating a fixed FOV LiDAR camera into robotic

systems marks a significant advancement, but several areas require improvement. Firstly, the sensor's limitations in depth measurement, RGB colour, and texture highlight the need for adopting 3D LiDAR-camera. Additionally, given that the L515 sensor provides RGB and depth maps in addition to point clouds, future work of this research could consider integrating these depth maps with RGB data during the 3D mapping process using novel SLAM algorithms, such as Gaussian Splatting-based SLAM, to improve 3D scene reconstruction accuracy. Lastly, advanced data acquisition planning optimisation is essential to guide the robot effectively through complex structural environments. These algorithms should also ensure adaptability to confined spaces, broadening the system's applicability across various use cases.

Acknowledgements

This research is supported by Singapore Ministry of Education under Academic Research Fund Tier 1 (A-8002138-00-00). Any opinions, findings and conclusions or recommendations expressed in the material are those of the author and do not reflect the views of the grantor. The author would like to thank Hu Difeng whose robotic work was referenced in the literature review, and the robot dog 'Ben Ben' for four years' experiments.

References

[1] D. Hu, V. J. Gan, T. Wang, and L. Ma. Multi-agent robotic system (MARS) for UAV-UGV path planning and automatic sensory data collection in cluttered environments. Building and Environment, 221:109349, 2022.

[2] P. Tang, et al. Automatic reconstruction of as-built building information models from laser-scanned point clouds: A review of related techniques. Automation in Construction, 19(7):829–843, 2010.

[3] V. J. L. Gan, K. Li, M. Li, and L. B. E. Halfian. 3D reconstruction of building information models with weakly-supervised learning for carbon emission modelling in the built environment. Applied Energy, 377:124695, 2025.

[4] H. Son, C. Kim, and Y. Turkan. Scan-to-BIM: An overview of the current state of the art and a look ahead. In ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction, pages 12–34, Citeseer, 2015.

[5] S. I. El-Halawany and D. D. Lichti. Detecting road poles from mobile terrestrial laser scanning data. GIScience & Remote Sensing, 50(6):704–722, 2013.

[6] K. S. Yen, T. A. Lasky, and B. Ravani. Cost-benefit analysis of mobile terrestrial laser scanning applications for highway infrastructure. Journal of Infrastructure Systems, 20(4):04014022, 2014.

[7] J. Gené-Mola, et al. Fruit detection in an apple orchard using a mobile terrestrial laser scanner. Biosystems Engineering, 187:171–184, 2019.

[8] J. R. Rosell-Polo, et al. Kinect v2 sensor-based mobile terrestrial laser scanner for agricultural outdoor applications. IEEE/ASME Transactions on Mechatronics, 22(6):2420–2427, 2017.

[9] D. Hu, V. J. L. Gan, and C. Yin. Robot-assisted mobile scanning for automated 3D reconstruction and point cloud semantic segmentation of building interiors. Automation in Construction, 152:104949, 2023.

[10] R. Zhai, et al. Semantic enrichment of BIM with IndoorGML for quadruped robot navigation and automated 3D scanning. Automation in Construction, 166:105605, 2024.

[11] V. J. Gan, D. Hu, R. Zhai, and Y. Wang. Automated As-built 3D reconstruction using quadruped robot and 3D LiDAR sensor. In ISARC Proceedings of the International Symposium on Automation and Robotics in Construction, IAARC Publications, 2024.

[12] J. Kim, D. Chung, Y. Kim, and H. Kim. Deep learning-based 3D reconstruction of scaffolds using a robot dog. Automation in Construction, 134:104092, 2022.

[13] T. Wang and V. J. Gan. Automated joint 3D reconstruction and visual inspection for buildings using computer vision and transfer learning. Automation in Construction, 149:104810, 2023.

[14] P. Kim, J. Chen, and Y. K. Cho. SLAM-driven robotic mapping and registration of 3D point clouds. Automation in Construction, 89:38–48, 2018.

[15] H. Kandath, T. Bera, R. Bardhan, and S. Sundaram. Autonomous navigation and sensorless obstacle avoidance for UGV with environment information from UAV. In 2018 Second IEEE International Conference on Robotic Computing (IRC), IEEE, 2018.

[16] P. Kim, J. Park, Y. K. Cho, and J. Kang. UAV-assisted autonomous mobile robot navigation for as-is 3D data collection and registration in cluttered environments. Automation in Construction, 106:102918, 2019.

[17] J. Peterson, et al. Online aerial terrain mapping for ground robot navigation. Sensors, 18(2):630, 2018.

[18] D. Hu, V. J. Gan, and R. Zhai. Automated BIM-to-scan point cloud semantic segmentation using a domain adaptation network with hybrid attention and whitening (DawNet). Automation in Construction, 164:105473, 2024.

[19] A. Aryan, F. Bosché, and P. Tang. Planning for terrestrial laser scanning in construction: A review. Automation in Construction, 125:103551, 2021.

New collision avoidance algorithm approach based on social sciences

Shuntaro Tamura[1] **and Masahide Horita**[1]

[1]University of Tokyo, Japan.

tamura-syuntaro653@g.ecc.u-tokyo.ac.jp, horita@g.ecc.u-tokyo.ac.jp

Abstract -

In future automated construction sites, multiple autonomous agents must navigate safely and efficiently. While velocity obstacle-based algorithms like Optimal Reciprocal Collision Avoidance(ORCA) handle microlevel control, achieving efficiency also requires the management of macrolevel dynamics, similar to human social behavior. The existing C-Nav algorithm, which incorporates the concept of "consideration for others" to promote yielding behavior, shows promise, but its evaluation method can lead to inefficient or nonconsiderate actions. This study found out that insufficient space for yielding causes these issues, and proposes a new evaluation method to encourage social behaviors. The results indicate reduced stacking in narrow areas, although some inefficiencies persist. As a result, the "waiting behind" action functions as an auxiliary mechanism that reinforces the mutual yielding behavior intended by C-Nav, thereby enhancing overall operational efficiency. These insights suggest that further refining the algorithm, guided by a deeper understanding of human social interactions, can lead to more robust and efficient navigation strategies.

Keywords -

Multiagent navigation; Coordination; Robotics

1 Introduction

Owing to the growing shortage of skilled workers in the construction sector, the automation of construction machinery is advancing to improve productivity and safety. A noteworthy example is Taisei Corporation's T-iCraft®, which tests coordinated operations using four types of autonomous machines, namely backhoes, crawlers, dump trucks, and bulldozers.

Path planning for autonomous robots relies heavily on their ability to "pass each other." For instance, in environments in which robots cannot pass each other, all pathways must be one-way or single-track, which leads to inefficient movements.

Figure 1 illustrates a situation in a construction site where passageways are blocked by construction materials, making it difficult for robots to pass each other. Such situations are expected to frequently occur in construction sites, and developing methods to address them is crucial for advancing automation.

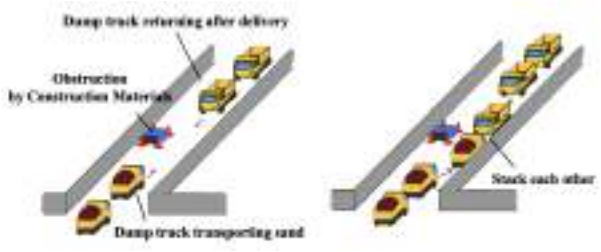


Figure 1. Illustration of the Situation in Which Robots Cannot Pass Each Other

Planning the movements of multiple autonomous construction machines (referred to as "agents") in continuous space to avoid collisions is not easy. Although humans can easily manage passing movements, autonomous machines require advanced algorithms. Velocity obstacle-based algorithms like Optimal Reciprocal Collision Avoidance (ORCA) [1], Reciprocal Velocity Obstacle (RVO) [2], Hybrid Reciprocal Velocity Obstacle (HRVO) [3], and Extended Velocity Obstacle (EVO) [4] treat other agents as dynamic obstacles and calculate safe velocities to prevent collisions.

Velocity obstacle-based algorithms use distributed systems in which each agent independently calculates its movement based on predefined rules. . This ability is aligned well with construction sites, where obstacles frequently appear, as illustrated in Figure 1. In addition, construction machines are typically large and lack maneuverability, making detours less feasible compared to waiting for clearance. This constraint is unique to construction sites, and under this premise, the avoidance-focused algorithm proposed in this study is expected to be highly effective.

However, the algorithms often leads to overall efficiency reductions such as traffic jams (referred to as "stacks"). By drawing inspiration from human interactions, such as crowd dynamics and mutual respect, and incorporating a social modeling framework into these algorithms, these types of problems may be mitigated. The literature on

game theory and other related disciplines in the social sciences has a crucial role in developing these enhanced algorithms.

C-Nav [5] is a collision avoidance algorithm that incorporates the concept of "empathy," voluntarily balancing agents' progression toward their goals in consideration of the freedom of others. While C-Nav improves movement efficiency in narrow passages, challenges remain in diverse site conditions and parameter dependencies. Strengthening the generalizability and robustness of such algorithms is essential for their practical deployment.

The purpose of this study is to analyze the effects of enhancing sociality, by introducing more social parameters within the existing C-Nav algorithm, on resulting group performances. It aims to apply analogies with human social behaviors to agent interactions and examine whether they improve the algorithm's performance. Methodologically, we conduct herein computer simulations based on existing ORCA and C-Nav codes, visualize agent behaviors, propose improvements, and compare results to evaluate enhancements. The novelty of this study is its emphasis of algorithmic robustness and strengthening of this robustness through the incorporation of new social behavior concepts.

2 Related Work

2.1 Overview of Collision Avoidance Algorithms

Various approaches exist for collision avoidance algorithms in a continuous space. The mainstream approach is that associated with the use of velocity obstacle-based algorithms. In these algorithms, agents calculate the velocity (to be adopted next) at each time step based on information about their surrounding environment. One of the advantages of this approach is its distributed nature. Instead of relying on a central decision-maker overseeing the entire environment, individual players plan their paths based on their judgments. This improves the computational efficiency and allows flexible responses to dynamic environments.

2.2 Overview of ORCA

Various formulations have been proposed to determine "safe" velocities, aiming to ensure safety while maintaining the maximum possible freedom for velocity selection. Among these, ORCA is currently one of the most extensively used methods. ORCA calculates the "collision velocity range" based on the velocities and positions of the surrounding agents. By "sharing" velocity vector adjustments to escape from this range with other agents, ORCA can estimate a "safe velocity range" with reasonable accuracy. Within this "safe velocity range," agents are free to select velocities (e.g., by choosing the direction and speed that brings them closest to their goal).

This method has been demonstrated to enable collision avoidance among thousands of agents, with the computation time scaling linearly with the number of agents. However, as ORCA focuses on guaranteeing safe velocities at each time step, there is room for improvement in motion planning from a macroscopic perspective. For example, in narrow passages, simply directing groups of agents toward their goals can lead to stacking when the two groups intersect.

Since the introduction of ORCA, research on velocity-based collision-avoidance algorithms in environments that guarantee accurate and complete communication has primarily focused on addressing these types of inefficiencies. Examples include the VGVO [6], proxemics [7], and GrpAvoid [8]. The mainstream approach proposes improvements in velocity decisions at a slightly higher macrolevel than ORCA. The same approach was adopted in this study.

2.3 Limitations of the ORCA Algorithm and the C-Nav Algorithm

When proposing algorithms to address the aforementioned limitations of ORCA, many researchers have drawn inspiration from human social behaviors and game-theoretic considerations. Among these, the C-Nav algorithm yields particularly interesting results. C-Nav is an algorithm that incorporates the advanced social behavior concept of "yielding." Explicitly integrating the concept of "empathy" from the related disciplines in social sciences, it predicts actions several steps ahead and rewards movements that reduce constraints on the actions of others. Consequently, it enables construction machines to pass each other in narrow passages, where only one machine can pass at a time.

3 Problem Formulation

3.1 Definition of the Velocity Obstacle Problem

According to Van Den Berg et al. [1], the velocity obstacle (VO) problem is formulated as described below. We consider an environment shared by n agents. Each agent was modeled as a disk moving in a two-dimensional plane. The following information is provided for agent A:

- Current position ($\mathbf{p}_A$): The agent's current location in the 2D plane
- Current velocity ($\mathbf{v}_A$): The agent's current speed and direction
- Radius ($\mathbf{r}_A$): The radius of the agent assumed to be a two-dimensional circle
- Maximum velocity ($v_A^{\max}$): The upper limit of the agent's speed

- Preferred velocity ($\mathbf{v}_A^{\text{pref}}$): The desired velocity in the absence of obstacles, typically directed toward the goal with magnitude $\mathbf{v}_A^{\max}$

The symbols $\mathbf{p}_A$, $\mathbf{v}_A$, and $\mathbf{r}_A$ are considered as external states that are observable through environmental sensing. By contrast, $\mathbf{v}_A^{\max}$ and $\mathbf{v}_A^{\text{pref}}$ are internal states that are not directly observable by the other agents. With a time-step interval τ, each agent selects a new velocity based on the parameters at each time step.

3.2 ORCA Algorithm Details

The primary objective of ORCA is for each agent to select a velocity that ensures that there are no collisions with other agents or obstacles in the subsequent time step. In addition, as a secondary objective, the agent may choose a velocity that is as close as possible to the preferred velocity v_{pref}^A. ORCA was designed to satisfy these requirements.

The collision velocity range refers to the range of velocities based on two agents that may collide within a specific timeframe. Selecting a velocity within this range introduces a risk of collision. The collision avoidance set is the set of velocities that avoids this range in the subsequent time step. By selecting the velocity from this set, we guaranteed that no collisions would occur. Based on this idea, each agent calculates the minimum necessary velocity adjustments to avoid collisions with all surrounding agents. From the convex set of safe velocities obtained through this process, Agent A selects the velocity closest to v_{pref}^A.

3.3 C-Nav Algorithm Details

In this section, we describe the setting of v_{pref}^A, which is a unique feature of C-Nav, according to the explanation of Godoy et al. [5].

At each time step, the agent first simulates hypothetical actions over several future steps for various possible actions it can take in the subsequent time step based on observations of the positions and velocities of the surrounding agents. Examples of possible actions include movements in eight directions based on the direction directly toward the goal.

Each action is then scored according to its degree of goal achievement and those of the surrounding agents. Herein, the agent's degree of goal achievement is defined as the distance moved closer to the goal. The degree of goal achievement for the surrounding agent j (more precisely, the extent to which the agent's action affects the agent's goal achievement) is given by the value $P_{a,j}$

$$P_{a,j} = v^{\max} - \|\mathbf{v}_j^{\text{intent}} - \mathbf{v}_j^{\text{new}}\| \tag{1}$$

where v_j^{intent} is assumed to be communicated by an agent j via a signal.

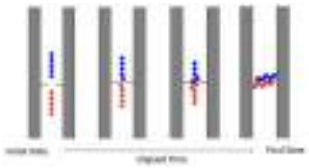


Figure 2. Behavior in the Case of Optimal Reciprocal Collision Avoidance

Consequently, the reward for each action is calculated using the following equations:

$$R_a = (1-\gamma)\cdot R_a^g + \gamma \cdot R_a^c \tag{2}$$

$$R_a^g = \frac{\sum_{t=0}^{T-1}\left(\frac{\mathbf{v}_i^{\text{new}}\cdot(\mathbf{g}_i-\mathbf{p}_i)}{\|\mathbf{g}_i-\mathbf{p}_i\|}\right)}{(T\cdot v^{\max})} \tag{3}$$

$$R_a^c = \frac{\sum_{t=1}^{T-1}\sum_{j\in C_{\text{rank}}} P_{a,j}}{(T-1)\cdot k\cdot v^{\max}} \tag{4}$$

where γ is a parameter that balances the agent's goal achievement and the consideration of other agents. A value of γ closer to zero places more emphasis on the agent's goal achievement, whereas a value closer to one places more emphasis on the consideration of other agents. This γ is known as the coordination factor.

Finally, the action with the highest R_a value is selected and passed to ORCA as v_{pref}^A to determine the final velocity.

4 Observation and Analysis

In this section, we determine what needs to be improved regarding the behavior of agent groups resulting from the existing C-Nav algorithm based on observations of our simulations.

4.1 C-Nav's Superiority over ORCA in Narrow Passages

In this section, we explain how C-Nav overcomes the challenges presented by ORCA and what concrete improvements it achieves in practice. In section 4.2, we present examples of undesirable behaviors in which C-Nav deviates from its intended principles, thereby clarifying directions for further refinement.

First, in the case of a simple ORCA, the two agents are set to move straight toward their respective goals; thus, they both enter a narrow passage. Because they collide, the leading agents of each line try to avoid each other by moving to the left and right, as calculated by the ORCA algorithm. However, because the narrow passage is adequately broad for only one agent, they cannot pass each other and become stuck in the passage (Figure 2).

We now explain the ideal avoidance behavior intended to be used in the original paper [5] in the case of C-Nav. In

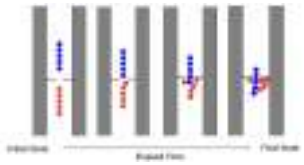


Figure 3. Behavior in the Case of C-Nav

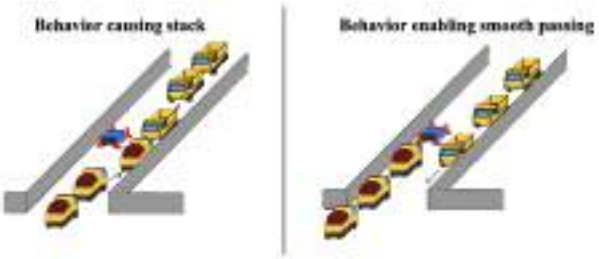


Figure 4. Illustration of the Real-world Narrow Passage Scenarios

Figure 3, the red agent determines that proceeding straight ahead would obstruct the blue agent, and will thus avoid moving to the right. Meanwhile, the blue agent notices that the red agent has moved to the right and proceeds straight ahead, resulting in "passing each other" behavior.

These simulated behaviors correspond to real-world movements observed at construction sites, as shown in Figure 4, specifically as these pertain to the passing of dump trucks heading to transport soil and those returning from transportation.

4.2 Unintended Stacking Caused by C-Nav in Narrow Passages

We also observed the behavior after modifying the width of the narrow passage to allow the two agents to pass through each other side-by-side. The ideal behavior in this field is for the agents to shift to the left and right directions and pass each other while maintaining their formation, as shown in Figure 5. However, when we conducted repeated trials under the same conditions, the following behavior often occurred (Figure 6). In their efforts to predict several

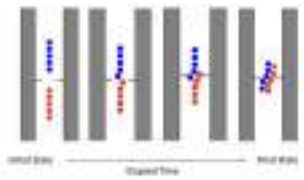


Figure 5. Smooth Behavior Attained in the Case of C-Nav

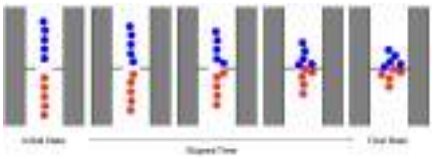


Figure 6. Stacking in the Case of C-Nav

steps ahead of their current positions in cases of straight motion trajectories, both the leading agents of the red and blue groups obstruct each other and fail to achieve their own goals; accordingly, they try to change their courses by moving to the right or left. However, the directions they chose to avoid were the same, resulting in a situation where they blocked each other's paths. In this case, the two leading agents that are "hesitating" restrict the width of the narrow passage, and the agents which follow can only pass through one at a time, reducing the efficiency of the crossing.

The fact that this inefficient behavior occurs often, presents a challenge when considering applications to actual construction sites and indicates a lack of robustness in the current algorithm. Therefore, modifications are required to the algorithm to avoid these types of inefficiencies.

4.3 Analysis of the Cause of Stacking

4.3.1 Changes in Behavior When Switching to Sequential Turns

First, let us consider an analogy to human behavior. In the narrow passage situation presented in Section 4.2.3, it is common to experience congestion in reality because the avoidance directions of oncoming individuals overlap. The essential cause in this example is that "the choice of avoidance direction is made simultaneously." If you knew (from your perspective) that another person would avoid it, you would not avoid it. Believing that this phenomenon also occurs in C-Nav, we modified the algorithm. We changed the evaluation of the actions in each direction, which was previously performed simultaneously based on the desired velocities of each agent at the start of the for-loop—to be performed sequentially. With this change, each agent performs evaluations while recognizing the preferred velocities of surrounding agents.

However, even when this algorithm was used, avoidance actions in the same direction occurred (Figure 7).

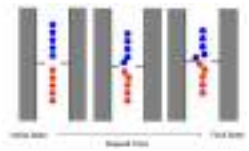


Figure 7. Same-direction Avoidance in the Case of Sequential Turns

Therefore, we concluded that this behavior was not due to the order of turns.

4.3.2 Changes in Behavior When the Number of Agents was Altered

We verified this by changing the number of agents and their intervals.

First, in the one-on-one case, the results were as follows: out of 1000 trials, the shortest and longest completion times were 98 and 112. In other words, severe stacks, such as those described in Section 4.2.3, did not occur.

In another case, the following trajectories were observed: The red and blue lines represent the agents moving upward and downward, respectively. If we focus on the points where each agent first changes its direction of movement, we can observe that both avoid the right side of the screen. However, after several adjustments in direction, these two agents eventually passed through (Figure 8).



Figure 8. Elimination of "Hesitation" Observed in the 1 on 1 Case

Based on the results presented earlier, we can infer that when there is sufficient space around oneself and the target agent that needs to be avoided, repeated directional adjustments are made; as a result, stacking due to mutual avoidance does not occur.

4.3.3 Specific Mechanism of Stacking in C-Nav and Opportunities for Algorithm Improvement

Based on the results of Sections 4.3.1, and 4.3.2, it can be inferred that the presence of the following agents influences the occurrence of stacks owing to mutual avoidance. Reinterpreting the behavior based on this, the sequence is as follows:

1. In Figure 6, both the red and blue agents first move to the right side of the screen to avoid obstacles.
2. The following agents proceeded straight into the space that had opened up.
3. The leading red and blue agents, who had initially avoided each other by moving to the right, lose the freedom to readjust their direction of movement as in the one-on-one case owing to the entry of the agents that followed; these two agents are continually pushed further to the right by these incoming agents.

It is considered that the stacks occurred because of the aforementioned mechanism.

Considering the meaning of this behavior, it implies that "even though the leading agents are trying to yield the way to the opposing agents, the following agents attempt to overtake and proceed." This contradicts one of the major themes of the C-Nav algorithm, which is "consideration for others."

Furthermore, this behavior was observed in cases where the

1. coordination factor was increased
2. reward target was limited to only the nearest agent
3. agents were allowed to select low-speed actions or stop at desired velocities

These observations indicate that with the current reward function, it is difficult to restrict "overtaking."

4.4 Proposed Improvements

As pointed out by Godoy et al. [5], an attempt to vary the way rewards are given, depending on whether the agents are moving in the same or opposite directions, is considered useful for constructing a more efficient algorithm. Based on this, we will revisit the current method of assigning rewards by focusing on the "direction of movement of the agent being considered."

The current method for assigning consideration rewards is as follows:

$$P_{a,j} = v^{\max} - \|v_j^{\text{intent}} - v_j^{\text{new}}\| \tag{5}$$

In other words, this method values "not hindering the movement of others."

Based on our observations thus far (e.g., passing through narrow passages in one-on-one situations), it has been confirmed that this is effective for agents moving in opposite directions. However, for agents moving in the same direction—especially the agent immediately ahead who is stalled—the possibility that one's actions will directly block the agent's desired path is low. Consequently, differences in the rewards for each of one's actions were less likely to occur. Moreover, as we have observed, blocking the space necessary for avoidance actions between that agent and agents facing it becomes a "hindrance" when viewed over a longer number of time steps. Therefore, a

more appropriate method for assigning rewards is required to avoid this.

We can draw an analogy with human social behavior. Suppose that at a place where only one person can pass—such as an automatic door at a building entrance—a few people from inside and another group from outside arrive and stumble across each other. In this situation, the leader of the group (coming from inside or outside) will step aside by the door and yield the way to the other group. At this time, the highest priority for the other members of the group that is yielding (assuming everyone in this group is considerate) is "not to step ahead of the person who is yielding."

Based on the above, a more desirable way of consideration is to "ensure freedom of action for agents moving in the opposite direction as in the existing method, but for agents moving in the same direction, if that agent yields the way or its movement stalls, to line up behind and wait."

In the existing method case, the reward calculation is performed for a specified number of agents in order from those with the largest speed change, regardless of the distances from each other. However, this method of assigning rewards depends on how the neighborhood is defined (e.g., the setting of the neighborhood radius); correspondingly, robustness against the problem situation is not guaranteed. In Godoy et al. [5], there is limited discussion on the order of consideration. When introducing the movement of "lining up behind a stalled agent ahead," it is more intuitive to consider agents as targets for this reward concept in order of increasing proximity to oneself.

Based on this, we propose changing the method for assigning consideration rewards as follows.

Existing reward assignment methods

Without distinguishing the agents, we calculate $P_{a,j}$ for each agent using Eq. 5, and estimate the average of a specified number of agents in order from the ones with the largest $P_{a,j}$ as a consideration reward.

Proposed method for reward assignment

We define whether the "flow directions are the same or opposite" based on whether the inner product of one's desired velocity vector and that of the other agent is positive or negative. In addition, we define an agent's movement as "slow" if its current speed is less than half its maximum speed. In this case,

- For agents j moving in the opposite direction, calculate $P_{a,j}$ using Eq. 5.
- For agents j moving in the same direction, calculate $P_{a,j}$ using the following equation:

$$P_{a,j} = \alpha \times \frac{(p_j - 4.0 \times r_j \times v_j^{\text{intent}}) - p_i}{\|(p_j - 4.0 \times r_j \times v_j^{\text{intent}}) - p_i\|} \cdot v_a \tag{6}$$

In this study, we refer to this method of providing rewards as "Q-Virtue," inspired by the Japanese queuing culture (the virtue of standing in line).

5 Experiments

We conducted experiments to verify whether the proposed changes in the reward rules led to a reduction in stacking in narrow passages.

5.1 Evaluation Method

Using C-Nav as a comparison point, we evaluated the C-Nav algorithm using the proposed reward allocation method.

In this study, we focused on actions to determine whether it is possible to avoid the phenomenon according to which the movement completion time becomes considerably longer than when ideal crossing occurs, owing to stacks resulting from the existing C-Nav. Therefore, we evaluated the method by performing 100 trials and comparing the frequency distribution of the movement completion time (the average time required for 10 agents to reach their goals).

Regarding the parameters of C-Nav, we fixed them at the following values, which showed the best performance in the existing algorithm as a result of the experiments; coordination factor (γ) = 0.9, and number of agents considered in reward = 1.

5.2 Results

For each scenario, we present the following for both the existing C-Nav and the proposed assignment method:

1. A distribution plot with the average completion time (in s) on the horizontal axis and the frequency on the vertical axis
2. A sequence of representative snapshots of observed behavior (the left image shows the initial state and the right image shows the state after some time has elapsed)

5.2.1 Narrow Passage Scenario

First, we consider a situation in which agents can pass through two narrow passages simultaneously.

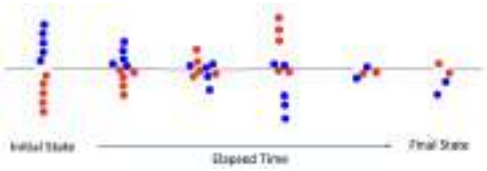


Figure 9. Behavior in the Case of C-Nav

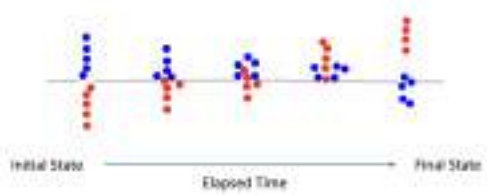


Figure 10. Behavior in the Case of Q-Virtue

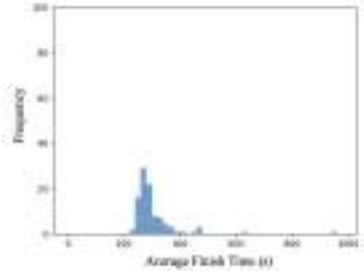


Figure 11. Completion Times in the Case of C-Nav

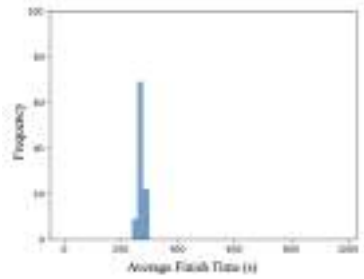


Figure 12. Completion Times in the Case of Q-Virtue

In contrast to the existing C-Nav method, where Figure 9 illustrates numerous instances of agents blocking each other's paths in narrow sections, the Q-Virtue demonstrates movements where groups on the avoiding side align in orderly rows, as shown in Figure 10. Additionally, virtually no time was spent obstructing each other's paths. Consequently, compared with the existing C-Nav (Figure 11), the distribution of movement completion times for the Q-Virtue (Figure 12) is skewed to the left. This shift led to a decreased number of trials in which crossing completion required a long time.

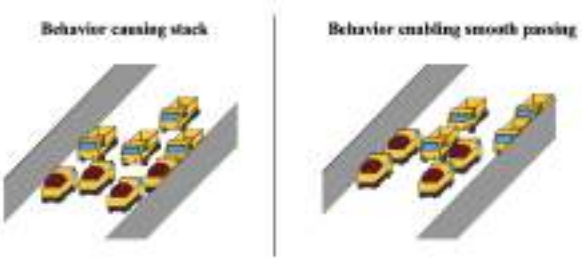


Figure 13. Illustration of the Real-world Situations Associated With the Group Crossing Scenario

5.2.2 Group Crossing Scenario

We focus next on the crossing of two groups consisting of multiple agents in a crowded situation and compare the behaviors resulting from each method handled so far. We verified whether Q-Virtue, which focuses on improving behavior in narrow passages, is also efficient in other situations. These simulated behaviors correspond to the real-world movements observed at construction sites, as shown in Figure 13.

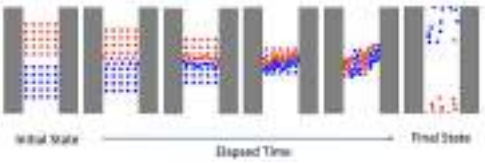


Figure 14. Behavior in the Case of C-Nav

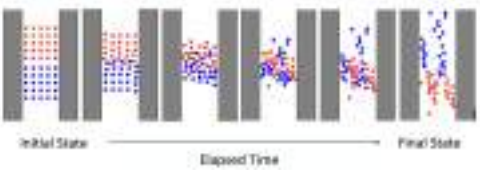


Figure 15. Behavior in the Case of Q-Virtue

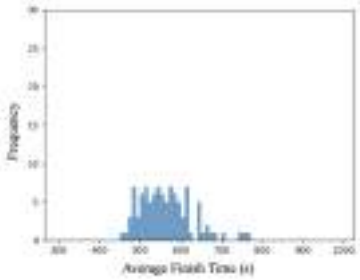


Figure 16. Completion Times in the Case of C-Nav

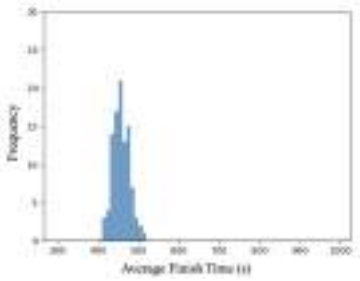


Figure 17. Completion Times in the Case of Q-Virtue

When comparing the existing C-Nav method (Figure 14) with the Q-Virtue (Figure 15), the latter exhibits formations at an earlier stage, which consequently alleviates crowding during crossings. Consequently, the distribution of movement completion times for the Q-Virtue (Figure 17) is skewed to the left compared with the existing C-Nav method (Figure 16). This shift leads to a decreased number of trials in which crossing completion requires a long time.

5.3 Summary

In all situations, the distribution of the average movement completion time shifted to a side with a smaller length because of the Q-Virtue.

6 Conclusion and future work

The existing C-Nav yielded interesting results by designing the algorithm to incorporate information related to the "consideration for others," generating yielding behaviors. However, it was found to be parameter-dependent, especially concerning behaviors during crossings in narrow passages, revealing robustness issues as areas for improvement in practical applications. To elicit the unique performance intended by C-Nav, which is "mutual yielding," we proposed the addition of auxiliary processes and consequently achieved an improvement in robustness. These findings provide new insights into refining velocity obstacle-based collision avoidance algorithms and can help alleviate congestion and facilitate more efficient passage among agents in future automated construction sites.

In this study, we demonstrated that for C-Nav (and for many other collision avoidance algorithms proposed as improvements of ORCA), the action of "waiting behind" is an important factor in achieving efficient movement. However, discussions on methodologies to induce the behavior of "waiting behind" and the conditions under which it becomes effective are still insufficient, and more detailed analyses and improvements are required.

Due to limitations in research duration and equipment, this study did not conduct experiments in actual construction sites. Future work will focus on testing the algorithm in real environments to assess the impact of construction-specific physical and social factors.

Acknowledgement

This study was supported by JSPS KAKENHI Grant Number JP22H01561. We would like to thank Editage (www.editage.jp) for English language editing.

References

[1] M. Lin J. Van Den Berg, S. J. Guy and D. Manocha. Reciprocal n-body collision avoidance. In *Robotics Research: The 14th International Symposium ISRR*, pages 3–19. Springer, 2011. doi:10.1007/978-3-642-19457-3_1.

[2] M. Lin J. Van Den Berg and D. Manocha. Reciprocal velocity obstacles for real-time multi-agent navigation. In *2008 IEEE international conference on robotics and automation*, pages 1928–1935. Ieee, 2008. doi:10.1109/ROBOT.2008.4543489.

[3] S. J. Guy J. Snape, J. Van Den Berg and D. Manocha. The hybrid reciprocal velocity obstacle. *IEEE Transactions on Robotics*, 27(4):696–706, 2011. doi:10.1109/TRO.2011.2120810.

[4] S. Engel A. Levy, C. Keitel and J. McLurkin. The extended velocity obstacle and applying orca in the real world. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 16–22. IEEE, 2015. doi:10.1109/ICRA.2015.7138974.

[5] M. Gini J. Godoy, S. J. Guy and I. Karamouzas. C-nav: Distributed coordination in crowded multi-agent navigation. *Robotics and Autonomous Systems*, 133: 103631, 2020. doi:10.1109/ROBOT.2008.4543489.

[6] V. G. Santos and L. Chaimowicz. Cohesion and segregation in swarm navigation. *Robotica*, 32(2):209–223, 2014. doi:10.1017/S0263574714000563.

[7] W. Wang L. He, J. Pan and D. Manocha. Proxemic group behaviors using reciprocal multi-agent navigation. In *2016 IEEE international conference on robotics and automation (ICRA)*, pages 292–297. IEEE, 2016. doi:10.1109/ICRA.2016.7487147.

[8] J. Ma L. Luo, X. Wang and Y.-S. Ong. Gr-pavoid: Multigroup collision-avoidance control and optimization for uav swarm. *IEEE Transactions on Cybernetics*, 53(3):1776–1789, 2021. doi:10.1109/TCYB.2021.3132044.

Enhancing Robotic Path Planning in Industrialized Construction: A Novel Benchmarking Approach

Amirpooya Shirazi[1] and Aladdin Alwisy[2]

[1,2] M.E. Rinker, Sr. School of Construction Management, University of Florida, United States
a.shirazi@ufl.edu, aalwisy@ufl.edu

Abstract –

Over the past decade, Industrialized Construction (IC) has experienced increasing attention, with numerous studies examining the integration of various technologies to boost productivity and efficiency through manufacturing-oriented strategies. Among these, robotic arms have emerged as a dominant solution. However, further research is required, as the interactions between robotic arms and the surrounding environment often remain suboptimal or even hazardous. This challenge is primarily attributed to the absence of a practical benchmarking tool for selecting appropriate path-planning pipelines under different scenarios. Moreover, existing benchmarking frameworks often disregard the complexity inherent in IC robotic stations, producing path-planning rankings that are not feasible for the required assembly strategies. To address these gaps, this study proposes a novel framework for generating a benchmarking tool that accounts for both the specific robotic arm task and user-defined criteria, enabling a ranking system for evaluating multiple path-planning pipelines. The proposed framework was validated in a simulated robotic station designed for manufacturing structural wooden assemblies. Findings highlight the framework's robustness by demonstrating how a criteria-driven evaluation metric can effectively compare diverse path-planning pipelines during a wooden stud pick-and-place task featuring complex geometric dimensions.

Keywords –

Industrialized Construction; Prefabrication; Robotics; Automation; Modular Construction

1 Introduction

Industrialized construction (IC) and prefabrication have gained significant traction over the past few decades primarily involving the off-site assembly of structural components, which are subsequently transported and installed at the construction site [1]. However, several persistent challenges continue to hinder the widespread adoption of IC, including issues related to design and quality, as well as limitations in technology adoption and the readiness of current practices [2]. In response, numerous studies have attempted to integrate advanced technologies into this sector. Among these advancements, the introduction of robotics into the IC is particularly promising, as it can substantially reduce both construction costs and time, in addition to enhancing labor productivity by alleviating or even eliminating labor-intensive tasks [3].

In this context, laying the groundwork for the incorporation of robotics into industrialized construction, thereby advancing it toward industrialized methods, has been acknowledged as a feasible outcome. Several robotic stations have been developed and assessed to facilitate the sector's transition toward industrialization and manufacturing [4]. Additionally, research has explored the use of robotic arms as components within mass production (or manufacturing line) stations [5]. Despite the recognized enhancements in productivity achieved through the integration of robotic arms within these sectors, substantial barriers continue to impede the broader implementation of robotic stations in IC practice.

These impediments are rooted in the unique attributes of raw construction materials (e.g., wooden studs), which display characteristics—particularly dimensional asymmetries in their geometries—that diverge significantly from the standardized components commonly utilized in robotic manufacturing lines [6]. Such discrepancies frequently often lead to unpredictable, and potentially hazardous, robotic movements within the manufacturing environment when performing collision-free path planning while carrying construction elements.

In response to these challenges, various studies have introduced new and innovative path-planning algorithms [7] or have developed robotic stations adapted from commercial solutions [5], [8]. Such planning pipelines, commonly referred to as inverse kinematic path planning pipelines, are designed to compute and generate collision-free trajectories for a robotic arm, transitioning from its initial configuration to a specified goal state while avoiding contact with environmental obstacles.

However, these newly proposed algorithms generally

lack robust evaluation metrics, often relying solely on binary success-failure criteria, which stand in contrast to the algorithms widely tested using the existing simulation packages and software like MoveIt [9] and Isaac Sim [10]. These simulation platforms are designed to visualize the collision-free trajectories of robotic arms using diverse path-planning algorithms, thereby ensuring that the real robotic arm can replicate these validated movements efficiently once deployed. Moreover, the commercial path planning solutions—such as the ABB path planner integrated into RobotStudio®—have not demonstrated sufficient practicality when handling construction components. This shortcoming arises because these proprietary solutions typically do not support online predictive planning or widely tested adaptation.

To overcome these limitations, this research aims to establish a standardized validation framework capable of evaluating a variety of robotic arm path planning algorithms within IC strategies. By selecting validation metrics that align with key performance criteria (e.g., safety, efficiency, and real-time responsiveness), this framework can systematically prioritize each metric in relation to specific testing scenarios (e.g., performing pick-and-place operations on construction components) by comparing nominated path planning algorithms. As a result, these ranked metrics serve as a reliable benchmarking tool, enabling informed selection of appropriate path planning algorithms for each IC task—an essential step in effectively integrating robotic technologies into the IC sector.

2 Literature Review

Previous investigations into path planning algorithms for robotic applications have typically utilized a range of benchmark scenarios, varying in complexity and evaluation metrics. In some of the earliest works, for instance, Zhao et al. [11] presented a straightforward scenario in which a robot moves from a start to a goal configuration without incorporating either environmental complexity or object manipulation tasks. Subsequent other studies introduced simple geometric obstacles (e.g., spheres and cubes) to evaluate path planning performance [12], [13], while others extended this approach by incorporating more complex obstacles, such as shelves [14], or by progressively increasing scene complexity from cuboidal obstacles to shelves [15]. Despite these advancements, two significant limitations are apparent in the existing literature. First, although dimensional asymmetry and shape complexity are critical factors when considering objects to be manipulated, most studies have not fully integrated picking and placing tasks into their test scenarios. As a result, the reported evaluation metrics—often limited to basic measures—may not reliably represent conditions encountered in robotic IC environments. Second, although some research has examined more intricate obstacle arrangements, these scenarios still fail to approximate the full complexity of manufacturing lines associated with IC strategies [5]. Accurately reflecting the intricacies of such environments is essential for ensuring that path planning algorithms are not only collision-free but also contextually suitable for real-world assembly tasks.

Several studies have begun to acknowledge these issues by introducing pick-and-place tasks into their experimental setups, although the objects remained relatively simple and symmetrical [6], [16]. To bridge this gap, [7] proposed a series of case studies involving pick-and-place operations with progressively more complex scenes, approaching the richness found in IC environments. Nonetheless, their evaluation remained restricted to fundamental metrics such as planning time and success rate, leaving room for more comprehensive assessments. Another noteworthy contribution is found in Metvaei et al.'s work [5], which presented a robotic station capable of automatically performing component pick-and-place tasks for manufacturing structural wall panels. However, this framework relied on RobotStudio® and an offline conversion of paths into RAPID code, limiting adaptability and responsiveness. In IC workflows, the ability to dynamically adjust paths is crucial due to component tolerances and variations that arise on the production floor. Offline programming and proprietary software constrain real-time adaptability, potentially leading to misalignments and decreased product quality. These findings underscore the necessity of employing open-source path planning algorithms that offer greater flexibility and integrative evaluation. By doing so, researchers can more thoroughly assess a range of metrics—beyond mere planning time and collision avoidance—ensuring that the proposed solutions are well-suited to the demanding conditions of robotic IC tasks.

3 Objectives

This study aims to address the existing limitations in evaluating path planning algorithms within robotic IC tasks by developing a standardized evaluation framework. By integrating this framework, researchers can systematically assess the robustness and effectiveness of various path planning algorithms across construction tasks with differing complexity levels. To achieve this, the proposed framework will be applied to a simulated robotic station equipped with an ABB IRB6620 robotic arm featuring a custom multifunctional end effector. A representative pick-and-place operation involving wooden studs is employed as a case study to illustrate the framework's utility, enabling the comparative analysis of path planning algorithms by informing the refinement of

validation criteria.

4 Methodology

4.1 Overview

The proposed framework is structured into three consecutive phases, each designed to collectively yield a robust benchmarking tool for assessing path planning algorithms in IC tasks. As illustrated in Figure 1.A, the first step focuses on deriving a comprehensive set of evaluation metrics. These metrics are obtained by synthesizing those previously employed across various studies, each of which has assessed one or multiple path planning algorithms. The selection of metrics is guided by predetermined qualitative criteria, prioritizing those that best align with the requirements and constraints of the designated robotic testing environment. The second step (Figure 1.B) entails the systematic identification and selection of experimental procedures to serve as evaluation case studies. This involves deconstructing the overall strategy for producing construction assemblies into a series of discrete tasks that highlight critical points of interaction between robotic systems and construction components. Among these identified tasks, those that expose the limitations of previously studied scenarios are chosen to evaluate the candidate path planning algorithms. Finally, in the third step (Figure 1.C), each selected path planning algorithm is assessed through direct observation of its generated motions. By comparing these observed trajectories against the established qualitative criteria from the first step, it becomes possible to assign appropriate weights to the extracted metrics. These weighted metrics thus enable a nuanced evaluation of path planning algorithms, ensuring that their performance can be rigorously compared and optimized for diverse qualitative requirements across a range of IC tasks.

4.2 Evaluation Metrics Selection

Although no standardized benchmarking tool currently exists for evaluating the efficiency of various path planning algorithms, numerous studies have employed diverse metrics to compare newly proposed algorithms with existing methods or to benchmark different planning pipelines against each other. To establish a more consistent framework, this section not only introduces a metric selection pipeline to be used as the evaluation metrics in testing scenarios, but also compiles a reference of metrics commonly found in the literature (Table 1). By providing researchers with a streamlined approach for selecting appropriate evaluation measures, this work facilitates more meaningful comparisons of path planning pipelines within various case studies involving the implementation of robotics in the IC sector.

The fundamental evaluation metric employed in many path planning studies is the binary assessment of whether a given pipeline succeeds or fails in solving a case study scenario. For instance, You et al. [17] introduced a GPT-based robotic task solver that uses ChatGPT4.0 to interpret provided 2D and 3D data—such as the initial state of a Hanoi Tower Puzzle—and

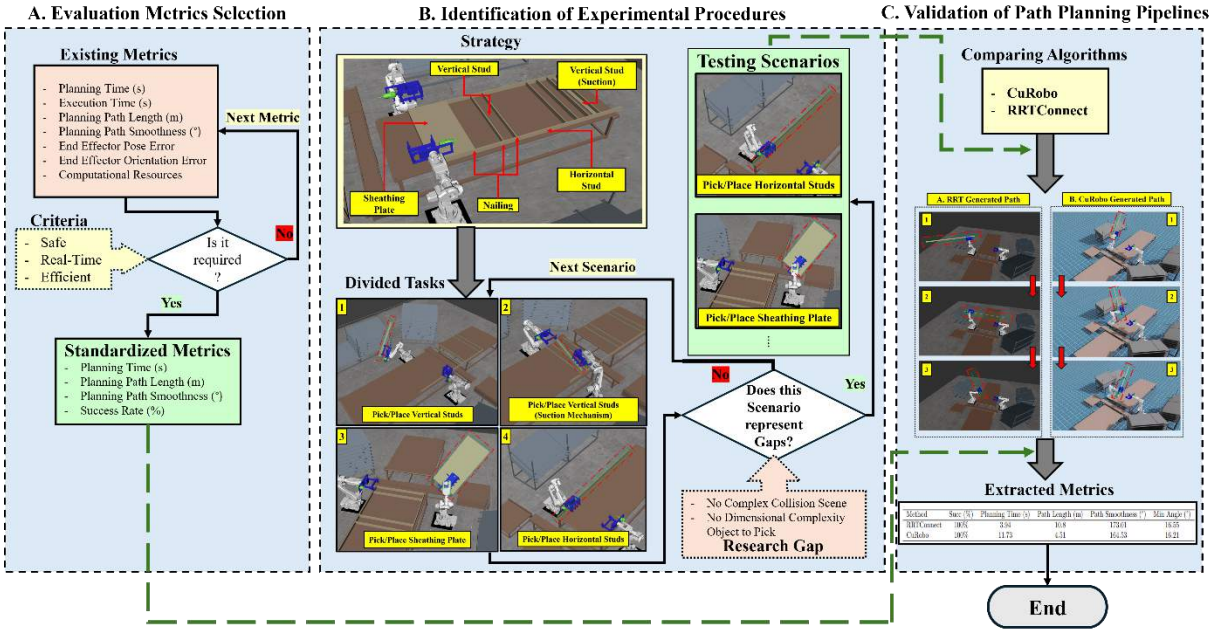


Figure 1. Methodology Overview.

generate a sequence of movements. These movements are then integrated into a ROS-based path planning framework to execute the robot's tasks. Although their work offers an initial exploration of employing Large Language Models (LLMs) for identifying task sequences in IC robotic assembly, it does not validate the path planning algorithm itself. Beyond the fundamental success/failure metric, other measures—including the accuracy of pick-and-place operations—would be valuable for such construction-oriented applications. In a more comprehensive comparative study, Ying et al. [6] developed a deep learning–based path planner optimizer and evaluated its impact on several path-planning pipelines using two principal metrics of planning time and the travel distance of the robotic tool. Similarly, Yi et al. [13] examined these metrics—planning time and path length—when proposing an enhanced P_RRT* [18] path planner. They utilized the average number of sampling nodes as a metric, reflecting computational resource demands [12]. However, within robotic IC tasks, where computational resources are less constrained, this measure is less relevant. Other studies have also considered cost function optimality, particularly in reinforcement learning–based planners [16], [19]. While cost optimality is crucial for path-planning algorithmic evaluations, it does not substantially impact IC tasks, where metrics affecting path characteristics are more informative.

In addition to these established metrics, Liu and Zuo [14] introduced path smoothness as another critical evaluation criterion. This concept involves assessing the continuity of directional changes along the generated trajectory. While they did not formally quantify smoothness for direct algorithmic comparison, their visualization underscored its importance, indicating that smoother directional changes can enhance both safety and efficiency. Building upon this notion, Zhou et al. [16] proposed a quantifiable smoothness metric derived from three consecutive path points and the angle formed between them; values approaching zero indicate trajectories closer to a straight line, thus denoting a smoother path. Furthermore, they introduced "path clearance" as an additional metric, defined as the average distance between each path point and the nearest obstacle or invalid state. Larger clearance values correlate with safer and more reliable paths.

Moreover, some studies have considered additional metrics, such as evaluating the end effector position and orientation error relative to the desired coordinates [11] or measuring the energy consumption of the robotic arm under different planning algorithms [20]. While these metrics offer insights into precision and efficiency in certain contexts, they are less directly applicable to the nuanced demands of IC tasks.

In conclusion, while certain metrics employed in previous studies may have been limited to specific robotic tasks within IC strategies, the key metrics for effectively selecting an appropriate path planning pipeline include success rate, planning time, path length (which can also reflect execution time), path smoothness, path clearance, and end effector coordination error (applicable only to real-world contexts). Accordingly, when utilizing the proposed framework to identify a suitable path planning pipeline for robotic IC applications, it is essential to select relevant metrics from the compiled reference set (Table 1)—encompassing those drawn from prior evaluation studies—based on the predetermined criteria established for comparing planning pipelines.

Table 1. Existing Metrics.

Metric	Parameter	Source
Success Rate	(%)	[6], [13], [16], [17], [19]
Planning Time	(s)	[6], [13]
Path Length	(m)	[6], [13]
Path Smoothness	(°)	[14], [16]
Path Clearance	(m)	[16]
End Effector Error	(mm)/(°)	[11]

4.3 Identification of Experimental Procedures

The overarching objective of this framework is to develop a benchmarking tool that prioritizes relevant metrics for assessing diverse path planning pipelines within different IC tasks. To achieve this, it is necessary to identify and incorporate those scenarios where the limitations of existing metrics become evident, thereby revealing gaps in current benchmarking approaches. This entails deconstructing and categorizing assembly strategies into more elementary operational tasks performed by the robotic arm. By focusing on tasks that address these identified research gaps, the testing scenarios can be enriched to more accurately reflect real-world demands. As a result, the framework can generate a metric prioritization scheme that evaluates the performance of path planning pipelines within the corresponding tasks.

4.4 Validation of Path Planning Pipelines

In the final step, the standardized metrics selected should be used to evaluate the chosen path planning pipelines. These pipelines, identified for validation, must showcase the defining characteristics of either optimal or nonoptimal solutions according to the criteria outlined by the standardized metrics. To facilitate this assessment, the trajectories generated by each selected algorithm should be visualized within a simulated environment,

allowing a comparison of the outcomes against the established criteria and the subsequent designation of each pipeline as either optimal or nonoptimal.

Moving forward, the standardized metrics identified in the literature should be calculated for each nominated path planning pipeline after generating the corresponding robotic arm trajectory in the chosen experimental scenario. By comparing the resulting metrics and ordering them according to the specified satisfaction criteria, a comprehensive benchmarking system can be established. This system will enable researchers to systematically compare any path planning pipeline within the defined testing context.

4.5 Case Study

Among the various approaches to robotic IC, the production of wooden structural assemblies has gained traction due to the relative ease of manufacturing and assembly compared to systems relying on steel or concrete components. In this study, the selected manufacturing strategy focuses on the fabrication of wooden structural panels, encompassing tasks such as the precise picking and placement of wooden studs with varying dimensions and functions, the positioning of wooden sheathing plates, and the nailing of these components. The simulated robotic workstation employs a multifunctional end effector—incorporating a gripper, suction system, and nail gun—in conjunction with conveyor belts that enhance the reachability of materials. This integrated setup enables the efficient production of wooden structural assemblies that may differ in length, include openings, and exhibit other variable design features. Figure 2.A illustrates the overall configuration of the robotic workstation, which comprises two ABB IRB6620 robots, two conveyor belts, and multiple material tables, thereby allowing the robots to access wooden studs of various lengths and thicknesses as required.

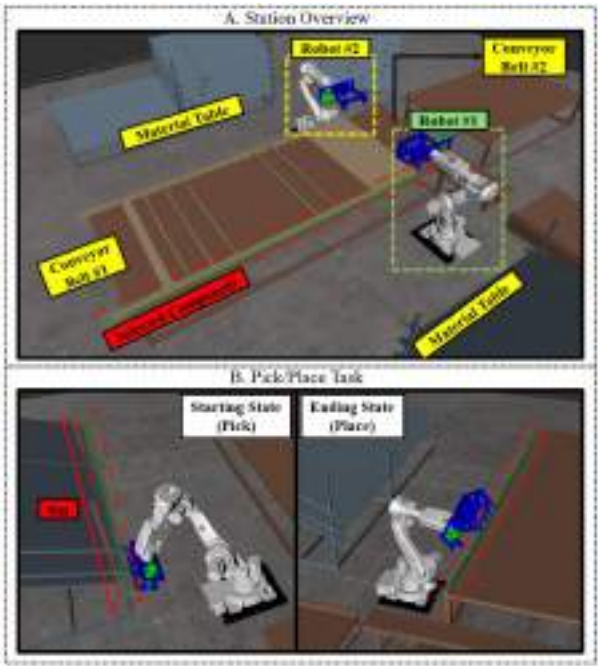


Figure 2. A) Robotic station overview. B) Selected pick/place task

4.5.1 Metrics Selection

In this study, the primary objective is to ensure the safe and reliable handling of components by simultaneously achieving the desired final configuration and minimizing potential hazards associated with both motion trajectories and tool–component interactions. To achieve this end, the most critical metrics for evaluating the effectiveness of path planning are considered success/failure, planning time, path length, and path smoothness. These metrics reflect the practical priorities of safety and efficiency more accurately than alternatives such as end effector positioning error, energy consumption, or computational optimality, which are more relevant in other domains or under different operational constraints. Furthermore, while path clearance is a common metric in standard pick-and-place tasks, it becomes less informative in scenarios involving dimensionally asymmetric construction components—such as wooden studs or sheathing plates—since the tool center point often fails to align symmetrically with the component's volume.

4.5.2 Identification of Experimental Procedures

The manufacturing of a wooden structural panel can be divided into three primary tasks, each informing the selection of path planning algorithms. The first task involves picking and placing wooden studs—both vertical and horizontal—using a robotic gripper. These studs vary in length, resulting in pronounced dimensional asymmetries that must be accommodated by the path planning strategy. The second task entails the handling of both wooden studs and sheathing plates through suction mechanisms. Although these components also exhibit dimensional asymmetries, their centers of volume are comparatively closer due to the more uniform distribution of their length and width, thereby posing a simplified set of path planning challenges compared to picking and placing wooden studs utilizing gripper. The third task centers on nailing the wooden components and involves two robots working in parallel: while one robot stabilizes the studs, the other applies the nails. Notably, this task does not include pick-and-place operations and thus does not address the dimensional asymmetry issues present in the first two tasks.

To specifically address the limitations identified in previous studies the chosen testing scenario focuses on

placing the top plate stud. The top plate represents the component with the greatest length, thereby maximizing the offset between its center of volume and the robot's tool center point. By incorporating the top plate placement task into the experimental setup, this scenario directly confronts the challenges associated with handling asymmetrical components (Figure 2.B). Furthermore, the simulated environment depicted in Figure 2.A introduces more complex spatial constraints, thereby rectifying the shortcomings of earlier studies that relied on overly simplistic obstacle configurations.

4.5.3 Validation of Path Planning Pipelines

While initially deploying open-source path planning algorithms from the MoveIt framework (OMPL library [21]), we confirmed their ability to generate collision-free trajectories. However, despite demonstrating acceptable validation metrics in previous studies, these algorithms did not consistently produce movement plans that could be characterized as safe and efficient (Evaluation Metrics Criteria). As shown in Figure 3.A, the RRTConnect-based [22] solution yielded abrupt rotational motions of the wooden stud, potentially increasing the risk of collisions in a dynamically changing environment. In contrast, the CuRobo [23] path planning algorithm—recently introduced and benchmarked favorably against other approaches—produced more stable, contextually appropriate trajectories (Figure 3.B). Based on this comparative assessment, we selected CuRobo to represent an effectively functioning path planning algorithm and RRTConnect as a pipeline for suboptimal performance within our robotic IC station. Although these two algorithms operate on different computational platforms (RRTConnect using the CPU and CuRobo leveraging GPU capabilities), planning time remains an essential performance metric. The implications of these differences in computational resources and their influence on algorithmic efficiency and suitability will be analyzed further in the Results and Discussion section.

5 Results and Discussion

In this study, 4 performance metrics were extracted to determine an appropriate path for robotic arms engaged in a pick-and-place scenario involving a 6-meter wooden stud. These metrics—Success Rate (%), Planning Time (s), Path Length (m), Path Smoothness (°), and the Minimum Angle (°) as derived from the average path smoothness—were assessed for two distinct path planning algorithms: RRTConnect (integrated with MoveIt and Rviz), serving as a motion generator misaligned with the desired criteria, and CuRobo (utilizing Isaac Sim), serving as a motion generator aligned with the desired criteria. As indicated by multiple studies evaluating path planner pipelines [15], RRTConnect is widely regarded as one of the most efficient planners in complex environments. Because offsite modular construction work zones often exhibit such complexity—particularly due to the presence of materials, components, and human workers—we selected RRTConnect as the representative graph-based planner and then compared it to a state-of-the-art planner announced by NVIDIA in 2024 to demonstrate our proposed benchmarking framework. Notably, the STOMP [24] planner was also tested but failed to compute a solution within the 60-second timeout period and was therefore excluded from our results.

Therefore, Each scenario was repeated ten times for both planners, and the results—summarized in Table 2—show that while both algorithms achieved a 100% success rate, CuRobo required, on average, more than 7.5 seconds longer to plan than RRTConnect.. Nonetheless, this additional time does not substantially impact the suitability of the generated paths for real-world pick-and-place operations, particularly given the possibility of reducing planning time by employing more powerful computational resources (e.g., GPU-based training seeds). Meanwhile, CuRobo generated significantly shorter paths, with a length of approximately 4.51 meters, less than half that achieved by RRTConnect, thus meeting the initial criteria favoring reduced path length. However, path length alone is insufficient for full

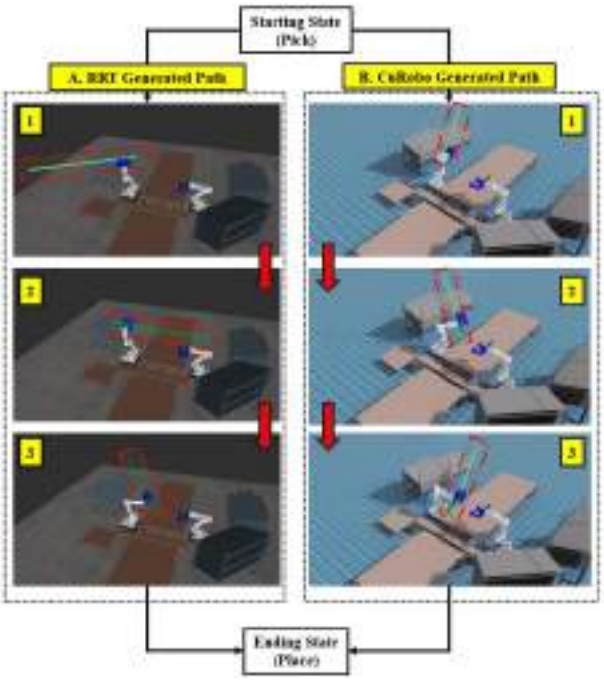


Figure 3. Representing the generated path utilizing RRT (A) and CuRobo (B) algorithms.

evaluation, and consideration must also be given to path

smoothness to avoid unsafe abrupt directional changes. Although both algorithms demonstrated negligible differences in path smoothness (less than 10° variation) and produced minimum angles of around 16°, the CuRobo algorithm ultimately proved more effective in this scenario. Accordingly, the metrics can be ranked in descending order of importance—beginning with Success Rate, followed by Path Length, then Path Smoothness, and subsequently Minimum Angle—while Planning Time emerges as comparatively less critical for the tested scenario.

Table 2. Experiment Results

Methods	Success Rate (%)	Planning Time (s)	Path Length (m)	Path Smoothness (°)	Min Angle (°)
RRTConnect [22]	100	3.94	10.8	173.01	16.55
CuRobo [23]	100	11.73	4.51	164.54	16.21

6 Conclusion

In conclusion, this study successfully identified, defined, and applied a set of weighted validation metrics to compare distinct path planning algorithms under conditions of dimensional asymmetry, which, although demonstrated here with a single robotic workstation, could be extended to various workstations and manufacturing strategies within the IC sector. Future efforts could involve categorizing these manufacturing strategies to form a comprehensive dataset of tasks for robotic stations, ultimately enabling the creation of a standardized, rule-based methodology for robotic operations in industrialized construction sector.

7 References

[1] F. M. Bademosi and R. R. A. Issa, "Automation and robotics technologies deployment trends in construction," *Automation and Robotics in the Architecture, Engineering, and Construction Industry*, pp. 1–30, 2022.

[2] B. Qi, M. Razkenari, A. Costin, C. Kibert, and M. Fu, "A systematic review of emerging technologies in industrialized construction," *Journal of building engineering*, vol. 39, p. 102265, 2021.

[3] M. Pan, Y. Yang, Z. Zheng, and W. Pan, "Artificial intelligence and robotics for prefabricated and modular construction: A systematic literature review," *J Constr Eng Manag*, vol. 148, no. 9, p. 03122004, 2022.

[4] B. M. Tehrani, S. Buhamdan, and A. Alwisy, "Robotics in industrialized construction: an activity-based ranking system for assembly manufacturing tasks," *Engineering Construction & Architectural Management*, vol. 31, Dec. 2022, doi: 10.1108/ECAM-02-2022-0143.

[5] S. Metvaei, K. Aghajamali, Q. Chen, and Z. Lei, "Developing a BIM-enabled robotic manufacturing framework to facilitate mass customization of prefabricated buildings," *Comput Ind*, vol. 164, p. 104201, 2025.

[6] K.-C. Ying, P. Pourhejazy, C.-Y. Cheng, and Z.-Y. Cai, "Deep Learning-based Optimization for Motion Planning of Dual-arm Assembly Robots," *Comput Ind Eng*, vol. 160, p. 107603, Aug. 2021, doi: 10.1016/j.cie.2021.107603.

[7] C.-H. Yang and S.-C. Kang, "Collision avoidance method for robotic modular home prefabrication," *Autom Constr*, vol. 130, p. 103853, 2021.

[8] A. Alwisy, "Framing and installation robotic systems and methods," WO2023150603A1, Aug. 10, 2023 Accessed: Dec. 18, 2024. [Online]. Available: https://patents.google.com/patent/WO2023150603A1

[9] S. Chitta, I. Sucan, and S. Cousins, "Moveit![ros topics]," *IEEE Robot Autom Mag*, vol. 19, no. 1, pp. 18–19, 2012.

[10] V. Makoviychuk *et al.*, "Isaac gym: High performance gpu-based physics simulation for robot learning," *arXiv preprint arXiv:2108.10470*, 2021.

[11] G. Zhao *et al.*, "A Tandem Robotic Arm Inverse Kinematic Solution Based on an Improved Particle Swarm Algorithm," *Front Bioeng Biotechnol*, vol. 10, p. 832829, May 2022, doi: 10.3389/fbioe.2022.832829.

[12] L. Jiang, S. Liu, Y. Cui, and H. Jiang, "Path Planning for Robotic Manipulator in Complex Multi-Obstacle Environment Based on Improved_RRT," *IEEE/ASME Transactions on Mechatronics*, vol. 27, no. 6, pp. 4774–4785, 2022, doi: 10.1109/TMECH.2022.3165845.

[13] J. Yi, Q. Yuan, R. Sun, and H. Bai, "Path planning of a manipulator based on an improved

P_RRT* algorithm," *Complex & intelligent systems*, vol. 8, no. 3, pp. 2227–2245, 2022.

[14] Y. Liu and G. Zuo, "Improved RRT Path Planning Algorithm for Humanoid Robotic Arm," in *2020 Chinese Control And Decision Conference (CCDC)*, 2020, pp. 397–402. doi: 10.1109/CCDC49329.2020.9164659.

[15] S. Liu and P. Liu, "Benchmarking and optimization of robot motion planning with motion planning pipeline," *The International Journal of Advanced Manufacturing Technology*, vol. 118, Jan. 2022, doi: 10.1007/s00170-021-07985-5.

[16] D. Zhou, R. Jia, H. Yao, and M. Xie, "Robotic Arm Motion Planning Based on Residual Reinforcement Learning," in *2021 13th International Conference on Computer and Automation Engineering (ICCAE)*, 2021, pp. 89–94. doi: 10.1109/ICCAE51876.2021.9426160.

[17] H. You, Y. Ye, T. Zhou, Q. Zhu, and J. Du, "Robot-Enabled Construction Assembly with Automated Sequence Planning Based on ChatGPT: RoboGPT," *Buildings*, vol. 13, p. 1772, Jul. 2023, doi: 10.3390/buildings13071772.

[18] S. Karaman, M. R. Walter, A. Perez, E. Frazzoli, and S. Teller, "Anytime motion planning using the RRT," in *2011 IEEE international conference on robotics and automation*, IEEE, 2011, pp. 1478–1483.

[19] T. Lindner, A. Milecki, and D. Wyrwał, "Positioning of the Robotic Arm Using Different Reinforcement Learning Algorithms," *Int J Control Autom Syst*, vol. 19, no. 4, pp. 1661–1676, 2021, doi: 10.1007/s12555-020-0069-6.

[20] K. Nonoyama, Z. Liu, T. Fujiwara, Md. M. Alam, and T. Nishi, "Energy-Efficient Robot Configuration and Motion Planning Using Genetic Algorithm and Particle Swarm Optimization," *Energies (Basel)*, vol. 15, p. 2074, Mar. 2022, doi: 10.3390/en15062074.

[21] I. A. Sucan, M. Moll, and L. E. Kavraki, "The open motion planning library," *IEEE Robot Autom Mag*, vol. 19, no. 4, pp. 72–82, 2012.

[22] J. J. Kuffner and S. M. LaValle, "RRT-connect: An efficient approach to single-query path planning," in *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No. 00CH37065)*, IEEE, 2000, pp. 995–1001.

[23] B. Sundaralingam *et al.*, "Curobo: Parallelized collision-free robot motion generation," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2023, pp. 8112–8119.

[24] M. Kalakrishnan, S. Chitta, E. Theodorou, P. Pastor, and S. Schaal, "STOMP: Stochastic trajectory optimization for motion planning," in *2011 IEEE International Conference on Robotics and Automation*, 2011, pp. 4569–4574. doi: 10.1109/ICRA.2011.5980280.

Exploring Extreme 3D Printed Clay Overhangs with Integrated Active Hot Air Drying System on Robotic Arm

Mohamed Abdelaziz[1], Ilija Vukorep[1], Deena El-Mahdy[2]

[1]Faculty of Architecture, Civil Engineering and Urban Planning, BTU Cottbus-Senftenberg, Germany
[2] Faculty of Engineering, The British University in Egypt, Egypt
abdelmoh@b-tu.de, ilija.vukorep@b-tu.de, deena.elmahdy@bue.edu.eg

Abstract –

3D printing has gained significant popularity in construction in recent years, along with a growing focus on earth-based materials. These materials must adhere to certain principles to ensure printability and extrudability, allowing for good consistency during the printing process. This would need extra water, resulting in a wet mixture that could decrease the production printing rate per day and limit overhang structures due to the wet materials.

One challenge facing these materials is their slow drying process, especially in low temperatures and humid climates. However, the strength and stiffness of these materials are gained by time when water evaporates. Given the limited studies on the drying process during 3D printing, this research aims to explore the structural stability of maximum support-free overhangs using a custom-designed heating system based on Active Hot Air Drying System (HADS) that mimics sun drying. This system accelerates and aids the drying process of the layers, ensuring greater structural stability without deformation.

According to the results, the contentious heating system increased layer stability by reaching more layers at once. The maximum overhang reached was an angle of 50 degrees before failure at 60 degrees. In conclusion, the research highlighted and promoted the use of earth-based materials and offered some solutions to ensure that printed buildings could be dried fast. This is a solution to establish earth as a reliable material in structures in a contemporary way with the use of 3D printing.

Keywords –

Additive manufacturing; automation; 3D printing; earthen materials; support-free overhang

1 Introduction

3D printing (3DP) has been gaining more popularity in the construction sector for its ability to reach complex customized geometries without any formwork or support structure [1]. When it comes to using 3DP techniques in construction, the effect of the environment on the materials is a crucial topic to consider regardless of the drying process, in which both humidity and temperature play the main role. Most earthen-based materials used in 3DP require more water than conventional construction techniques such as rammed earth, adobe, and cob, which allows good printability and extrusion. Thus, this could be a problem for cold climates where extra time will be needed until the materials dry and the water vapour. In addition, the wet material would limit the amount of printed layers per day, which might decrease the production rate. Complex geometries are another limitation while the materials are wet. For instance, overhanging geometry, as a self-supported structure, is one of the main challenges in 3DP, as certain overhanging parts can compromise structural stability, especially when using slow-drying materials like earth. This typically limits design flexibility, reduces printing speed, and consequently lowers the daily production rate of printed layers.

Many researchers have investigated the maximum self-supporting overhanging geometry of earthen-based materials for 3DP, finding that the angle typically ranges between 17 and 20 degrees before failure [2, 3, 4]. The importance of studying the material's rheology allows controlling the climatic conditions that affect their internal bonding, which is responsible for the mechanical strength [5]. In 3DP with concrete, the material dries quickly, generating more layers per day and increasing the production rate. This occurs when concrete is exposed to air and reacts with chemical agents, such as admixtures and plasticizers. These additives alter the chemical formulation, resulting in a stiffer bonding material [5]. However, the concrete faces high drying shrinkage due to the high exposure, especially when the aggregates are not carefully selected [6]. When earth-based materials are exposed to air, they become stiffer over time as water evaporates, while the chemical composition remains unchanged [5]. Although cementation materials have been explored for their durability and the effects of additives added, yet, limited

research covered the drying process of earthen materials during printing. Research done by [5] focused on the ”Forced Layer Drying” (FLD) technique to increase the mechanical properties of the wet mortar they used during printing. The results concluded that airflow, temperature, and geometry are factors affecting the evaporation rate, mechanical properties, and structure stability of the whole printed layers.

This research explores the feasibility of a custom-designed Active Hot Air Drying System (HADS) that aids in drying the layers during printing. The goal is to achieve the maximum stable overhanging angle with earthen materials, ensuring structural stability. The research method is tested on eight different inclination angles (Figure 2) of a modular overhanging arch using 3DP with clay as shown in Figure 1.



Figure 1: Overhanging example reaching 33° visualization inside 10mm grid. The object is 150 mm high and varies in length and depth depending on the overhanging angle

The printing and structural stability of the prototypes are tested both with and without an HADS. All the printing speed and flow rates remained constant. The results show that the heating system installed on both sides of the nozzle allows for achieving more overhanging arches at wider angles.

2 Related Work - 3DP drying process of earthen materials

Earthen materials typically require additional drying time to allow water to evaporate, unlike concrete. In 3DP, this challenge can impact daily productivity as layers rely on sunlight for drying. Unlike filament-based 3D printing, which requires cooling, the travel time between layers can affect bonding, leading to weaker connections and a drop in temperature [7].

Earthen material mechanical performance can be enhanced through several aspects; 1) adding admixture, alcohol, or alginate [5], 2) reducing the water content, which then could affect the extrudability of the layer. Yet both factors are affected by the climate, for example, a hot-arid climate allows water to evaporate faster than in a cold climate. Thus, the buildability will limit the printed layers from holding the upper layer until it dries and gets strength. It is important to mention the conflict between both printability and buildability for 3DP where pumping efficiency requires more wet materials which results in a lower strength, while buildability requires a material with a higher strength [5].

Architects are seeking designs that are innovative and cost-effective for their material optimization and stability without the need for any formwork. A research done by [8] produced a double-curvature façade panel by printing clay on a developed non-planar surface using a multi-forming machine with a movable actuator. This gives the flexibility to adapt the surface based on the customized geometry for each different tile without any waste formwork.

Research done by Motamedi et al. [5] implemented conductive heating through a heat gun with an 1800-watt and two heat guns with 3300 watts on two 30cm 3D printed clay blocks. The results show that the first experiment had a collapse at the 25th layer, while the second case reached higher layers with the 30th layer [5]. These results were done on a block with no overhanging consideration. A different drying process is done using infrared radiation to the deposited printed layer to allow increasing the past’s yield stress which the geometry keeps its shape [9]. Another study was conducted by [10] from IAAC, where a wooden formwork was used to support 3D-oriented clay overhanging at angles of 45°, 60°, and 75°, resulting in significant timber waste. Gomaa et al. [11] demonstrated in their research that cob, a mixture of earth and fibers, can be printed with overhangs up to 40 degrees. This finding indicates that fibers enhance the mechanical properties of the material, thereby improving structural stability. In our research, however, we will not incorporate fibers into the earth mixture.

2.1 State of the art

Building on previous studies, the effect of heat and self-supported overhanging angles on the stability of earthen materials remains unexplored. This will be the focus of our research. Accordingly, the heating process will be controlled after a specific number of layers to ensure the stability of overhanging geometries without causing cracking.

3 Method

To achieve the aim of this paper—creating structurally stable overhanging geometries with the maximum angle using earthen materials—an experimental method was developed. This method is based on a digital design-to-fabrication workflow and is divided into three main phases: geometry consideration, fabrication and assessment.

3.1 Geometry consideration and design parameters

The geometry was designed to achieve maximum overhanging through a balanced shape with a small area base of 27 x 64 mm, as shown in Figure 1 and 2. The total height is constant at 150 mm with a layer height of 1.7 mm and a total of 88 layers.

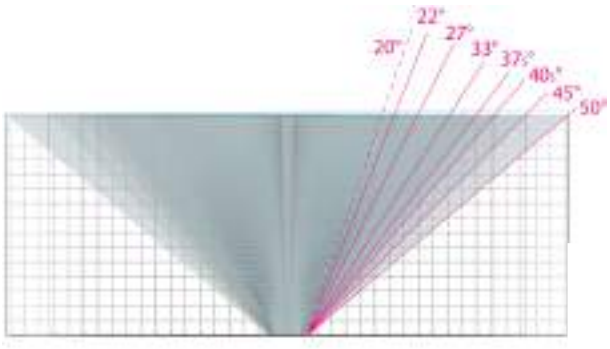


Figure 2. Inclination steps with a dotted line indicating the previously achieved maximum inclination of 20°

3.2 Fabrication

Nine prototypes were 3D printed using cla angles of 22°, 27°, 33°, 37.5°, 40.5°, 45°, 50°, and 60 assess which angle achieves the goal before failure, with and without HADS. A similar approach was ado based on the method outlined by [5]. They used tunnel experiment which act as a heating effec calculate the evaporation and drying rate of the mat while printing based on airflow speed and temperat with and without forced layer drying.

Accordingly, a HADS was added to both side the nozzle, with the assumption that it would impr structural stability by accelerating the drying of layers.

The first two angles, ranging from 17° to have been validated in previous research as the maxi achievable in clay printing [2, 3, 4]. This study, theref aims to explore new angles with the objective exceeding these limits to achieve stable geometries.

3.3 Setup

This robotic system is controlled by a computer running Rhino 3D, Grasshopper (GH), and the Machina plugin, which directly transmits URScript paths to the Universal Robot UR10. The robot is equipped with an end effector featuring a 3 mm nozzle and HADS on both sides. The clay is extruded from a cylinder cartridge by a motor, which is regulated by a controller (see Figure 4). This controller is directly connected to the UR's control unit. The parametric 3D model is sliced in Grasshopper, and the generated paths are merged and transferred via the Machina Bridge software.

3.4 Assessment

The structural stability of each prototype is measured within a setup (Figure 3 and 4) that hosts several tools. This begins with counting the maximum number of layers each prototype achieved before failure, using a grid installed behind the robot. A fixed camera was positioned in front of the model during printing. Additionally, a thermal camera was used to measure humidity, temperature, and water content in the layers to analyze the differences between each prototype.

3.5 Materials and Tools

The clay used is beige with Chamotte 20% 0 - 0,2 mm and JK Drop Spike [12] test of 3.5. A universal robotic arm (UR10) is used for 3D printing with a payload of 10kg and a nozzle size of 3mm. Flow rate was not calculated and is very depended on the nozzle size. The feed rate or the printing speed is 20mm/sec.

Two heater fans were installed on both sides of the nozzle to ensure consistent drying while printing. The heating system mimics the sun and air ventilation which increases the drying process of the layer. The heaters have not yet been designed to synchronize with the robotic arm, as each operates independently. Currently, the heater is manually activated. However, future adjustments could enable synchronization with both the robotic arm and the thermal camera, allowing for further experimentation.

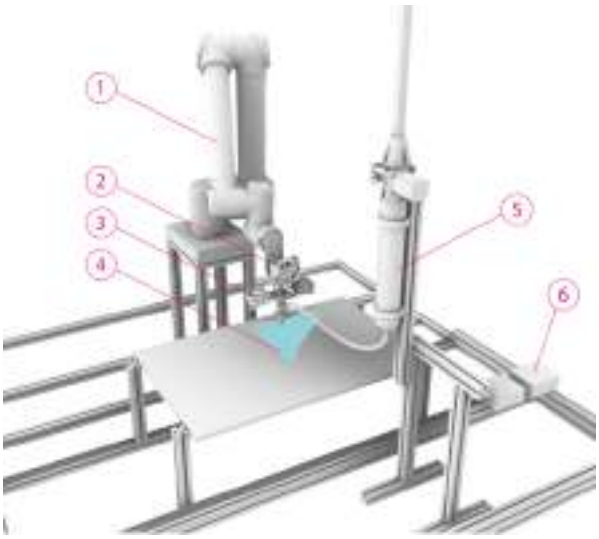


Figure 4. Experimental Setup: 1)robotic arm (UR10); 2)Screw-Extruder (custom-build); 3)PTC Fan Heater; 4)Test Object with inclined Overhangs; 5) Ram Extruder (custom-build); 6) Extruder controllers

4 Results & Discussion

4.1 The structure stability

Based on previous references, where the maximum stable inclination angle was documented as 20 degrees, it is acknowledged that this value can vary depending on the clay type and water ratio. The results showed that failure occurred at the maximum inclination angle tested. Our observations revealed that an angle of 22° remained stable without heating, while failure occurred at 27°. Table 1 compares the results for different inclinations with and without HADS, highlighting that all prints utilizing HADS were successfully completed, achieving angles up to 50°. Figure 5 illustrates the outcomes of the nine trials and the extreme overhangs achieved. Regarding stability, prototypes printed without HADS were observed to be overly wet, leading to loss of stability and self-deformation when moved from the bed post-printing. In contrast, prototypes exposed to HADS dried quickly, with layers bonding firmly to one another, resulting in greater stiffness. These prototypes retained their geometry and were easy to move, exhibiting behavior comparable to materials like PLA or ABS.

Although the research could have employed spiral continuous path planning, we deliberately decided to create a seam in the middle of the left overhang of the testing part. This decision was made to study the behavior of the printed object in the seam area, as shown in Figure 6. The 3DP process successfully achieved maximum overhang angles of 50°.

Table 1 Printing iterations

Angle	Drying System	Stable
22°	No	Yes
27°	No	No
27°	Yes	Yes
33°	Yes	Yes
37.5°	Yes	Yes
40.5°	Yes	Yes
45°	Yes	Yes
50°	Yes	Partially
60°	Yes	geometry fail

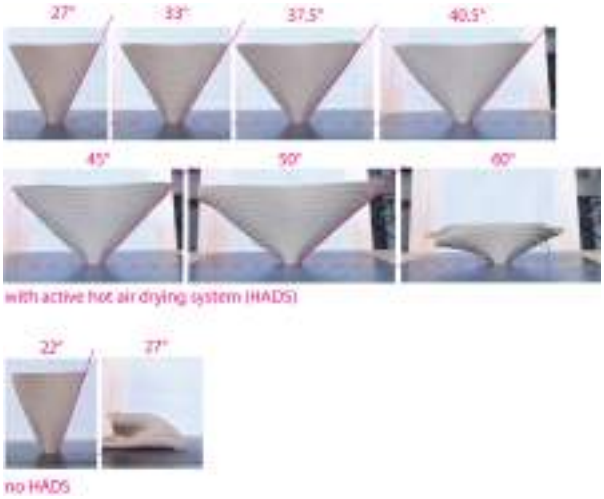


Figure 5. The different inclination angles printed with and without HADS

We assume that overhangs between 50° and 60° can be achieved which can be subject of further research. Both angles experienced failure on the seam side. This failure was attributed to the seam formed by the reverse extrusion after each layer, as illustrated in Figure 5.



Figure 6. The seam concept

This suggests that the HADS limit is 50° before failure occurs. While the 60° inclination did not result in failure, the layers did not align properly, which could be attributed to geometric limitations. At 45°, the layers overlap by 50%, indicating that further increases in inclination angles would likely lead to failure. This is due to the volumetric overlapping of the material, which shifts the geometry's center of gravity outside the footprint of its support. A similar issue could arise when validating the scalability of the prototypes on-site.

In the referenced paper by Motamedi, the overhang angle was achieved after 30 layers, heated with two heat guns, without collapsing. In contrast, our approach reached 88 layers. However, due to differences in the size of the printed material, these methods are not directly comparable. Scaling the nozzle increases both the material volume and the encapsulated humidity, requiring proportional adjustments in ventilation for effective heating. This should be further tested in future experiments.

Compared to the findings in [10], our prototypes underperformed, as their approach utilized formwork with angles ranging between 50 and 75 degrees.

4.2 Thermal Camera Temperature analysis

The monitoring of the prototypes during printing revealed a correlation between failures and the moisture levels of the layers. It was observed that the layers at the base were excessively wet, compromising their ability to support the weight of the layers above, leading to deformation. However, with the application of the HADS system, the temperature of the lower layers decreased, allowing them to solidify and better resist the loads from above. The HADS distributed heat evenly, ensuring that the entire printed body, rather than just the printed stripe, was surrounded by heated air, as shown in Figure 7. The thermal images show that the printing bed reflects heat, which supports the overall distribution of warm air. Although the surrounding area warms up to 36°C compared to the undried state at around 20°C, the object itself does not heat up and may even become cooler due to the heat-exchange effect. This assumption could be a subject for further research.

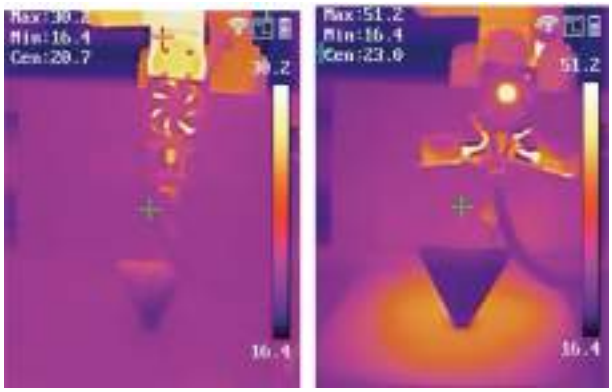


Figure 7. Thermal images. Left: 3DP without HADS. Right: 3DP with HADS

4.3 Limitation and further research

Although the printing speed was constant, it significantly impacted the structural stability especially when dealing with wet materials. Increasing the speed can lead to insufficient layer bonding, as the lower layers may remain too wet and unable to support subsequent layers. This can result in structural weaknesses and increased bucking and failure. Conversely, slower printing speeds allow for better moisture control and layer adhesion, enhancing the stability of the printed structure. Balancing print speed is, therefore, crucial for achieving optimal stability in 3DP clay components.

The room temperature was maintained at a constant 15°C; however, the same settings may exhibit different behavior when dealing with fluctuations in varying climates and humidity levels. Additionally, scaling up could impact the stability of the geometry, even if it is stable at a smaller scale. This is due to the increased volume in larger scales, which can affect the bonding between layers. Moreover, larger geometries may necessitate adjustments in printing parameters, such as speed and layer height, to ensure consistent quality and performance.

Furthermore, different clays can vary in their drying process which might cause cracks and uneven shirking problems by fast or unequal drying. Each clay should be tested and their percentage of water checked to observe the drying time and predict any cracks.

Concerning the technical issue, some failures occurred due to the seam observed on the one side of the object. Consequently, the design of the toolpath can influence the structural stability of the geometry, whether it follows a spiral path or a continuous path without reversal, as in our case.

Overall, scaling up necessitates careful consideration of these factors to ensure successful outcomes. HADS on a small scale could be durable, while on a large construction scale, we can rather find another way to absorb the humidity of the wall or ensure equal sun heat distribution across the whole building surface. The HADS process should be observed to prevent any cracking that might occur in the material.

Another recommendation for further research in large scale 3DP on-site is to add a larger heating system that can absorb the humidity from the large printed walls in the case of not controlling the sun distribution across the building surface especially when having the side of the building in the North direction where no sun is exposed on that side.

For further research, additional sensors could be integrated above the extruder to automate the HADS based on air humidity and the deformation of the printed output. Furthermore, a differentiated implementation of HADS could be explored, such as turning it off, or adjusting the temperature by reducing or increasing it on specific parts of the layer or object.

5 Conclusion

This research explored the maximum support-free overhanging angles that 3D-printed clay can achieve without additional support. Building on previous studies that reached a maximum of 20 degrees, our results surpassed this limit, achieving angles of up to 50 degrees with the implementation of the HADS. This new system ensured even heat distribution across the entire geometry, addressing challenges found in real-world applications where building surfaces, such as northern sides, are not naturally exposed to sunlight.

Our active hot air drying system, mounted on the printing head on a robot arm, ensures consistent heat distribution across the entire layer surface. This approach differs from previous studies that used handheld heat guns, which led to issues like cracking and shrinkage.

Looking ahead, this study raises further research questions, such as how the heating system would perform in humid or dry climates. Additional tests may be needed to adjust the water content in the clay and calibrate the heater temperature for optimal performance in different environmental conditions.

References

[1] Xie, F., Chen, L., Li, Z. & Tang, K. Path smoothing and feed rate planning for robotic curved layer additive manufacturing. *Robotics and Computer-Integrated Manufacturing* 65, 101967, 2020.

[2] IAAC, "Terraperforma Project," *IAAC - Institute for Advanced Architecture of Catalonia. Open Thesis Fabrication Program*, 2017. https://iaac.net/project/terraperforma/ (acc. Nov. 17, 2022).

[3] EL-Mahdy, D. & Ali, M. Assessing the solar radiation performance of self-shaded 3D-printed clay-based façades. *Architectural Engineering and Design Management* 20, pages 249–268, 2024.

[4] Cutajar, S. *et al.* Upscaling earth formworks: 3D printing strategies for material optimised reinforced concrete structures. *Constr Robot* 8, 3, 2024.

[5] Motamedi, M., Mesnil, R., Tang, A.-M., Pereira, J.-M. & Baverel, O. Structural build-up of 3D printed earth by drying. *Additive Manufacturing* 95, 104492, 2024.

[6] Mozaffari, S. *et al.* Digital design and fabrication of clay formwork for concrete casting. *Automation in Construction* 154, 104969, 2023.

[7] Hungria, E. B., Di Nisio, F. G., Cano, F. C., Voltolini, R. & Volpato, N. A study about weak intralayer bonding in extrusion-based additive manufacturing due to resumed extrusion during filling. *Polymer Testing* 140, 108595, 2024.

[8] Salem, I. Coupling Non-planar Robotic Clay Deposition with Multipoint Forming to Optimize the Manufacturing of Double Curved Façade Panels. in *Dokonal, W, Hirschberg, U and Wurzer, G (eds.), Digital Design Reconsidered - Proceedings of the 41st Conference on Education and Research in Computer Aided Architectural Design in Europe (eCAADe 2023) - Volume 1, Graz, 20-22 September 2023, pages 499–508*, 2023.

[9] Ghazanfari, A., Li, W., Leu, M. C. & Hilmas, G. E. A novel freeform extrusion fabrication process for producing solid ceramic components with uniform layered radiation drying. *Additive Manufacturing* 15, pages 102–112, 2017.

[10] Coutinho, B.,Polvi, F., and Bryson, Z., "Temporary embedded supports for cantilevers and vaults in earthen 3d printing architecture," IAAC blog, 2021. https://www.iaacblog.com/programs/temporary-embedded-supports-cantilevers-vaults-earthen-3d-printing-architecture/ (accessed Nov. 14, 2024).

[11] Gomaa, M., Jabi, W., Veliz Reyes, A. & Soebarto, V. 3D printing system for earth-based construction: Case study of cob. *Automation in Construction* 124, 103577, 2021.

[12] Keep, J., JK Drop Spike - Clay Consistency MeasuringTool.https://wikifactory.com/@jonathankeep/jk-drop-spike. (accessed Nov. 20, 2024).

Automation of the Pneumatic Caisson Method - Optimization of Excavation Using the Greedy Algorithm -

Hiroki Harada[1], Kiichiro Ishikawa[1], Taichi Terui[2], Yusen Inagawa[2], Kiyoshi Nakamura[2], Shinichi Tachibana[3], Shingo Tsugawa[3], Kotaro Nagahama[4] and Koji Terada[4]

[1] Department of Mechanical Engineering, Nippon Institute of Technology, Saitama, Japan
[2] OBAYASHI CORPORATION, Tokyo, Japan
[3] OHMOTO GUMI Co., Ltd., Okayama, Japan
[4] Preferred Robotics, Inc., Tokyo, Japan
2247024@stu.nit.ac.jp, k-ishika@nit.ac.jp

Abstract –

The pneumatic caisson method, which is used in the construction of bridge piers and other structures, is used in a variety of different sites. Currently, many processes in the pneumatic caisson method in Japan are carried out remotely. However, In Japan, the declining birthrate and aging population have created a shortage of experienced field workers. Therefore, research into the automation of the pneumatic caisson method is being actively conducted in Japan.

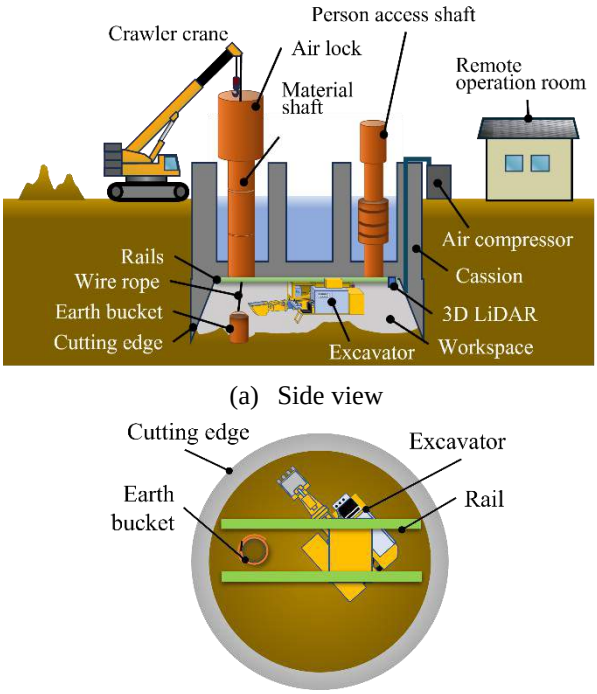
In this paper, we describe the results of experiments using an actual machine to test a method using the greedy algorithm with the purpose of automating the pneumatic caisson method. Our research group has previously proposed a method using the greedy algorithm for optimizing the excavation process. In this paper, we have reviewed the cost calculation equation and conducted experiments in a test field using an actual excavator. As a result of these experiments, we were able to achieve a construction speed equivalent to actual construction.

Keywords –

Automatic construction; The pneumatic caisson method; Actual machine test; Job scheduling; Process optimization; Greedy algorithm

1 Introduction

The pneumatic caisson method, which is a construction method for burying structures such as bridge piers and underground pumping plants, is used in a variety of different sites. In the pneumatic caisson method, as shown in Figure 1, a workspace is provided at the bottom of the concrete caisson. Rails are installed in the ceiling of the workspace, and an excavator is suspended from the rails. The excavator moves by traveling on and turning around these rails. By removing the soil at the bottom of the caisson using this excavator, the caisson sinks underground by its own weight and is buried. The excavated soil is loaded into an earth bucket and carried out aboveground through a material shaft. In the pneumatic caisson method, compressed air is blown into the workspace at the same pressure as the water pressure to prevent groundwater from entering the workspace. As a result, the pressure in the workspace increases as the caisson sinks. Construction work under high pressure increases the physical burden on field workers.



(a) Side view

(b) Diagram of workspace from above

Figure 1. Pneumatic caisson method overview

For this reason, many of the processes involved in the pneumatic caisson method in Japan are carried out

remotely. However, In Japan, the declining birthrate and aging population have created a shortage of experienced field workers. For this reason, research into the automation of the pneumatic caisson method is being actively carried out in Japan.

In our previous research, we defined the excavation field and temporary soil storage area and developed a system that loads soil into an automatically recognized earth bucket or disposes of soil at the temporary soil storage area by excavating the soil at the excavation positions within the field. We have confirmed the effectiveness of this system through experiments at construction sites [1, 2, 3, 4]. Also, with the purpose of achieving automatic excavation of the entire workspace, excluding excavation under the cutting edge, we proposed a method of optimizing the excavation position and soil disposal position given to the Automatic Excavation System using the linear programming method, and compared the results with actual construction in a simulation experiment. The results showed that the proposed method achieved good results, with a volume of soil of 28.5 m^3 taking 74 minutes, compared to 98 minutes in actual construction [5].

In this paper, we describe the results of experiments in a test field using a method for optimizing excavation positions and soil disposal positions for an Automatic Excavation System, which was developed for automatic excavation of the entire workspace, excluding excavation under the cutting edge.

2 Relational Works

In Japan, there has been a lot of research into construction automation due to the shortage of manpower in the construction and civil engineering industries caused by the declining birthrate and aging population. Much of this research has focused on automating general construction machinery used in outdoor environments such as dam construction [6, 7, 8, 9, 10]. Research related to excavation planning includes the work of TERENZI et al. [11, 12, 13]. TERENZI et al. propose a method for automating back shovel excavators that divides the wide-area excavation plan into a global plan and a local plan. The global plan is treated as a Coverage Problem, and the order of the cells is determined using graph navigation algorithms. There is also research on autonomous mobile robot behavior planning, such as the research of LI et al. [14, 15].

There has also been a lot of research into the automation of the pneumatic caisson method [16, 17, 18]. In particular, in the research by TAKADA et al. [18], a job scheduling algorithm for multiple excavators is proposed using the Munkres algorithm based on the positional relationship between the soil mountain and the excavator bucket. This method achieves cooperative work by multiple excavators using a 1/10 scale model and has achieved some results. However, the experiments using the model were conducted under optimal conditions, such as the shape of the soil mountain, and the excavation conditions did not match the actual conditions of the pneumatic caisson method. In addition, this method is one in which the location of the temporary soil storage and soil disposal is specified in advance, and the excavators are assigned to work at each location, and the definition of the temporary soil storage and soil disposal locations is done manually.

In this method, we propose a method for automatically selecting the optimal excavation positions or temporarily placing positions within the excavation field. In addition, we used an excavator that is actually used in construction for the experiment, so the experiment is in line with actual construction conditions.

3 System Configuration

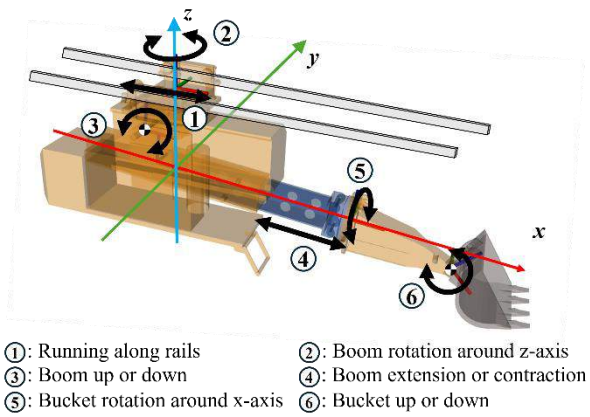


Figure 2. Diagram of excavator model

First, we will explain the excavator used in the pneumatic caisson method. Figure 2 shows a model diagram of the excavator. This excavator has six degrees of freedom: (1) running along rails, (2) boom rotation around the z-axis, (3) boom up or down, (4) boom extension or contraction, (5) bucket rotation around the x-axis and (6) bucket up or down. Basically, excavation is carried out using the (3) boom up or down, (4) boom extension or contraction and (6) bucket up or down axes, and movement is carried out using the (1) running along rails and (2) boom rotation around z-axis, and soil disposal is carried out using the (5) bucket rotation around x-axis. This excavator is operated by remote control in actual construction, so rotary encoders or potentiometers are attached to all axes. The physical quantities obtained from these are sent to the remote operation room by a wireless connection. In addition, this excavator is powered by hydraulics, as are other

construction machines. The amount of stick operation on the remote-control panel in the remote operation room is sent to the excavator by wireless connection, and the hydraulic actuator is operated by opening and closing the solenoid valve. These remote-control functions are used to automate the excavator. In addition, the excavation method of the excavator is different from that of the back shovel excavator, and it excavates by scooping up the soil from below.

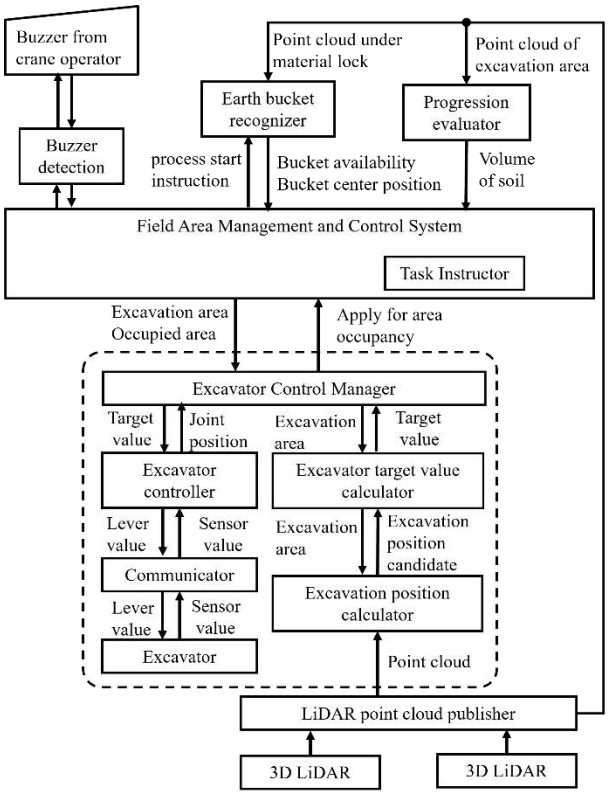


Figure 3. Diagram of the automatic excavation system configuration

Figure 3 shows an overview of the system. This system is based on the Automatic Excavation System [1] that we have been developing. It is composed of a group of software that mainly consists of the Field Area Management and Control System that manages the entire field, and a group of software that mainly consists of the Excavator Control Manager that manages the excavator. This system is a new addition to the former Field Area Management and Control System, adding a Task Instructor to instruct excavators on their tasks and where to move to. The Task Instructor instructs excavators on their tasks and where to move to, based on the volume of soil in the excavation area, the condition of the excavator and the earth bucket.

The process is as follows: First, the Excavator Control Manager requests the Task Instructor to perform a task. The Task instructor, having received the request, instructs the excavator on the work to be done, the excavation positions or the destination of the soil disposal, according to the volume of soil in the excavation area and the state of the earth bucket. Next, the Excavator Control Manager requests the Field Area Management and Control System to provide a route to the destination received from the Task Instructor. The Field Area Management and Control System then returns to the Excavator Control Manager the shortest route to the destination without coming into contact with any obstacles.

4 Methodology

In the normal excavation process of the pneumatic caisson method, excavation begins in the center of the workspace, leaving the soil near the inner wall, called the cutting edge, in place. After that, the sinking of the caisson begins by removing the soil under the cutting edge. The purpose of this research is to achieve automatic excavation of the entire central area of the workspace, excluding excavation under the cutting edge. Our research group proposed a method of optimization by replacing the excavation process of the pneumatic caisson method with a transport problem of soil [5]. The following constraints are set in this replacement:

1. Constraints on the excavator
   - The excavator can move only on rails.
   - The excavator moves by traveling and turning on the rails, at which time the speed of movement is assumed to be constant.
   - The time required for one excavation and disposal of the soil is assumed to be constant.
   - Not to excavate under the caisson cutting edge or to fill up the soil around the cutting edge.
   - The volume of soil that can be moved in one excavation depends on the size of the excavator bucket, and the excavator that our research group is targeting is 0.25 m^3.
   - When there are multiple excavators on the same rail, each one has a limited range of movement.
2. Constraints on the Earth Bucket
   - While there is no earth bucket in the workspace, soil cannot be carried out of the caisson.
   - The capacity of the earth bucket is constant at 1 m^3.
   - Only one excavator can load soil into one earth bucket at a time.
   - The time required for a full earth bucket to rise and be returned to the workspace again is

assumed to be constant. The time required depends on the length of the material shaft. Furthermore, the length depends on the amount of sinking of the caisson, but the amount of sinking is only a few tens of centimeters per day. Therefore, note that the time required does not change over one day.

- The number of earth buckets that can be loaded at one time is limited to the number of material shafts.
- The location of the earth bucket is limited to directly below the material shaft.

3. Constraints on the excavation field
   - Excavation cells are cells that divide the excavation field into grid-like sections, and the size of each cell is 0.5 x 0.5 m. Here, the cell size is set to 0.5 m to ensure that there is no excavation left when excavating adjacent cells, as the excavator bucket width is 880 mm.
   - The goal of excavation is to excavate all cells to the target depth.

We defined an objective function like Equation 1 for the above constraints. Here, N is the number of times the earth bucket is raised during construction, T_B is the time when the earth bucket is not in the workspace, T_{Ei} is the time taken for excavation and disposal of soil, T_{movei} is the time taken to move to the excavation and disposal location, and the subscript i indicates one operation.

$$\text{Minimize } T = \sum_{i=1}^{I} \left(T_{E_i} + T_{move_i}\right) + T_B N \qquad (1)$$

The objective function was solved using the greedy algorithm. The greedy algorithm is one of the methods used to solve optimization problems. The greedy algorithm obtains the optimal solution by dividing the elements of the problem into subproblems, evaluating each of them independently, and incorporating them in order of their evaluation values. For example, the greedy algorithm is used in the scheduling problem of trucks and yard cranes at quay [19, 20]. The reason we chose the greedy algorithm in this method is that it is easier to implement than other algorithms and has a fast computational time.

In previous research, we have proposed a method of dividing the excavation field into a grid, calculating the movement cost of each cell with coordinates (i, j) using Equation 2, and selecting the cell with the lowest cost, and have evaluated this method through simulation [5].

$$cost_{(i,j)} = L_{(i,j)} - (G_{(i,j)}/R) \qquad (2)$$

Here, $L_{(i,\ j)}$ is the distance from the excavation cell with coordinates (i, j) to the earth bucket, $G_{(i,\ j)}$ is the volume of soil in the excavation cell with coordinates (i, j), and R is the available capacity of the earth bucket.

However, when the cost calculation formula in Equation 2 was implemented in an actual machine and experiments were conducted, when the available capacity of the earth bucket became low, the excavator would select an excavation cell that was far away from the earth bucket and had a large volume of soil, and it did not operate efficiently.

In this paper, we will revise this movement cost calculation formula and propose a method that uses Equation 3 for excavation and Equation 4 for soil disposal. Here, $L_{\mathrm{mt}(i,\ j)}$ is the Euclidean distance from the excavation cell at coordinates (i, j) to the material shaft, $V_{(i,\ j)}$ is the volume of soil in the excavation cell at coordinates (i, j), and $D_{(i,\ j)}$ is the minimum Euclidean distance from coordinates (i, j) is the minimum value of the Euclidean distance from the coordinate (i, j) to the cell where the soil is temporarily placed, and $L_{\mathrm{rovo}(i,\ j)}$ is the Euclidean distance from the excavation cell at coordinates (i, j) to the excavator's cutting edge, and α, β, γ and δ are weighting coefficients. In addition, $L_{\mathrm{mt}(i,\ j)}$, $V_{(i,\ j)}$, $D_{(i,\ j)}$, and $L_{\mathrm{rovo}(i,\ j)}$ are normalized by their respective maximum values. Also, u is a coefficient that changes depending on whether or not an earth bucket is present and is when there is one and when there is not. The earth bucket shape is recognized from the point cloud data measured by the 3D LiDAR attached to the ceiling. When recognized, the position of the opening at the top of the earth bucket and the volume of soil loaded are obtained. The recognition program is not described in detail in this paper, but it is a high-speed version of the automatic recognition program developed by OSAKI et al. [2].

The following actions are selected based on whether or not an earth bucket is present, as determined by the recognition program:

1. When an earth bucket is placed under the material shaft.

   When excavating, the cell with the lowest cost is selected using the calculation in Equation 3. When disposing of soil, the position of the recognized earth bucket is returned.

2. When the earth bucket is not placed under the material shaft.

   When excavating, the cell with the lowest cost calculated using Equation 3 is selected. When disposing of soil, the cell with the lowest cost calculated using Equation 4 is selected. If the soil mountain is allowed to continue being temporarily placed in the same cell, it will become too high and interfere with the movement of the excavator. Therefore, if the distance from the highest point of the soil mountain to the ceiling becomes 1.9 m or

less, the soil is disposed of in the excavation cell with the lower movement cost. This value was determined with a margin of safety, as the bottom of the excavator is 1.6 m from the ceiling.

$$cost_{e_{(i,j)}} = \alpha_e {L_{mt}}_{(i,j)}^{\ u} + \beta_e {V_{(i,j)}}^{-1} + \gamma_e {D_{(i,j)}}^{u} + \delta_e {L_{rovo}}_{(i,j)} \quad (3)$$

$$cost_{d_{(i,j)}} = \alpha_d {L_{mt}}_{(i,j)} + \beta_d V_{(i,j)} + \gamma_d D_{(i,j)} + \delta_d {L_{rovo}}_{(i,j)}^{\ -1} \quad (4)$$

5 Evaluation Experiment

5.1 Experiment Overview

To evaluate this method, we conducted an evaluation test using an actual machine in a test field. Figure 4 shows the excavation field and the location of the material shaft. The excavation field is a 3 x 3 m square centered on (x, y) = (0.0 m, –1.0 m), and the target excavation depth is 0.5 m from the ground surface. The size of the excavation field was determined based on the maximum size of our test field. The size is small, but in actual construction sites where we have conducted experiments, the size has been the same or smaller than this. In this experiment, we measured the time required to excavate all the cells in the excavation field to the target depth.

The material shaft is a circle with a radius of 1.5 m centered on (x, y) = (0.0 m, 3.0 m), and the soil disposal bucket is placed within this circle. The earth bucket is placed and carried out directly below the material shaft by crane. The time from when a full earth bucket is carried out of the workspace to when it is placed in the workspace again is about 4 minutes. This is the same as in actual construction. The earth bucket can load 1 m^3 of soil at a time. In this experiment, the total volume of soil excavated is calculated based on the number of full earth buckets.

In addition, the weighting coefficients for Equation 3 are α = 0.2, β = 0.3, γ = 0.3, δ = 0.2, and the weighting coefficients for Equation 4 are α = 0.2, β = 0.0, γ = 0.5, δ = 0.3. These weighting coefficients were determined by repeated experiments.

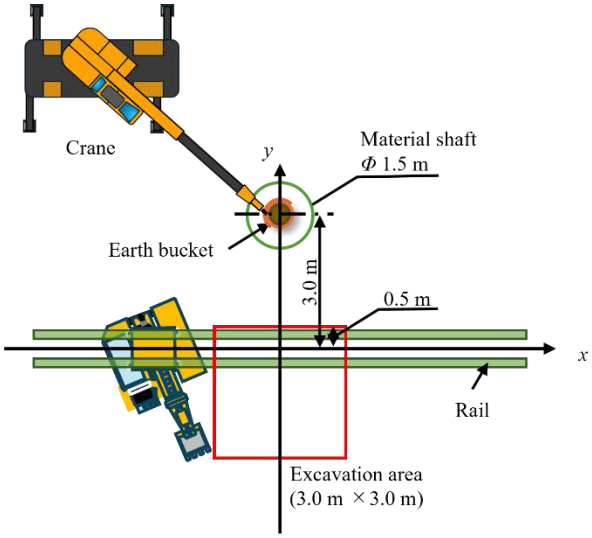


Figure 4. Diagram of Evaluation Experiment

5.2 Results and Discussion

The experiment is shown in Figures 5 and 6. The right side of both figures shows the test field. The left side of both figures shows the display screen of the Automatic Excavation System. The red frame shows the excavation field, and the green frame shows the excavation cells selected by the method. The Automatic Excavation System controls the excavator so that the blade tip is at the center of the green frame, and then performs excavation or soil disposal at that point. The heat map in both figures shows the height of the ground, with red indicating low areas and purple indicating high areas.

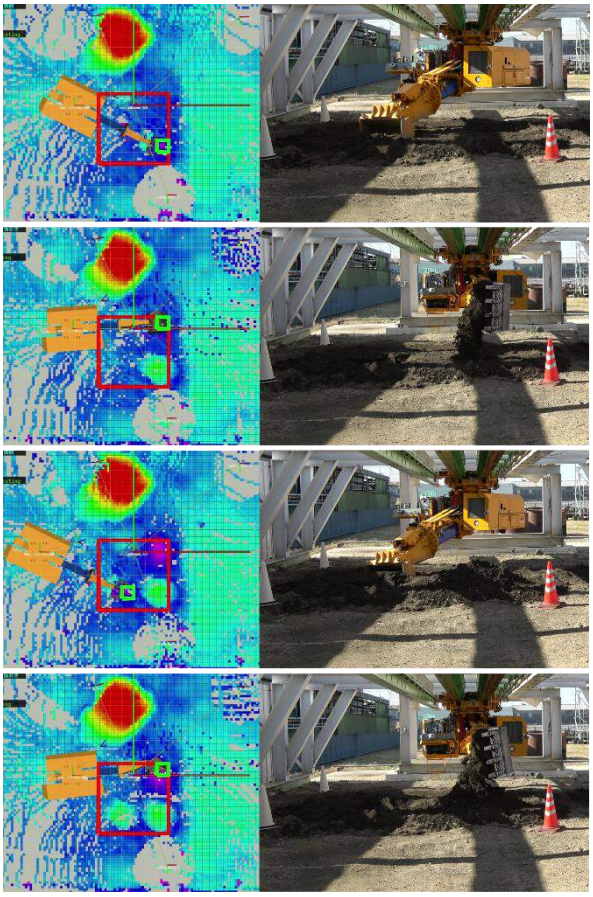


Figure 5. Experimental result when the earth bucket is not placed under the material shaft

Figure 5 shows the situation when the earth bucket is not placed under the material shaft. In the top row, the excavator is digging the cell with the lowest cost calculated using Equation 3. In the second row from the top, the excavator is temporarily placing the earth in the cell with the lowest cost calculated using Equation 4. The two bottom rows are the same as the top row, but the cell that was excavated the first time is no longer selected due to a decrease in the volume of soil. This process is repeated until the earth bucket is placed under the material shaft. Looking at the entire process, the earth in the excavation field is moved towards the material shaft. When disposing of soil in temporary soil storage, cells with high soil mountains were passed over in the calculations, and cells with low costs were selected next. As a result, the distance from the highest point of the soil mountain to the ceiling was never less than 1.6 m, and there were no problems with the movement of the excavator.

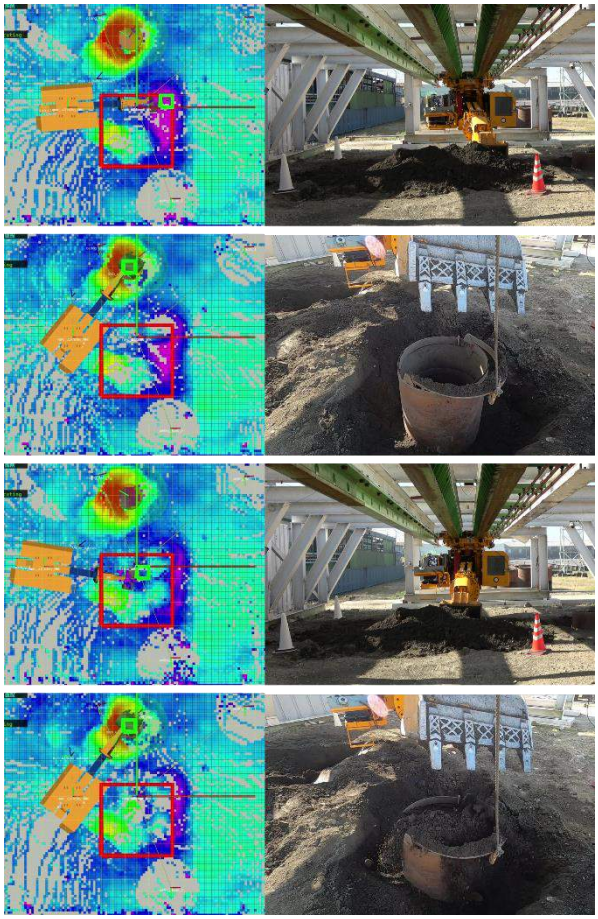


Figure 6. Experimental result when the earth bucket is placed under the material shaft

Figure 6 shows the situation when the earth bucket is placed under the material shaft. The earth is loaded until the Automatic Excavation System judges that the bucket is full. The top row in Figure 6 shows the state after the earth bucket is placed under the material shaft. The excavation cell with the lowest cost, as calculated by Equation 3, is excavated. The second row from the top shows the situation where the earth is loaded into the earth bucket. The location where the earth is directly thrown is the location of the recognized earth bucket, which also matches the actual location where it is placed. In addition, the recognition rate for the earth bucket was 100% out of the 30 times it was loaded during the evaluation experiment. The accuracy of the recognition posture was an error of 0.07 m in position and 5.73 degrees in orientation.

Table 1. Required time for each state of the excavator

Step	Case	Time [h:min:s]	Excavated soil volume [m^3]
1~6	Temporary	0:06:22	—
7~11	Disposal	0:06:27	1.00
12~16	Temporary	0:05:20	—
17~22	Disposal	0:09:42	1.00
23~27	Temporary	0:04:42	—
28~35	Disposal	0:08:49	1.00
36~40	Temporary	0:05:22	—
41~45	Disposal	0:08:19	1.00
46~49	Temporary	0:04:47	—
50~56	Disposal	0:10:20	1.00
Total		1:10:00	5.00
Actual construction process time		1:07:06	5.00

Table 1 shows the results of measuring the time required for each operation of the excavator. In this table, the process from when the excavator excavates the soil and moves it to when it is temporarily placed or directly thrown is considered to be one step. As a result, it took 56 steps and 1 hour and 10 minutes. During actual construction, the same conditions were completed in 1 hour, 7 minutes, and 6 seconds, so it can be said that effective results were obtained.

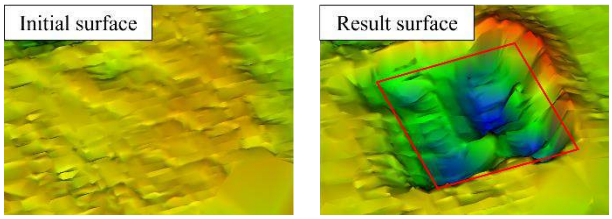


Figure 7. Ground surface before and after the experiment

Figure 7 shows the changes in the ground before and after excavation, with the left side showing the state before the experiment began and the right side showing the state at the end of the experiment. The excavator is located on the left side of both figures, and the material shaft is located on the upper side. The colors of the ground in the figures indicate the height of the ground, with blue indicating low ground and red indicating high ground. The red frame in the figure on the right shows the excavation field. Looking at the excavation field in the ground in the figure on the right, you can see that the left side of the excavation field is higher than the right side. Also, looking at the area outside the excavation field in the figure, you can see that the left side has been excavated and the right side has been piled up with soil. The standard deviation of the depth to the ground surface after excavation is 0.27 m with respect to the target depth.

Because the excavation operation of this system measures the ground height using 3D LiDAR, the bucket cutting edge is placed at that point, and excavation is carried out to the target depth from there. However, the excavation positions are not the center of the excavation cells, but the starting points of excavation in front of the excavation cells as seen from the excavator. Therefore, when the excavator excavates a cell that is further away from the excavator than the cell that has already been excavated, it is thought that the excavator will excavate deeper than the excavation cell. From this, there is still room for improvement in the excavation method of this system.

6 Conclusion and Future Works

In this paper, we propose a method using the greedy algorithm for the purpose of automating the pneumatic caisson method, and we confirmed that we achieved a construction speed equivalent to actual construction in experiments using an actual machine.

In the future, we plan to review weighting coefficients and cost calculation equations, and consider other optimization algorithms, with the purpose of minimizing the required time. In addition, in the large-scale pneumatic caisson method, multiple excavators are used. Our research group has achieved the automatic excavation and soil disposal of multiple excavators without causing them to collide with each other. Therefore, we plan to propose a method for multiple excavators.

References

[1] K. Ishikawa, H. Harada, H. Osaki, S. Tsugawa, S. Tachibana, H. Fujisawa, T. Terui, K. Nakamura and Y. Inagawa. Automatic Excavation System with Multiple Excavators in the Pneumatic Caisson Method. *Journal of Robotics and Mechatronics*, 36(4):961–972, 2024.

[2] H. Osaki, K. Ishikawa, Y. Inagawa, T. Terui, S. Tachibana and S. Tsugawa. Development of Automatic Excavation System for Pneumatic caisson method -Overhead Traveling Excavator Automation and Earthen Bucket Recognition-. In *Proceedings of the Symposium on Construction Execution and Construction Machinery*, pages 159–164, Tokyo, Japan, 2022. In Japanese.

[3] OBAYASHI CORPORATION. Demonstration of the automated excavation system for the pneumatic caisson method at the construction site of the Hanshin Namba Line in Osaka, Japan. On-line:

https://www.obayashi.co.jp/news/detail/news20230616_2.html, Accessed: 12/01/2024. In Japanese.

[4] Y. Betsui, Y. Inagawa, T. Terui, K. Nakamura, T. Ogata, H. Fujisawa, S. Tachibana, and S. Tsugawa. A Study on Overhead Travelling Excavator Automation -An Excavator Simulator Development and Bucket Trajectory Control-. In *SICE SI 2021*, pages 2088–2089, Online, Japan, 2021.

[5] H. Harada, K. Ishikawa, T. Terui, Y. Inagawa, S. Tachibana and S Tsugawa. Optimization of Excavation Process in the Pneumatic Caisson Method. In *SICE Festival 2024 with Annual conference*, pages 514–517, Kochi, Japan, 2024.

[6] L. Zhang, J. Zhao, P. Long, L. Wang, L. Qian, F. Lu, X. Song and D. Manocha. An autonomous excavator system for material loading tasks. *Science Robotics*, 6(55):eabc3164, 2021.

[7] E. Halbach, A. Kolu and R. Ghabcheloo. Automated Pile Transfer Work Cycles with a Robotic Wheel Loader. In *17th International Conference on Computing in Civil and Building Engineering*, Tampere, Finland, 2018.

[8] Y. Yang, L. Zhang, X. Cheng, J. Pan and R. Yang. Compact Reachability Map for Excavator Motion Planning. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2308–2313, Macau, China, 2019.

[9] H. Murayama, T. Kawabe, M. Inagawa, K. Yoshizawa, M. Ishibashi, T. Takei and K. Nagatani. Demonstration of Snow Removal Work by Wheel Loader in an Environment Surrounded by Obstacles. *Journal of Robotics and Mechatronics*, 35(5):1251–1266, 2023.

[10] H. Yoshida, T. Yoshimoto, D. Umino and N. Mori. Practical Full Automation of Excavation and Loading for Hydraulic Excavators in Indoor Environments. In *2021 IEEE 17th International Conference on Automation Science and Engineering (CASE)*, pages 2153–2160, Lyon, France, 2021.

[11] L. Terenzi and M. Hutter. Towards autonomous excavation planning. *IEEE Transactions on Field Robotics*, 1(1):292–317, 2024.

[12] J. Seo, S. Lee, J. Kim and S. Kim. Task planner design for an automated excavation system. *Automation in Construction*, 20(7):954–966, 2011.

[13] S. Park, J. Kim, K. Cho and J. Seo. Integrating Task Component Design in Excavator–Truck Operation Planning: Enhancing Automation Systems through a Web-Based System. *Applied Sciences*, 14(14):6052, 2024.

[14] W. Li, H. Li, Y. Wang, and Y. Han. Optimizing flexible job shop scheduling with automated guided vehicles using a multi-strategy-driven genetic algorithm. *Egyptian Informatics Journal*, 25(1):100437, 2024.

[15] B. Xin, S. Lu, Q. Wang, F. Deng, X. Shi, J. Cheng and Y. Kang. Simultaneous scheduling of processing machines and automated guided vehicles via a multi-view modeling-based hybrid algorithm. *IEEE Transactions on Automation Science and Engineering*, 21(3):4753–4767, 2023.

[16] N. Negishi, A. Iwasaki and T. Kataoka. Development of a construction management system for the pneumatic caisson method -Construction management aids and caisson excavator automation-. In *Proceedings of the Symposium on Construction Execution and Construction Machinery*, pages 126–129, Tokyo, Japan, 2023. In Japanese.

[17] N. Negishi, T. Tsuneki, K. Kikuchi, T. Kondo, T. Koyo, A. Kamei and K. Hayakawa. A Study on Autonomous Operation System of Caisson Shovels in High Air Pressure and Narrow Underground Space (Demonstration of Trajectory Tracking Control). In *The 4th International Conference on Design Engineering and Science: ICDES2017*, Aachen, Germany, 2017.

[18] K. Takada, R. Sato, S. Sato, M. Ito, R. Tsuchiya, K. Kikuchi, M. Shindo, A. Kamei and T. Kondo. Autonomous Operation System of Caisson Shovels in Narrow Underground Space at High Air Pressure (Development of a Job Scheduling Algorithm using a 1/10 Caisson Platform). *Journal of Japan Society for Design Engineering*, 57(8):381–396,2022.

[19] K. Li and L. Tang. A Tabu Search Algorithm for the Integrated Truck Scheduling and Storage Allocation in Quay. In *2009 Second International Symposium on Computational Intelligence and Design*, pages 275–278, Changsha, China, 2009.

[20] L. Kong, M Ji, and Z Gao. Joint optimization of container slot planning and truck scheduling for tandem quay cranes. *European Journal of Operational Research*, 293(1):149–166, 2021.

Human-Inspired Inverse Reinforcement Learning for Autonomous Excavator Operations in Lunar Environments

Minguk Kim[1] and Youngjib Ham[2]

[1]Department of Construction Science, Texas A&M University, USA
[2]Department of Civil and Environmental Engineering, Seoul National University, South Korea
mingukkim@tamu.edu, yj.ham@snu.ac.kr

Abstract –

Construction robots must perform complex construction tasks robustly under extreme and unpredictable conditions. Although current reinforcement learning methods, including imitation learning, show promise for this goal, they face inherent challenges, such as time-consuming reward function design or substantial data requirements that limit their generalizability to unstructured environments. These limitations are exacerbated by energy constraints of deep neural network-based algorithms, making them less viable for lunar surface construction, where data collection and power supply are significantly constrained. This study integrates inverse reinforcement learning with neuroscientific theories of human decision-making to equip construction robots with robust autonomy. The proposed approach modularizes a given task into simpler subtasks governed by learned parameters and explicitly models state features such as spatial and temporal distances, enabling the agent to anticipate and adapt to diverse scenarios while reducing computational overhead. Experimental results indicate that the agent modeled by the proposed approach achieved a 73% success rate in a lunar-like simulation environment, exceeding the 30 % success rate of an agent modeled by a conventional method. The findings demonstrate that the proposed approach can enhance adaptability for autonomous infrastructure construction, especially in lunar surface construction missions constrained by minimal human oversight and limited resources. This study contributes to the advancement of autonomous construction robotics, offering a novel and scalable solution for extreme construction settings.

Keywords –

Construction Robotics; Lunar Surface Construction; Inverse Reinforcement Learning; Autonomous Excavation

1 Introduction

Driven by prospects such as abundant resource acquisition and advances in cutting-edge technologies, the global space exploration ecosystem is shifting its paradigm toward human habitation and resource extraction on the Moon [1]. This ambitious shift hinges on developing infrastructure and capabilities to support sustainable operations beyond Earth. However, the lunar environment is harsh, dynamic, and fraught with unpredictable events [2,3], thereby necessitating the deployment of construction robots capable of minimizing human intervention while reliably performing complex construction tasks on the lunar surface.

Nevertheless, it remains challenging to imagine current commercial construction robots operating stably under such dynamic and adverse conditions. These robots have been primarily designed to perform repetitive tasks in factories or other relatively controlled and static environments, thereby depending substantially on human oversight. Indeed, human operators are often required to intervene in unexpected situations directly or engage in indirect collaborative forms, such as sequential task execution. In light of these considerations, the development of fully autonomous construction robots that can function robustly in demanding and unpredictable lunar environmental settings becomes imperative.

Reinforcement learning (RL) approaches, including imitation learning (IL), present a potential avenue to surmount these challenges, as they allow robotic agents to be trained through interactions with their environment to accomplish assigned tasks, either by exploring and exploiting independently or by utilizing cues provided by experts. By leveraging such approaches, robots can potentially deliver robust and adaptive performance, facilitating stable execution of complex construction tasks in dynamic and harsh environments [4]. However, conventional RL methods have inherent limitations, such as the time-consuming process of reward function design and an excessive dependence on training data [5,6].

Prevailing IL methods still require extensive datasets of high-quality demonstrations, even though they circumvent these limitations [7]. These constraints may prevent construction robots from overcoming scenarios that deviate from their training set or from performing reliably in environments characterized by limited resources and accessibility [8]. Moreover, many state-of-the-art algorithms, which inherently rely on deep neural networks (DNN), demand substantial energy consumption. These challenges pose significant obstacles to lunar surface construction where accessibility is highly constrained.

To mitigate these challenges, this study draws inspiration from the remarkable adaptability and responsiveness that humans exhibit across diverse conditions [9]. To embed these capabilities into construction robots, this study proposes an approach grounded in robust biological evidence, integrating a human-like decision-making architecture with inverse reinforcement learning (IRL) framework. Through this integration, the resulting construction robots will be endowed with the capacity to operate robustly under lunar environments, ultimately contributing to the realization of autonomous infrastructure construction in outer space.

2 State of the Art

RL approaches have recently gained traction in construction robotics because of their ability to learn robust and adaptive behaviors through continuous interaction with the environment ([5,10–12]). By updating policies or value functions based on trial-and-error, agents can theoretically handle dynamic and uncertain scenarios, making them appealing candidates for tasks conducted in extreme environmental settings, such as those on the lunar surface. Generally, they employ DNN to approximate policies, and in many instances, they leverage expert trajectories to enhance learning efficiency [13].

However, these predominantly policy-based RL approaches suffer from several drawbacks that can be especially pronounced in extraterrestrial construction contexts. First, they often rely on extensive, high-quality datasets—an issue exacerbated by the inaccessibility of the lunar surface, where data collection is highly constrained [5]. Second, their dependence on DNN results in high energy consumption, necessitating substantial computational capabilities in hardware [4]. Lastly, in conventional RL methods, the need for sophisticated reward function design and the challenge of sparse rewards can make it difficult for agents to obtain the necessary signals to guide their learning in a timely manner [6], particularly when tackling complex, long-duration tasks under extreme conditions. As a result, these challenges can impede the scalability and reliability of construction robots, preventing them from swiftly attaining the necessary performance for diverse and uncertain tasks, such as those encountered in the lunar environment.

In contrast, IRL has not yet been widely explored within construction robotics despite its promising potential to provide computational models of human and animal behavior [14]. Unlike policy-based RL approaches, IRL aims to uncover the underlying reward structure from expert demonstrations, thereby distilling implicit cues that can guide learning toward higher-level objectives [15,16]. This characteristic of IRL makes it especially attractive for complex tasks where reward design is difficult and for scenarios where sample efficiency is paramount due to data collection constraints. In lunar construction, these strengths of IRL could enable robust adaptation to diverse or novel circumstances, alleviating the burdens of large-scale data gathering and intricate reward engineering.

A particularly relevant extension is modular IRL, which draws from neuroscientific findings indicating that humans decompose tasks into simpler subtasks that can be evaluated separately [17]. By framing a complex construction activity as multiple subtasks, the approach can assign simplified Q-values to each subgoal. These modular Q-values can then be estimated using Bayesian IRL [18]. For instance, Zhang et al. [19] demonstrated the potential of such an approach for reproducing human sensorimotor decisions by incorporating a discounting factor that regulates impulsive behavior in virtual reality settings. Compared to policy-based RL approaches, this value-based technique shows greater sample efficiency and a reduced computational footprint [4], making them particularly well suited for physically and energetically constrained environments like the Moon. Nevertheless, while modular IRL has proven its capabilities primarily in spatial navigation tasks, its broader application to industrial contexts, including automation in construction, has yet to be fully explored.

In summary, contemporary research into RL-based robotic control indicates great potential for autonomous operation in construction tasks. Still, multiple hurdles, ranging from data scarcity to energy constraints, must be addressed to realize robust performance in lunar environments. In this regard, an IRL-based approach integrating human-like decision-making processes can be a promising way to overcome the hurdles. Building on these insights, the next section proposes a novel approach that integrates modular IRL to endow construction robots with robust, adaptive capabilities suited to the rigors of lunar surface construction.

3 Methods

Lunar construction imposes unique constraints due to limited on-site supervision, power restrictions, and complex terrain. To demonstrate the proposed IRL-based approach under these demanding conditions, this study focuses on an excavator swing operation, a frequent maneuver for material transport and debris clearance. Continuous horizontal swinging has been identified as a primary source of struck-by hazards on Earth-based sites, and this risk is magnified in off-Earth contexts where immediate human intervention is challenging [20]. As such, the excavator swing task represents an ideal testbed for validating an autonomous decision-making scheme that effectively balances safety and task efficiency.

3.1 Problem Formulation

Initially, the excavator's swing is modeled using a Markov Decision Process (MDP):

- States (S): Each state describes the excavator's relevant kinematic features (joint angles and velocities of boom, stick, cab) and key environmental cues (e.g., nearby obstacles, target locations).
- Actions (A): Discrete maneuvers include moving the boom up/down, the stick in/out, rotating the cab left/right, or performing no movement. Actions that exceed joint limits or risk an immediate collision are excluded from consideration.
- Reward (R): A single reward function penalizes collisions and rewards movement toward a defined dump area, implicitly treating all safety and positioning objectives as one.
- Discount Factor (γ): A scalar in (0,1) that weighs future returns; γ near 1 values cautious, long-term strategies, whereas smaller γ emphasizes immediate outcomes.

A SoftMax policy, as shown in Equation (1), governs action selection:

$$\pi(s,a) = \frac{\exp\left(Q(s,a)/\eta\right)}{\sum_{a' \in A} \exp\left(Q(s,a')/\eta\right)} \quad (1)$$

where $Q(s,a)$ is the action-value function, and $\eta \in (0,1]$ manages exploration versus exploitation. Although this single-task MDP is straightforward, it does not fully capture human-like strategies that partition tasks into sub-goals and consider spatial and temporal dimensions when avoiding collisions.

3.2 Task Modularization with Explicit State Features

Drawing on empirical findings in human motor control and decision-making [17,19], a main task is replaced with multiple subtasks. For instance, the swing operation might split into:

- Obstacle Avoidance (Subtask 1): Minimizing collisions.
- Targeted Material Placement (Subtask 2): Reaching or depositing material in the designated area.

Each subtask n is assigned:

- A subtask weight $r^{(n)} \in, with\ \sum_n r^{(n)} = 1$.
- A unit reward $u^{(n)}$, reflecting baseline significance for that objective.
- A sub-discount factor $\gamma^{(n)} \in (0,1)$, enabling each subtask to handle near- vs. long-term gains differently.

Hence, the total action-value $Q(s,a)$ is shown in Equation (2).

$$Q(s,a) = \sum_{n=1}^{N} Q^{(n)}(s^{(n)}, a) \quad (2)$$

where $Q^{(n)}$ focuses on a particular sub-goal. This modularization can make the overall decision process more transparent and potentially more adaptable, as each subtask can be updated independently.

Beyond subdividing tasks, psychological distance [21] is introduced to emulate how skilled operators anticipate future collisions or time-critical steps:

- Spatial Distance: The Euclidean distance between the excavator's bucket and key objects (e.g., a rock, the dump area).
- Temporal Distance: The estimated time until a collision (or successful arrival) if a chosen action is executed.

These distances are computed at each timestep, using a brief forward simulation. By incorporating distance-based features explicitly, one can reduce reliance on DNN, thus lowering computational demand—a crucial advantage in energy-constrained lunar construction missions. This approach also strengthens safety, as the agent "looks ahead" before finalizing an action.

3.3 Parameter Estimation

The following steps outline the process of estimating subtask weights $r^{(n)}$ and discount factors $\gamma^{(n)}$ from expert demonstrations:

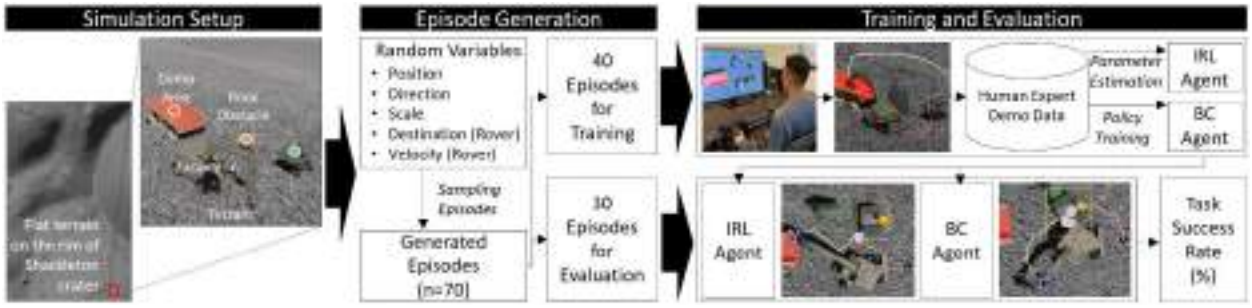


Figure 1. Overall experimental steps for training and evaluating IRL and BC agents in a simulation environment

1. Data Collection: Human experts run the excavator in a simulation, completing the swing operation multiple times under simulation environments. This yields state–action pairs $\{(s_1, a_1), \dots, (s_T, a_T)\}$, where T is the end of an episode.
2. Likelihood Computation: For a trial parameter set $\Theta = \{r_1, \dots, r_N, \gamma_1, \dots, \gamma_N\}$, the likelihood $\mathcal{L}$ of human expert demonstration data is calculated by Equation (3), where Q sums up each subtask's value function $Q^{(n)}$, weighted by $r^{(n)}$ and discounted by $\gamma^{(n)}$.

$$\mathcal{L} = \prod_{t=1}^{T} \frac{exp^{Q(s_t, a_t)\eta}}{\sum_{a \in A} exp^{Q(s_t, a)/\eta}} \quad (3)$$

3. Parameter Set Sampling: Each subtask's reward and discount factor are estimated by a sampling method. This study uses a Hamiltonian Monte Carlo (HMC) technique to explore the parameter set spaces. Key constraints, $\sum_n r^{(n)} = 1$ and $0 < r^{(n)}, \gamma^{(n)} < 1$, are enforced to ensure valid solutions.
4. Convergence to Final Parameter Set: Parameter sets aligning closely with expert actions are accepted, leading to a convergent estimate $\widehat{\Theta}$. This final parameter set grants the excavator robot agent.

4 Experimentation

4.1 Experimental Goal and Setup

The primary goal of this experiment is to determine whether the IRL agent modeled by the proposed IRL-based approach can robustly execute a construction task compared to an agent modeled by behavior cloning (BC), a conventional IL method. To do so, as shown in Figure 1, A total of 70 episodes were devised in a simulated lunar environment constructed based on real lunar terrain data [22] in the Unity game engine [23], with 40 episodes for training and 30 episodes for evaluation. Each episode included five key objects—an excavator, a dump truck, a rock obstacle, a moving rover, and the surrounding terrain, which are randomly arranged to generate varied spatial layouts and dynamics using the Latin Hypercube Sampling method [24]. To ensure dynamic interaction throughout each episode, constraints were applied during the generation process. These constraints included maintaining a minimum distance between the moving rover and other objects, as well as selecting episodes where the rover's linear path did not result in collisions with other objects, thus preserving the intended episode dynamics. Episodes are terminated upon the excavator bucket reaching the designated dump area or upon reaching the time limit of 150 timesteps, which is approximately three times longer than an expert requires. The workstation on which the experiments were conducted has the following specifications: Intel Xeon W-2123 CPU, 64 GB RAM, NVIDIA GeForce RTX 2080 GPU.

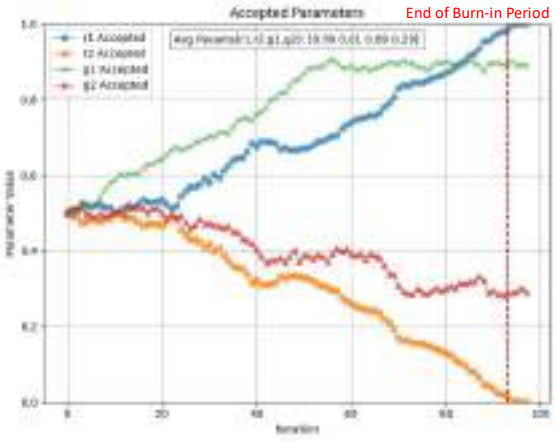


Figure 2. Accepted parameter trajectories during HMC sampling

4.2 Experimental Procedure

Human expert demonstrations were recorded in a simulation equipped with joystick controls and a first-person view. Each demonstration began with five seconds of situational awareness from an overhead perspective, allowing participants to plan their route

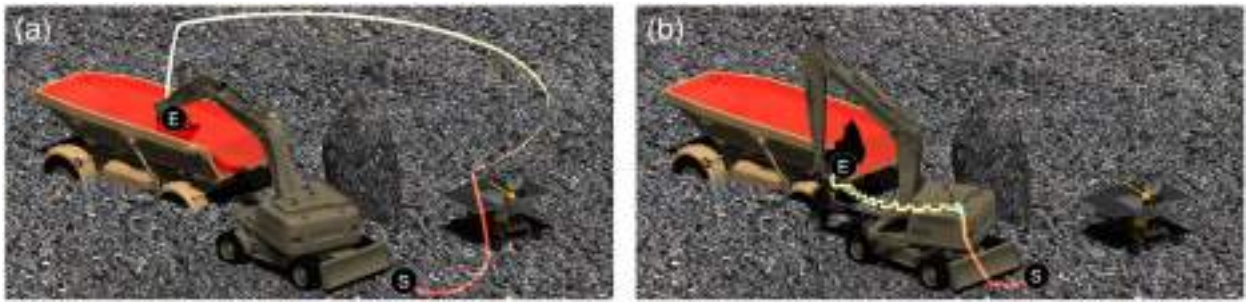


Figure 3. Trajectory comparison of a successful episode between (a) a human expert and (b) the corresponding IRL agent employing different detour strategies

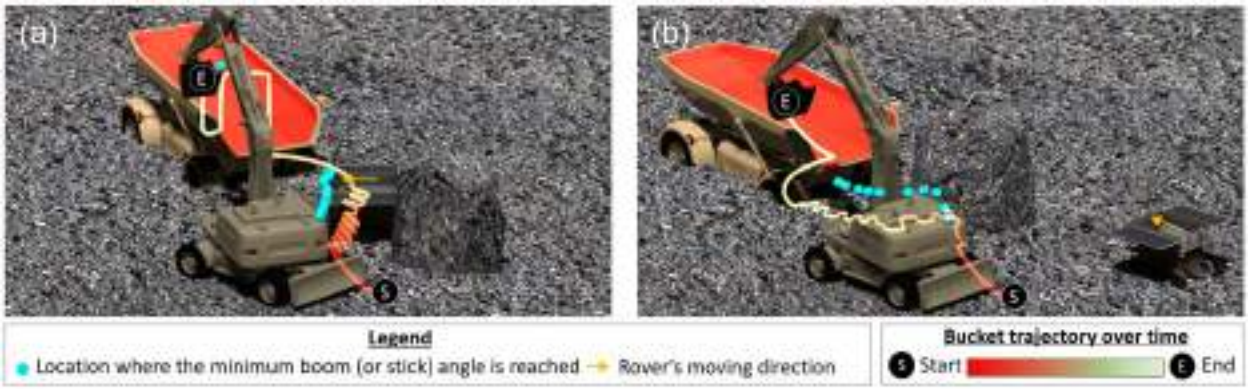


Figure 4. Failure cases of IRL agents: (a) and (b) illustrate inefficient movement leading to episode termination

[25,26]. Participants then navigated the excavator bucket to the dump area using what they perceived as the safest, most efficient path. The collected state–action pairs formed the training data for parameter estimation in the IRL agent and supervised training in the BC agent.

Afterward, the IRL agent's parameters, $\Theta = \{r^{(1)}, r^{(2)}, \gamma^{(1)}, \gamma^{(2)}\}$, were inferred from the recorded demonstration data via HMC. After discarding an initial burn-in, the final accepted samples were aggregated to define a robust parameter set. Figure 2 illustrates the evolution of these parameters over iterations, highlighting their convergence after the burn-in phase. On the other hand, the BC agent was trained using a neural network architecture with optimal hyperparameters determined through a grid search: dropout probability of 0.1, learning rate of 1e-4, neural network dimension of 256, and depth of 3.

Both agents were subsequently evaluated in the 30 remaining episodes, and the task success rate was measured as the proportion of runs where the bucket reached its target without a collision.

5 Results and Discussion

5.1 Robustness Compared to BC Agent

The IRL agent demonstrated a success rate of 73%, significantly outperforming the BC agent, which achieved only 30%. This notable difference underscores the limitations of relying solely on policy imitation, as seen in the BC agent's performance in dynamic environments. In contrast, the IRL agent exhibited superior adaptability to diverse layouts, including a moving rover. Additionally, the IRL agent achieved a traveled distance of 15.71 meters on average (S.D.=3.91), compared to the BC agent's average traveled distance of 17.98 meters (S.D.=4.61). These findings highlight that the proposed IRL-based approach enables construction robots to robustly perform construction tasks more efficiently than the existing RL methods. This finding aligns with prior literature emphasizing that conventional policy-based IL methods can struggle in unseen or high-variability contexts [8]. On the other hand, the IRL agent exhibited robust performance and more efficient path selection.

Interestingly, as shown in Figure 3, the IRL agent often exhibited detour strategies that differed from human experts. Both tended to detour around obstacles; however, the strategy employed diverged. While the human expert consistently chose the safest and farthest routes to avoid collisions, as depicted in Figure 3(a), the IRL agent opted to detour in the opposite direction, as shown in Figure 3(b). This strategic variation resulted in the IRL agent adopting a more efficient trajectory, reducing the required timesteps and minimizing redundant movements. Consequently, the IRL agent not only maintained a higher success rate but also enhanced overall task efficiency compared to the BC agent.

Figure 4 illustrates two distinct failure cases of the

IRL agent. In Figures 4(a) and 4(b), the IRL agent repeatedly reaches the minimum boom angle without progressing toward the target, leading to the termination of the episode at the maximum timestep limit. These failures occur not due to collisions but because the agent's movements become inefficient, highlighting a lack of comprehensive exteroceptive perception of the excavator's body. This inefficiency suggests that the IRL agent's decision-making process does not fully account for the excavator's physical constraints, leading to suboptimal movement strategies.

These observations underscore the critical role of parameter diversity within the IRL framework. Unlike traditional IL approaches that merely replicate observed behaviors, the proposed IRL approach models human-like decision-making strategies through modular subtask decomposition and explicit distance-based state features. This enables the agent to develop its strategic approaches to overcome harsh and complex environments and tasks. This capability to generate unique strategies, rather than simply imitating human actions, highlights the advantage of the proposed modular IRL approach in fostering more resilient and effective autonomous agents.

5.2 Summary and Implications

This study illustrates how the proposed IRL-based approach that models human-like decision-making can outperform a traditional IL approach in a demanding robotic construction task, echoing the importance of deeper reward representations and highlighting that direct behavior imitation can lead to failures when the environment is dynamic or partially observable.

In practical terms, the proposed approach could be adopted in extraterrestrial construction scenarios or similarly constrained environments where robotic autonomy must manage unpredictable conditions. By leveraging task modularization and employing explicitly modeled state features, the IRL agent demonstrated adaptability and robustness that could reduce mission risks and resource waste. In this regard, autonomous construction robots modeled by the proposed approach could perform construction tasks (i.e., excavation or debris clearance) with minimal human oversight, saving time and costs.

Several assumptions and abstractions were made to manage complexity and computational overhead. The excavation task was described with a limited number of subtasks, which may not capture all subtasks of the human decision-making process. Additional constraints, including simplified collision models and a uniform gravitational field, might not reflect every aspect of lunar physics. These simplifications allowed the method to converge more quickly and highlight possible limitations when extending to higher-fidelity simulations or more intricate tasks. Future work might consider expanding the subtask framework or incorporating fine-grained terrain and machine dynamics.

These experimental findings indicate that the proposed IRL-based approach can significantly improve construction robots' robustness. By recognizing both the strengths and constraints of the proposed approach, researchers and practitioners can refine this approach for even more robotic applications in extreme construction settings.

6 Conclusion

This study was motivated by the escalating complexity and risk inherent in construction robotics, particularly in extreme environments like the Moon. The current body of research emphasizes the need for construction robot modeling methods that can adapt to uncertain conditions with minimal human oversight. However, traditional RL approaches often replicate narrow expert behaviors and can fail when the environment diverges from the training set.

In response to these challenges, this study proposes an IRL-based approach integrated with neuroscientific theories of human decision-making. By identifying underlying reward structures rather than imitating actions outright, the IRL agent demonstrated a capacity for more adaptive and reliable decision-making, even in scenarios characterized by moving objects and variable layouts. The proposed approach’s novelty is that it leads to an agent that can synthesize multiple expert strategies and yield robust performance. The experimentations showed that the agent modeled by the proposed IRL-based approach outperformed an agent modeled by BC, a conventional IL approach regarding task success rates under a simulated lunar environment setting.

Several assumptions were made in this study. The limited number of subtasks used in this study simplifies the complexity of real-world construction tasks, which often involve numerous interdependent and dynamic processes. Moreover, the high-level representations of state features may not fully translate to highly unstructured or rapidly changing environments without further customization. Future studies should consider extending the subtask framework, refining state features, or integrating more detailed physics to bridge the gap between simulated experiments and real deployment.

This study demonstrates the effectiveness of an IRL-based approach in automating construction tasks, particularly in scenarios like extraterrestrial excavator operations, where supervision is limited. The integration of human-inspired decision-making with IRL establishes the foundation for the development of robust and adaptable robotic systems capable of thriving in extreme environments. This work contributes to the broader vision of achieving fully autonomous infrastructure

development in space by addressing the unique demands of extraterrestrial construction environments.

References

[1] NASA, The Artemis plan, NASA's lunar exploration program overview, 2020. https://go.nasa.gov/3WrOLjW (accessed January 5, 2024).

[2] C. Wong, E. Yang, X.-T. Yan, D. Gu, An overview of robotics and autonomous systems for harsh environments, in: 2017 23rd Int. Conf. Autom. Comput. ICAC, IEEE, Huddersfield, United Kingdom, 2017: pp. 1–6. https://doi.org/10.23919/IConAC.2017.8082020.

[3] D. Jud, P. Leemann, S. Kerscher, M. Hutter, Autonomous free-form trenching using a walking excavator, IEEE Robot. Autom. Lett. 4 (2019) 3208–3215. https://doi.org/10.1109/LRA.2019.2925758.

[4] J. Manuel Davila Delgado, L. Oyedele, Robotics in construction: A critical review of the reinforcement learning and imitation learning paradigms, Adv. Eng. Inform. 54 (2022) 101787. https://doi.org/10.1016/j.aei.2022.101787.

[5] R. Li, Z. Zou, Enhancing construction robot learning for collaborative and long-horizon tasks using generative adversarial imitation learning, Adv. Eng. Inform. 58 (2023) 102140. https://doi.org/10.1016/j.aei.2023.102140.

[6] H.B. Ammar, M.E. Taylor, K. Tuyls, G. Weiss, Reinforcement Learning Transfer Using a Sparse Coded Inter-task Mapping, in: M. Cossentino, M. Kaisers, K. Tuyls, G. Weiss (Eds.), Multi-Agent Syst., Springer Berlin Heidelberg, Berlin, Heidelberg, 2012: pp. 1–16. https://doi.org/10.1007/978-3-642-34799-3_1.

[7] M. Zare, P.M. Kebria, A. Khosravi, S. Nahavandi, A Survey of Imitation Learning: Algorithms, Recent Developments, and Challenges, ArXiv Prepr. (2023) 1–15. https://doi.org/10.48550/arXiv.2309.02473.

[8] K. Lee, B. Vlahov, J. Gibson, J.M. Rehg, E.A. Theodorou, Approximate Inverse Reinforcement Learning from Vision-based Imitation Learning, in: 2021 IEEE Int. Conf. Robot. Autom. ICRA, IEEE, Xi'an, China, 2021: pp. 10793–10799. https://doi.org/10.1109/ICRA48506.2021.9560916.

[9] Y. Qin, T. Menara, S. Oymak, S.N. Ching, F. Pasqualetti, Representation Learning for Context-Dependent Decision-Making, in: Proc. Am. Control Conf., American Automatic Control Council, Atlanta, Georgia, USA, 2022: pp. 2130–2135. https://doi.org/10.23919/ACC53348.2022.9867204.

[10] B. Belousov, B. Wibranek, J. Schneider, T. Schneider, G. Chalvatzaki, J. Peters, O. Tessmann, Robotic architectural assembly with tactile skills: Simulation and optimization, Autom. Constr. 133 (2022) 104006. https://doi.org/10.1016/j.autcon.2021.104006.

[11] C.-J. Liang, V.R. Kamat, C.C. Menassa, W. McGee, Trajectory-Based Skill Learning for Overhead Construction Robots Using Generalized Cylinders with Orientation, J. Comput. Civ. Eng. 36 (2022) 04021036. https://doi.org/10.1061/(ASCE)CP.1943-5487.0001004.

[12] L. Huang, Z. Zhu, Z. Zou, To imitate or not to imitate: Boosting reinforcement learning-based construction robotic control for long-horizon tasks using virtual demonstrations, Autom. Constr. 146 (2023) 104691. https://doi.org/10.1016/j.autcon.2022.104691.

[13] K. You, C. Zhou, L. Ding, Deep learning technology for construction machinery and robotics, Autom. Constr. 150 (2023) 104852. https://doi.org/10.1016/j.autcon.2023.104852.

[14] S. Russell, Learning agents for uncertain environments, in: Proc. Annu. ACM Conf. Comput. Learn. Theory, ACM, New York, NY, USA, 1998: pp. 101–103. https://doi.org/10.1145/279943.279964.

[15] S. Arora, P. Doshi, A survey of inverse reinforcement learning: Challenges, methods and progress, Artif. Intell. 297 (2021) 103500. https://doi.org/10.1016/j.artint.2021.103500.

[16] F. Torabi, G. Warnell, P. Stone, Behavioral cloning from observation, in: IJCAI Int. Jt. Conf. Artif. Intell., Stockholm, Sweden, 2018: pp. 4950–4957. https://doi.org/10.24963/ijcai.2018/687.

[17] C.A. Rothkopf, D.H. Ballard, Modular inverse reinforcement learning for visuomotor behavior, Biol. Cybern. 107 (2013) 477–490. https://doi.org/10.1007/s00422-013-0562-6.

[18] D. Ramachandran, E. Amir, Bayesian inverse reinforcement learning, in: IJCAI Int. Jt. Conf. Artif. Intell., IJCAI, Hyderabad, India, 2007: pp. 2586–2591.

[19] R. Zhang, S. Zhang, M.H. Tong, Y. Cui, C.A. Rothkopf, D.H. Ballard, M.M. Hayhoe, Modeling sensory-motor decisions in natural behavior, PLoS Comput. Biol. 14 (2018) 1–22. https://doi.org/10.1371/journal.pcbi.1006518.

[20] U.S. Department of Labor, Bureau of Labor Statistics, Census of Fatal Occupational Injuries, (2022). https://www.bls.gov/iif/fatal-injuries-

tables/fatal-occupational-injuries-table-a-9-2021.htm (accessed June 30, 2024).

[21] F. Pardo, A. Tavakoli, V. Levdik, P. Kormushev, Time Limits in Reinforcement Learning, in: Proc. Mach. Learn. Res., PMLR, Stockholm, Sweden, 2018: pp. 4045–4054. https://doi.org/10.48550/arXiv.1712.00378.

[22] M.K. Barker, E. Mazarico, G.A. Neumann, D.E. Smith, M.T. Zuber, J.W. Head, Improved LOLA elevation maps for south pole landing sites: Error estimates and their impact on illumination conditions, Planet. Space Sci. 203 (2021) 105119. https://doi.org/10.1016/j.pss.2020.105119.

[23] A. Juliani, V.-P. Berges, E. Teng, A. Cohen, J. Harper, C. Elion, C. Goy, Y. Gao, H. Henry, M. Mattar, D. Lange, Unity: A General Platform for Intelligent Agents, ArXiv Prepr. ArXiv180902627 (2018). https://doi.org/10.48550/arXiv.1809.02627.

[24] M.D. McKay, Latin hypercube sampling as a tool in uncertainty analysis of computer models, in: Proc. 24th Conf. Winter Simul. - WSC 92, ACM Press, Arlington, Virginia, United States, 1992: pp. 557–564. https://doi.org/10.1145/167293.167637.

[25] H.A. Taylor, S.J. Naylor, N.A. Chechile, Goal-specific influences on the representation of spatial perspective, Mem. Cognit. 27 (1999) 309–319. https://doi.org/10.3758/BF03211414.

[26] D. Sirkin, N. Martelaro, M. Johns, W. Ju, Toward Measurement of Situation Awareness in Autonomous Vehicles, in: Proc. 2017 CHI Conf. Hum. Factors Comput. Syst., ACM, Denver Colorado USA, 2017: pp. 405–415. https://doi.org/10.1145/3025453.3025822.

Safe Reinforcement Learning for Objects Manipulation in Safety-Critical Coordinated Tasks

Fawad Khan[1], Wei Feng[2], Zhiyong Wang[1], Tianlun Huang[1], Liu Xiao[1], Asad Ali Shahid[3], Weijun Wang*[1]

[1]Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China.
[2]Guangdong Provincial Key Lab of Construction Robotics and Intelligent Manufacturing, Shenzhen, China.
[3]Dalle Molle Institute for Artificial Intelligence (IDSIA), Lugano, Switzerland.

fawad, zy.wang, tl.huang1, xiao.liu1@siat.ac.cn, wj.wang@giat.ac.cn, wei.feng@siat.ac.cn, asadali.shahid@idsia.ch

Abstract -

This paper presents a safety-focused, closed-loop grasping approach for object manipulation in dynamic environments, leveraging simulation data to ensure adaptive and safety-aware operations. We extend the OpenAI Safety Gym library by integrating a robotic arm model and propose 'Grasp Mechanics,' a novel method for adaptive gripping. Using a constrained variant of Proximal Policy Optimization (cPPO), we emphasize secure object manipulation in human-robot interaction scenarios. Preliminary experiments show that cPPO, while requiring longer training, achieves policy performance comparable to baseline PPO and significantly outperforms it in safety compliance. Our work highlights the potential of integrating advanced RL techniques with robust safety mechanisms, advancing the capabilities of robotic systems for real-world applications. These findings lay the groundwork for future advancements in safe autonomous robotics, emphasizing the importance of integrating safety considerations from the ground up.

Keywords -

Safe reinforcement learning, Robotic grasping, Autonomous learning, safety-critical coordination.

1 Introduction

performance using task success, constraint adherence, and safety violation costs. We adopt these metrics to balance task efficiency and safety compliance.

Building on Safety Gym, we integrate the Panda robotic arm, creating a novel safety-aware environment for object manipulation. Unlike prior work relying on predefined policies [5], our approach combines obstacle avoidance, movement planning, and grasping through learning-based policies. By adapting Safety Gym's constraint mechanisms, we develop customized reward and cost functions that balance task completion with safety.

In this study, we introduce a novel approach for robotic grasping in dynamic environments. Our method employs 3D translation control with a strategically fixed orientation, optimized for grasping common household objects such as a Lipton tea box or a teacup on various surfaces. While full 6D grasping (with dynamic orientation control) presents significant challenges in motion planning and feedback integration, our simplified approach achieves remarkable effectiveness while reducing computational complexity. Unlike typical bin-picking implementations where top-down approaches suffice, our method addresses more complex scenarios through adaptive motion planning and real-time feedback. Previous work has largely focused on grasp pose generation rather than complete motion trajectories, limiting effectiveness in dynamic environments [6, 7]. In contrast, we propose a novel closed-loop grasping and motion planning policy that integrates proprioceptive and tactile feedback, enabling the robot to adapt to real-time object interactions without relying on high-dimensional real-world data (such as images). Our key innovation lies in the hierarchical RL framework [1], which decomposes complex grasping tasks into manageable sub-goals, significantly reducing learning complexity while maintaining robust performance across diverse objects and tasks.

uisite for successful robotic manipulation. This aligns with recent approaches such as the teacher-student training framework [8], where a teacher policy first trained with privileged information then guides a student policy relying on image-based inputs.

A video demonstration of the simulation experiments can be found at: `https://www.youtube.com/watch?v=9oEhdw1F3aI&t=28s`.

2 Related Works

Recognizing the challenges of ensuring safe interactions, we draw upon the consensus that safe behavior in robots equates to avoiding hazardous states, a perspective supported by [9] and further elaborated through various safety and learning strategies in the literature. Significant contributions to safety in RL include the development of constraints to prevent unsafe actions, as suggested by [10], and the incorporation of concepts like reachability and stability in algorithm design. Advanced techniques emphasize aligning robot actions with human preferences for enhanced safety [4], explored through ergodicity criteria and the mitigation of adverse "side effects" in agent behaviors. In addressing the intricacies of robotic grasping and task achievement, our study is inspired by and aims, in the future, to rely on Learning from Demonstration (LfD) and RL, with a specific focus on blending these methodologies for optimized task performance. Influential works by [11] demonstrate the potential of merging human demonstrations with RL for nuanced strategy development. This hybrid approach, along with advancements in deep learning, offers new pathways for robotic control and manipulation, particularly in the realm of object grasping where recognizing and handling objects of various geometries remains a complex challenge. Recent advancements in DRL, including PPO and its constrained variant, highlight improved efficiency and safety in robotic tasks. Leveraging these innovations, our research focuses on enabling robots to safely grasp and manipulate objects in dynamic, unstructured environments, mitigating risks in human-robot interaction and advancing autonomy in safety-critical scenarios.

2.1 Our Contributions

- **Safety Gym Extension:** Integrated a robotic arm into OpenAI's Safety Gym, creating a novel environment for safe object manipulation.
- **cPPO Advantages:** Through a rigorous comparison with standard PPO, demonstrated cPPO's capability to maintain safety constraints during robotic manipulation, while achieving comparable task performance.
- **Zero-Shot Generalization:** Achieved effective policy transfer to unseen objects after training on a simple cube.
- **Efficient Learning:** Demonstrated rapid convergence within 200,000 steps while maintaining safety constraints.

Our approach addresses a critical gap in robotic manipulation by maintaining strict collision avoidance with obstacle while preserving task performance, a balance often overlooked in existing systems that prioritize task success metrics while overlooking operational safety considerations.

2.2 Finding Technical Solutions

The availability of Python resources and simulators based on various physics engines offers significant advantages but also introduces integration challenges. Our goal was to demonstrate constrained RL algorithms using a robotic arm in human-robot interaction scenarios. However, the Safety Gym library [4] lacks robotic arm agents, posing integration challenges. Initially, we attempted to integrate a robotic arm from a Gym environment into Safety Gym, assuming compatibility due to their shared MuJoCo foundation [12]. However, this revealed dependency conflicts and required older versions of Python and TensorFlow. We also explored integration with CoppeliaSim via PyRep but faced similar compatibility issues. Ultimately, we adopted a PyBullet-based Gym environment [13], which supports robotic arms and provided comprehensive documentation and examples, simplifying the configuration process.

3 Problem Formulation

The state is an 8-dimensional vector encoding the end-effector's position, the object's position, and the gripper's status, enhanced by Grasp Mechanics for adaptive and precise manipulation.

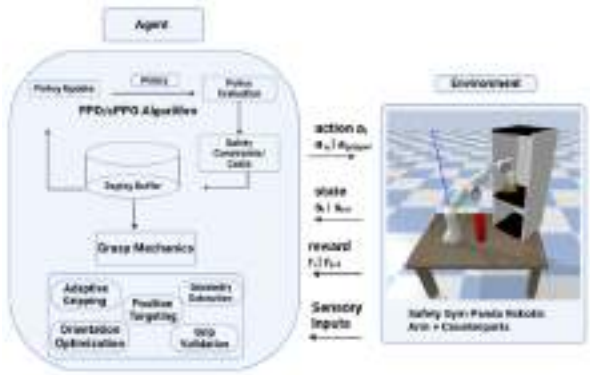


Figure 1. Agent Equipped with Grasp Mechanics

4 METHODOLOGY

4.1 Safe Reinforcement Learning Through Constraints

RL[4] is an advanced methodology focused on training RL agents to operate within predefined safety constraints, ensuring safe behavior both during training and deployment. An optimal policy in safe RL is defined as:

$$\pi^* = \arg\max_{\pi \in \Pi_C} J_r(\pi) \quad \text{where} \quad \Pi_C = \left\{\pi : J_{c_i}(\pi) \leq d_i,\ i = 1, \ldots, k\right\} \tag{1}$$

Where $J_r(\pi)$ represents an objective function driven by rewards and each J_{c_i} signifies a cost-oriented constraint function, with thresholds d_i chosen as hyper-parameters by humans. This Constraint mechanisms define a valid set of permissible policies Π_C that meet the constraints, within the framework of CMDPs. The Lagrangian technique employs a composite loss function, integrating both reward-driven and cost-driven elements. In the equation below, cost-related aspect is incorporated within the set of allowable policies Π_C. This optimization task is formulated as follows:

$$\max_{\theta} \min_{\lambda \geq 0} L(\phi, \theta) = f(\theta) - \lambda g(\theta) \tag{2}$$

Within the loss function, the two components pertain to the reward and cost, incorporating the policy model parameters θ and the Lagrangian hyper-parameter λ.

4.1.1 Pre-Learning Phase

to define requirements, establishing boundaries for task performance and safety.

Expert Knowledge: Expert input encodes operational context aligned with real-world applications, embedding task requirements via reward functions and safety constraints [16], defining action spaces leveraging existing robot skills [17], establishing state spaces with necessary perception capabilities, and incorporating demonstrations to accelerate learning [18].

We partially automate the encoding process using hierarchical RL, where sub-goals guide the agent toward specific sub-tasks. The environment uses standard robot description models (URDF), and the RL agent is trained to identify hazardous states while maintaining performance.

4.1.2 State Space

Our framework integrates simulation data with proprioceptive state information, including robot states (forces on the end-effector), fingertip states, object states, force sensor data, object geometry extraction, and distance measurements. These elements provide tactile feedback for adaptive gripping, ensuring a continuous state space essential for effective manipulation task learning. In continuous state spaces, deep neural networks (DNNs) excel at learning low-dimensional representations that map states to optimal control actions.

4.1.3 Action Space

In Human-Robot Collaboration (HRC) scenarios, the action space is defined by complex interactions in a six-dimensional space (the robot's end-effector pose). This continuous action space poses significant challenges for RL agents, often requiring millions of time steps for training, even with robust exploration and exploitation algorithms [19]. To address this, recent approaches constrain the action space by leveraging Inverse Kinematics (IK) solvers instead of directly learning joint angles [20].

This method reduces the action space dimensionality from seven (corresponding to the robot arm's seven degrees of freedom) to three, focusing only on positional changes in Cartesian space $(\Delta x, \Delta y, \Delta z)$ with fixed orientations. This reduction accelerates learning, enhances task efficiency and safety, and promotes robust output behaviors by minimizing unintended actions. Furthermore, it facilitates the transferability of learned behaviors across different robots, improving generalization and adaptability in diverse HRC applications.

4.1.4 Reward Function

60% of development effort, directly influencing learning efficiency and task performance.

4.2 Hierarchical RL Framework

In recent years, Hierarchical RL (HRL) has emerged as a powerful framework for solving complex decision-making tasks by decomposing them into a hierarchy of simpler subtasks [1]. This approach is particularly beneficial for robotic manipulation tasks, where a robot must perform a sequence of actions with long-term dependencies and complex goals. In this work, we apply HRL to a robotic arm tasked with manipulating objects in an environment. The task is divided into phases approaching, grasping, and lifting each governed by a low-level policy, while the high-level policy coordinates the transitions between these phases, with each phase having its corresponding reward function to guide learning. This hierarchical structure enhances both learning efficiency and execution performance by allowing the robot (agent) to focus on simpler sub-problems sequentially.

5 IMPLEMENTATION SETTINGS

We employ the PPO algorithm with the PPO-Clip variant [21] to enhance policy gradient stability and efficiency.

To address safety concerns within the RL framework, our work integrates the PPO algorithm with the Lagrangian method for constrained optimization. This integration allows us to define a feasible set of policies (πc) that satisfy predefined safety constraints $J_{ci}(\pi) \leq d_i$, CMDPs incorporate a set of cost functions $c_1, ...c_k$, which are distinct from the reward function, as outlined in the framework of CMDPs [4] (see subsection 4.1 in Methodology). In our scenario, if the agent makes contact with an obstacle at time step t, it incurs a cost of $c(s_t) = 1$; otherwise, the cost is 0. This is represented by $d_i = 1$ if the robotic arm violates the safety constraint, and $d_i = 0$ otherwise. This is essential to constrain hazardous behavior within the Safety Gym environment. Our integration of PPO and its Lagrangian extension advances both policy optimization and safety-aware human-robot interactions, ensuring robotic agents operate within safety margins in simulated environments, a foundation for constrained RL solutions applicable to real-world scenarios.

5.1 Multi-step Dense Reward System

This section describes a multi-step dense reward system for training agents in object manipulation tasks, providing continuous feedback to guide the arm through approaching, grasping, and lifting phases.

The core of the reward system is a distance-based reward, incentivizing the agent to minimize the Euclidean distance between the end-effector and the target object. This reward shaping [22], ensures incremental improvement in positioning. Additionally, a contact stability reward encourages stable grasping: $r_{\text{contact}} = 2$ for first contact with both fingers, 3 for maintaining stable contact, -2 for persistent single-finger slip, and -1 for losing established contact.

Once the object is lifted above a threshold, a significant positive reward reinforces the desired behavior: $r_{\text{grasp_lift}} = 5$. Safety is ensured through collision detection: $c_{\text{collision}} = \mathbb{1}(\text{contact between robot and obstacle})$, with info = $\{\text{cost} : c_{\text{collision}}\}$, aligning with CMDPs [23] to evaluate both task performance and safety.

5.2 Model Structure and Training

Our method employs two RL algorithms: PPO [21] and its Lagrangian variant. The agent explores the environment through on-policy learning, collecting data via controlled policy rollouts rather than random exploration. While we use actor-critic algorithms, our approach is generalizable to any deep Q-learning algorithm, whether online or offline.

PPO: An on-policy actor-critic algorithm that optimizes cumulative rewards using a clipped surrogate objective, known for its simplicity and stability in continuous control tasks.

PPO Lagrangian Variant: Extends PPO by incorporating a Lagrangian multiplier to enforce safety constraints during training, making it ideal for safety-critical tasks. Both algorithms utilize low-dimensional state information and tactile feedback, enhanced by Grasp Mechanics for adaptive manipulation.

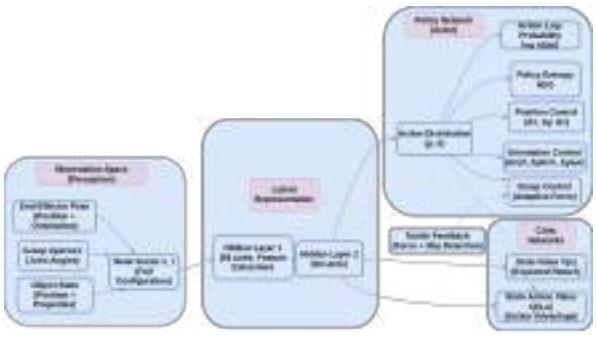


Figure 2. Actor-Critic Architecture for Robotic Manipulation with Tactile Feedback.

`cost_gamma` (0.99), `lam` (0.97), `cost_lam` (0.97), `ent_reg` (0.01), `cost_lim` (25), `penalty_init` (1.0), `target_kl` (0.01), and `vf_iters` (80). Optimization techniques include 80 gradient descent iterations per update cycle, a clipping ratio to prevent drastic policy changes, and backtracking steps for efficient optimization.

Algorithm 1: Constrained PPO

Data: Environment (env)
Result: Trained Policy
Initialization: Initialize logger, seeds, environment, buffers, parameters,and optimizers.
Policy Learning: Compute likelihood ratio, surrogate advantage, and cost.
Main Loop: Initialize environment variables.
Epoch Loop: for *each epoch* **do**
  Update penalty;
  for *each step* **do**
    Render env, get action, step in env;
    Store experience, update stats;
    if *terminal state or end of epoch* **then**
      Finish path, reset env;
    end
  end
  Save model, update policy and value functions;
  Log performance and stats;
end
Main Function: Parse args, initialize logger and agent. Execute `run_polopt_agent(parameters)`.

5.3 Grasping Policy via Grasp Mechanics

5.4 Task Success and Evaluation Metrics

Task success is evaluated using three metrics from the Safety Gym protocol [4]: (1) **Task Performance**, measured as the percentage of successfully completed tasks (approaching, grasping, and lifting) out of total attempts (Figure 3), Success Rate $= \frac{\text{Successful Tasks}}{\text{Total Tasks}}$; (2) **Constraint Satisfaction**, tracking safety violations (e.g., collisions), with fewer violations indicating better safety; and (3) **Average Regret**, reflecting cumulative safety costs during training, Average Regret $= \frac{1}{N}\sum_{i=1}^{N} C_i$, where C_i is the safety cost at step i, and N is the total number of steps. These metrics balance task efficiency and safety compliance.

5.5 Zero-shot Policy Transfer Across Objects

We trained our robotic manipulation system through 500,000 iterations using a cube as the reference object. The system's MLP actor-critic architecture processed state vectors containing object pose, gripper state, and spatial relationships. To evaluate generalization capabilities, we tested the trained policy on unseen objects: a Lipton tea box, teacup, soap bar, and remote control. The system successfully grasped all test objects, with highest success rates for the Lipton tea box 95% and soap bar 90%. While performance remained robust on objects with regular geometries, success rates decreased slightly for the teacup 87% and remote control 85% due to their complex or asymmetrical shapes challenging gripper alignment. The system demonstrated effective zero-shot transfer of manipulation policies without object-specific training.

5.5.1 Performance Metrics

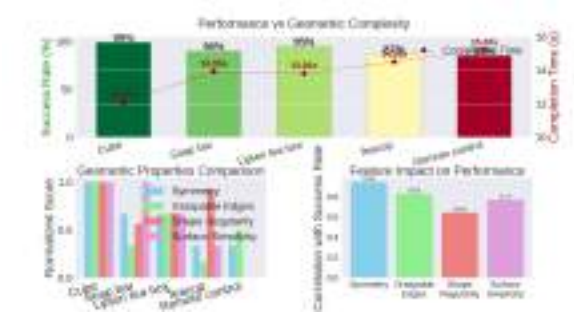


Figure 3. Performance evaluation of zero-shot policy transfer across objects with varying geometries, demonstrating the system's ability to generalize and highlighting the impact of features like symmetry, graspable edges, shape regularity, and surface simplicity on success and completion time. Note: Results averaged over 40 attempts per object.

5.5.2 Comparative Analysis

We compared our closed-loop grasping approach to object manipulation, which fundamentally differs from traditional fixed-policy methods [7]. Unlike these fixed-sequence approaches [6],[7], our system utilizes a purely learning-based framework that dynamically adapts within a sequential decision-making architecture. The key innovation lies in its ability to monitor and adjust behavior based on the current state. For example, in failure scenarios, such as when a grasp attempt fails and the object is not secured, the system reassesses and retries, rather than continuing with the predefined sequence. As shown in Figure 3 and Table 1, our approach achieves higher success rates and adapts more effectively to varying object shapes and positions.

Comparison with Existing Approaches

We empirically compared our approach with baselines across several key features as shown in Table 1.

Table 1. Comparison of Object Grasping and Manipulation Approaches

Feature	Our Approach	Baselines [7, 6]
Learning Paradigm	Goal-Directed Reinforcement Learning	Imitation Learning (predefined behavior + learning from data)
Decision Making	Sequential, state-dependent with learned policy	Fixed action sequences with predefined policy
Failure Handling	Dynamic reassessment with real-time adaptation	Continue with predefined sequence, static behavior
State Awareness	Continuous monitoring of gripper state	Limited state feedback
Task Execution	Safety-aware operation while maintaining performance	Performance optimization without safety considerations

6 EXPERIMENT DESIGN AND REALIZATION

Our analysis of robotic arm control in 3D Cartesian space reveals that increasing the state vector's dimensionality does not significantly improve learning efficiency for adaptive grasping. However, integrating 'Grasp Mechanics' as a feedback mechanism enhances both learning efficiency and overall effectiveness. A comparative study of PPO and its constrained variant, cPPO, shows that cPPO incorporating Lagrangian safety methods achieves lower average costs, indicating safer training despite slightly longer training periods. Figure 4 illustrates this trade-off, comparing reward values and costs across training cycles. The graph differentiates between average cumulative rewards and episode-wise costs for PPO and cPPO, demonstrating the impact of constrained optimization on safety and effectiveness in robotic control.

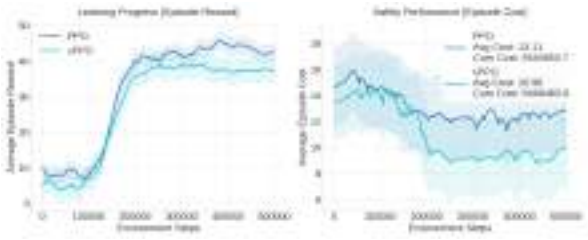


Figure 4. PPO vs. cPPO Rewards and Costs Comparison in Object Manipulation.

The visual comparison highlights cPPO's gradual learning curve and consistently lower cost metrics, reflecting its prioritization of safety. This systematic minimization of costs ensures safer operational dynamics for the robotic arm.

The Table 2. compares the average costs per run for the standard PPO algorithm and its constrained variant (cPPO). While PPO incurs an average cost of 13.11, cPPO achieves a lower cost of 10.96, reflecting its enhanced safety performance. This evaluation is based on a task involving the manipulation of various geometries and securing them at target locations under strict safety constraints, using a 3D action representation in Cartesian space.

Table 2. Average comparison costs per run for PPO and cPPO.

Algorithms	Average Cost Incurred
PPO	13.11
cPPO	10.96

6.1 Discussion and Conclusion

robots must interact with diverse objects while adhering to safety constraints. Our approach leverages cPPO to develop policies that balance task performance with safety. We introduce Grasp Mechanics, a novel closed-loop grasping policy that combines compact state information acquisition with precise positional control. By relying primarily on proprioceptive, low-dimensional kinematic data, our system achieves effective manipulation without requiring detailed object shape information, a key advantage in real-world environments where high-dimensional data, such as images, may be unavailable.

Our experimental results validate two key contributions. First, the cPPO implementation demonstrates consistent improvement in safety performance, reducing the average episode cost from approximately 14 to 10.96 while maintaining task effectiveness. This reduction in safety cost, combined with the stable learning progress shown in the reward curve, indicates that our approach successfully balances safe navigation with task completion. Second, the system demonstrates an 89% success rate in grasping previously unseen objects using a purely learning-based policy, highlighting that safe manipulation can be achieved without compromising task effectiveness. In contrast, previous work relied on predefined policies [5], which may have limited flexibility and adaptability.

The learning curves from our experiments reveal both strengths and limitations of the current approach. While cPPO demonstrates consistent convergence to safe policies within 200k training steps, we observed that the initial exploration phase requires careful tuning of safety boundaries to prevent constraint violations. This highlights a trade-off between learning efficiency and safety guarantees that warrants further investigation.

6.1.1 Limitations and Future Work

menting a teacher-student framework [8], where our safe manipulation policies serve as teachers for training vision-based students. This approach could reduce the computational cost of learning from visual inputs while maintaining safety, with significant implications for real-world applications. Our research contributes to the field of autonomous robotics by demonstrating that effective object manipulation can be achieved while strictly adhering to safety requirements. The combination of constrained RL and proprioceptive feedback provides a particularly promising foundation for future sim-to-real transfer. Unlike vision-dependent approaches that often struggle with the reality gap due to visual discrepancies, our proprioceptive-focused policy relies on physical interactions and low-dimensional state representations that theoretically transfer more reliably across domains. This design choice deliberately addresses a critical challenge in robotics: policies that perform well in simulation often degrade significantly when deployed on physical hardware. By prioritizing contact dynamics, force-based feedback, and physics-grounded policy constraints, our approach is inherently more resilient to the domain transfer challenges that typically hinder real-world deployment. Future work will focus on systematically bridging the simulation-reality gap through domain randomization and progressive transfer techniques, with particular emphasis on preserving safety guarantees while improving sample efficiency across the full spectrum of manipulation tasks from controlled laboratory environments to dynamic human-centered settings.

Acknowledgment

This work was supported in part by the National Key R&D Program of China (No. 2023YFB4705002), in part by the National Natural Science Foundation of China (U20A20283), in part by the Guangdong Provincial Key Laboratory of Construction Robotics and Intelligent Construction (2022KSYS013), in part by the CAS Science and Technology Service Network Plan (STS) - Dongguan Special Project (Grant No. 20211600200062), and in part by the Science and Technology Cooperation Project of Chinese Academy of Sciences in Hubei Province Construction 2023.

References

[1] Benjamin Beyret, Ali Shafti, and A Aldo Faisal. Dot-to-dot: Explainable hierarchical reinforcement learning for robotic manipulation. In *IROS*, pages 5014–5019. IEEE, 2019.

[2] Xue Bin Peng, Erwin Coumans, Tingnan Zhang, Tsang-Wei Lee, Jie Tan, and Sergey Levine. Learning agile robotic locomotion skills by imitating animals. *arXiv preprint arXiv:2004.00784*, 2020.

Jakub Pachocki, Arthur Petron, Matthias Plappert, Glenn Powell, Alex Ray, et al. Learning dexterous in-hand manipulation. *The International Journal of Robotics Research*, 39(1):3–20, 2020.

[4] Alex Ray, Joshua Achiam, and Dario Amodei. Benchmarking safe exploration in deep reinforcement learning. *arXiv*, 2019.

[5] Lirui Wang, Yu Xiang, Wei Yang, Arsalan Mousavian, and Dieter Fox. Goal-auxiliary actor-critic for 6d robotic grasping with point clouds. In *CoRL*, pages 70–80, 2022.

[6] Lirui Wang, Yu Xiang, and Dieter Fox. Manipulation trajectory optimization with online grasp synthesis and selection. *arXiv preprint arXiv:1911.10280*, 2019.

[7] Joao Carvalho, An T Le, Philipp Jahr, Qiao Sun, Julen Urain, Dorothea Koert, and Jan Peters. Grasp diffusion network: Learning grasp generators from partial point clouds with diffusion models in so (3) xr3. *arXiv preprint arXiv:2412.08398*, 2024.

[8] Tao Chen, Jie Xu, and Pulkit Agrawal. A system for general in-hand object re-orientation. In *CoRL, pages=297–307, year=2022, organization=PMLR*.

[9] Alexander Hans, Daniel Schneegaß, Anton Maximilian Schäfer, and Steffen Udluft. Safe exploration for reinforcement learning. In *ESANN*, pages 143–148, 2008.

[10] Moshe Haviv. On constrained markov decision processes. *Operations research letters*, 19(1):25–28, 1996.

[11] Yuke Zhu, Ziyu Wang, Josh Merel, Andrei Rusu, Tom Erez, Serkan Cabi, Saran Tunyasuvunakool, János Kramár, Raia Hadsell, Nando de Freitas, et al. Reinforcement and imitation learning for diverse visuomotor skills. *arXiv preprint arXiv:1802.09564*, 2018.

[12] Emanuel Todorov, Tom Erez, and Yuval Tassa. Mujoco: A physics engine for model-based control. In *2012 IEEE/RSJ international conference on intelligent robots and systems*, pages 5026–5033. IEEE, 2012.

[13] Erwin Coumans and Yunfei Bai. Pybullet, a python module for physics simulation for games, robotics and machine learning, 2016.

[14] Marvin Rausand. *Reliability of safety-critical systems: theory and applications*. John Wiley & Sons, 2014.

[16] Gal Dalal, Krishnamurthy Dvijotham, Matej Vecerik, Todd Hester, Cosmin Paduraru, and Yuval Tassa. Safe exploration in continuous action spaces. *arXiv preprint arXiv:1801.08757*, 2018.

[17] Marcin Andrychowicz, Filip Wolski, Alex Ray, Jonas Schneider, Rachel Fong, Peter Welinder, Bob McGrew, Josh Tobin, OpenAI Pieter Abbeel, and Wojciech Zaremba. Hindsight experience replay. *Advances in neural information processing systems*, 30, 2017.

[18] Xue Bin Peng, Pieter Abbeel, Sergey Levine, and Michiel Van de Panne. Deepmimic: Example-guided deep reinforcement learning of physics-based character skills. *ACM Transactions On Graphics (TOG)*, 37(4):1–14, 2018.

[19] Peter Henderson, Riashat Islam, Philip Bachman, Joelle Pineau, Doina Precup, and David Meger. Deep reinforcement learning that matters. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

[20] Mohamed El-Shamouty, Xinyang Wu, Shanqi Yang, Marcel Albus, and Marco F Huber. Towards safe human-robot collaboration using deep reinforcement learning. In *ICRA*, pages 4899–4905. IEEE, 2020.

[21] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. In *Proceedings of the 34th International Conference on Machine Learning*, pages 1–12. PMLR, 2017.

[22] Andrew Y Ng, Daishi Harada, and Stuart Russell. Policy invariance under reward transformations: Theory and application to reward shaping. In *Icml*, volume 99, pages 278–287, 1999.

[23] Eitan Altman. *Constrained Markov decision processes*. Routledge, 2021.

CoNav Chair: Design of a ROS-based Smart Wheelchair for Shared Control Navigation in the Built Environment

Yifan Xu[1], Qianwei Wang[2], Jordan Lillie[3], Vineet Kamat[4] and Carol Menassa[5]

[1]Department of Civil and Environmental Engineering, University of Michigan, United States of America
[2]College of Literature, Science, and the Arts, University of Michigan, United States of America
[3]Wheelchair Seating Service, University of Michigan Health, United States of America
[4]Department of Civil and Environmental Engineering, University of Michigan, United States of America
[5]Department of Civil and Environmental Engineering, University of Michigan, United States of America

yfx@umich.edu, qweiw@umich.edu, jlillie@med.umich.edu, vkamat@umich.edu, menassa@umich.edu

Abstract -

With the number of people with disabilities (PWD) increasing worldwide each year, the demand for mobility support to enable independent living and social integration is also growing. Wheelchairs commonly support the mobility of PWD in both indoor and outdoor environments. However, current powered wheelchairs (PWC) often fail to meet the needs of PWD, who may find it difficult to operate them. Moreover, existing research on robotic wheelchairs typically focuses either on full autonomy or enhanced manual control, which can lead to reduced efficiency and user trust. To address these issues, this paper proposes a Robot Operating System (ROS)-based smart wheelchair (CoNav Chair) that incorporates a shared control navigation algorithm and obstacle avoidance to support PWD while fostering efficiency and trust between the robot and the user. Our design is divided into hardware and software components. Experimental results conducted in a typical indoor social environment demonstrate the performance and effectiveness of our smart wheelchair's hardware and software design. This integrated design fosters trust and autonomy, crucial for the acceptance of assistive mobility technologies in the built environment.

Keywords -

Smart Wheelchair; Navigation; Shared Control

1 Introduction

With the number of people with disabilities (PWD) increasing annually, more and more people will have to rely on wheelchairs to facilitate their mobility to live their lives independently. In the United States, nearly 2.7 million people experience mobility challenges leading them to rely on wheelchairs in their daily lives [1]. That is expected to increase by 7% annually [2] leading to a growing demand for mobility support systems that enable independent living and social integration. Wheelchairs, especially powered wheelchairs (PWC), are the most common means of assisting PWD in both indoor and outdoor environments where these devices provide crucial mobility assistance, allowing users to navigate their surroundings more easily [3]. In addition, smart wheelchairs created to introduce autonomy in PWC operation have demonstrated significant potential in improving the quality of life for PWD [4].

Current smart wheelchairs face a notable limitation: they often focus exclusively on either full autonomy or enhanced manual control [4]. This imbalance can result in inefficiencies during navigation and a lack of trust in the technology—critical factors that influence the successful adoption of these devices [5]. Shared control enables users to partially override the autonomous navigation system, allowing them to contribute to decision-making during navigation. This approach not only enhances the mobility of PWD but also fosters user trust by offering a balanced partnership between autonomy and manual control.

Building on insights from previous smart wheelchair designs [6, 7, 8, 9, 10, 11] and shared control-based navigation structures [12], we developed both the hardware and software for such a shared control-based smart wheelchair that we call the CoNav Chair.

For hardware design, We based our hardware design on a widely used commercial PWC. The key elements include the integration of LiDAR, a camera, an inertial measurement unit (IMU), and two encoders. These sensors collectively provide comprehensive input data for the system. To accommodate all the sensors and the controller board, we designed a dedicated sensor mounting board located under the wheelchair. This sophisticated hardware integration lays the foundation for the implementation of the advanced shared control-based navigation algorithm, ultimately contributing to a more intuitive, reliable, and efficient mobility solution for people with disabilities. Additionally, we implemented a closed-loop proportional–integral–derivative (PID) [13] motor control module to manage the wheelchair's motors effectively.

The software design centers around a ROS-based Simultaneous Localization and Mapping (SLAM) and navigation framework, which facilitates navigation and interactive motor control. For the SLAM module, we employ Faster-LIO [14] to estimate the wheelchair's states and positions and RTAB-Map [15] to construct



Figure 1. Development Timeline of Smart Wheelchairs

a detailed map of the indoor environment. This 2D map is then used in our shared control-based navigation system. By integrating the user's control input signals with the autonomous navigation system, the wheelchair can navigate effectively through the mapped area.

In general, our proposed system has the following contributions:

- We developed the CoNav Chair, a shared control-based smart wheelchair platform that bridges the gap between full autonomy and manual control by leveraging our proposed shared control algorithm, setting a new benchmark for assistive mobility solutions.
- At the hardware level, we developed a ROS-based smart wheelchair hardware platform built from a widely used commercial PWC with essential control electronics and sensors to enable advanced perception, mapping, autonomous navigation, and precise motor control.
- At the software level, we proposed a novel shared control-based navigation framework for assistive mobile robots integrating a user's desire for path preference and control during navigation to improve the efficiency and human-robot trust.
- We have tested our wheelchair system in a built environment and demonstrated that our robot system outperforms other autonomous wheelchair navigation systems both in effectiveness and user comfort.

2 Related Work

The development of smart wheelchairs has been a focal point in the literature aimed at improving the mobility and independence of PWD. A lot of research on wheelchair hardware instrumentation and software integration has been conducted and implemented to address the limitations of the traditional PWC.

The research on smart wheelchairs started in the 1990s, and a lot of work related to smart wheelchairs has been developed such as VAHM [6], Vulcan1.0 [7], Wheelsley [8] and NavChair [9], Vulcan2.0 [10] and Cyberwork Wheelchair [11]. The development timeline of the smart wheelchair system is shown in Figure 1. VAHM is equipped with a contact sensor, ultrasonic and infrared sensors to detect collisions, and wall-following algorithms to allow the wheelchair to navigate by following the wall. Wheelsley allows wheelchair users to give navigation commands by using an eye-tracking system. Vulcan1.0 uses 2D LiDAR and RGB camera input to build a visual 2D map for the wheelchair user to visualize their current location and their surroundings to let them better operate the wheelchair. The NavChair is an assistive wheelchair navigation system that integrates autonomous navigation, obstacle avoidance, door passage, and wall-following which allow the wheelchair to navigate autonomously. Vulcan2.0 is the improved version of Vulcan1.0 which integrates autonomous navigation algorithms and reduces the cost by using 2D LiDAR only. The Cyberwork Wheelchair is a commercially available smart wheelchair which is equipped with multiple RGB-D cameras to get the surrounding information and use it to navigate autonomously while avoiding obstacles. The performance comparison is shown in Table 1.

While these wheelchair platforms offer significant navigational assistance to individuals who struggle with traditional powered wheelchairs, they rely heavily either on sensor data or human input, which can have limitations and sensitivity to environmental changes [16]. For user controlled wheelchair systems, such as those using

Table 1. Comparison between the previous wheelchair models and the CoNav Chair

Wheelchair Models	SLAM compatible	Autonomous navigation	Shared control by user	Sensors
Powered wheelchair	N/A	N/A	N/A	N/A
VAHM [6]	N/A	N/A	N/A	Contact, Ultrasonic sensor
Wheelsley [8]	N/A	N/A	✓	Eye tracking system
Vulcan 1.0 [7]	✓	N/A	N/A	2D LiDAR and camera
Nav-Chair [9]	N/A	✓	N/A	12 ultrasonic sensors
Vulcan 2.0 [10]	✓	✓	N/A	2D LiDAR
Cyberwork [11]	✓	✓	N/A	3 Cameras and eye
Co-Nav Chair(Ours)	✓	✓	✓	3D LiDAR and RGB-D camera



Figure 2. Hardware Design of the CoNav Chair

eye-tracking control [8], they require substantial cognitive effort or fine motor skills, which can be fatiguing or unreliable over extended periods. Fully autonomous systems, on the other hand, can fall short in dynamic and unpredictable environments due to sensor limitations or algorithmic shortcomings [17]. Our proposed system is equipped with multiple sensors and a shared control-based navigation system that allow users to partially control the wheelchair which increases the reliability of the intelligent wheelchair system, addressing the limitations of fully autonomous systems in dynamic and unpredictable environments.

3 Hardware Design

For this project, we have equipped a commercially available PWC, Quantum Q6 Edge 2.0, with control electronics, optical encoders, and multiple sensors such as VLP-16 LiDAR, ZED stereo camera, and Witmotion IMU. Although LiDAR and cameras are fundamental for these robotic functions, most commercial PWCs (e.g., Quantum Q6 Edge 2.0, Pride Mobility) are not manufactured with such sensing systems by default, and many current "smart" wheelchairs on the market remain prohibitively expensive or do not implement shared control. Inspired by the widely used Turtlebot3 platform [18], our goal is to develop a universal, research-oriented standard for shared-control smart wheelchairs. While our prototype employs a 3D LiDAR and an RGB-D camera to demonstrate robust performance, the architecture is modular—it can readily accommodate alternative sensors such as a 2D LiDAR or a single RGB-D camera. This flexible design allows researchers and manufacturers to balance cost and functionality as needed, thereby fostering accessible, shared-control mobility solutions for a broader range of users. The hardware architecture is shown in Figure 2.

As shown in Figure 3, Our hardware structure follows the Sensing - Reasoning - Acting structure. For the Sensing part, all of the sensors, such as the LiDAR and the camera will provide the robot with environmental information about the surroundings and publish the sensor data to a laptop PC for further processing. For the Reasoning module, the laptop PC is connected to the sensors and motors, which run a Linux operating system and ROS. Our shared control-based navigation program will run on this PC and send actuator control commands to motors. For the Acting module, a motor control system based on R-Net is designed, and a closed PID control system is integrated to ensure the stability of the lower-level control system.

3.1 Motion Control Module

For the motor control module, we are using the alternative control interface embedded in the wheelchair. The alternative control interface is on the enhanced display, which has a 9-way pin-out. We design our motor controller board, which can output two ways of voltage ranging from 4.8V - 7.2V to both control the forward and rotation velocity. The board is programmed to have a *rosserial* module to take command velocity signals sent from our laptop by publishing and subscribing nodes and topic protocol. Inside the board, two digital Potentiometers are used to convert the control message to the voltage signal and send it to the alternative control interface. After the alternative control interface gets the voltage output from the controller board, the interface will convert the voltage signals to CAN messages and send

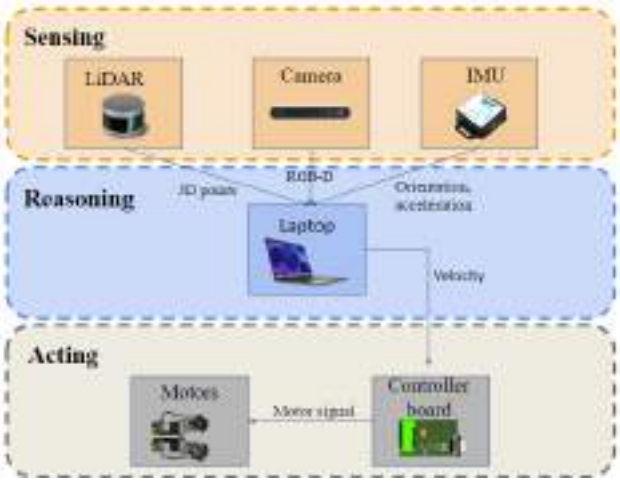


Figure 3. Hardware Dataflow of CoNav Chair



Figure 4. Motor Control Module

them to the wheelchair base. The way we connect the motor controller module is shown in Figure 4.

3.2 Sensor Module

The LiDAR and camera were mounted on the wheelchair using T-slotted aluminum extrusions to avoid obstructing leg movement. Both extrusions were attached to the wheelchair's armrest and seat rail using L brackets and square nuts, as shown in Figure 5. Cables were routed along the extrusions and secured with zip ties for a clean installation. The LiDAR and camera were mounted at the front of the wheelchair to prioritize forward-facing perception, since most powered wheelchair navigation involves moving and avoiding obstacles in the forward direction. Given that our wheelchair is a differential-drive system, it can rotate in place and thus rarely needs to travel backward; this reduces the necessity for rear coverage.

The wheelchair's motors were modified to retrofit new encoders while retaining the functionality of the integrated electromagnetic brakes. This involved preparing the motor shafts by drilling and threading them to accommodate an encoder mount. The encoders were securely installed



Figure 5. Sensor Module

using existing mounting points, following standard guidelines, to ensure precise and reliable operation.

4 Software Design

Our proposed SLAM and shared control-based navigation system of the smart wheelchair is based on ROS [19]. ROS provides a robust and flexible framework for writing robot software, allowing developers to leverage a vast ecosystem of tools, libraries, and conventions designed for robot applications. The general structure of the software arrangement is shown in Figure 6. The structure consists of three parts: a 3D mapping module, a Localization module, a Shared control-based navigation module, and a Motor control module.

4.1 Mapper and Localizer

For the mapping module, the system employs RTAB-Map [15], which is a graph-based SLAM algorithm capable of generating both 3D maps and their 2D projections of the surrounding environment. The RGB-D sensor serves as the primary input to RTAB-Map, enabling the construction of detailed 3D maps through real-time point cloud data processing. This map data is stored in a database, which is also used for managing obstacle point clouds for effective collision avoidance. The RTAB-Map framework integrates with a map server to publish the map information, accessible under the `/map` topic, which is crucial for high-level path planning and navigation tasks.

Simultaneously, the system utilizes Faster-LIO [14], a LiDAR-inertial odometry framework, for accurate localization. Faster-LIO processes data from the 3D LiDAR and IMU inputs. Initially, the raw LiDAR

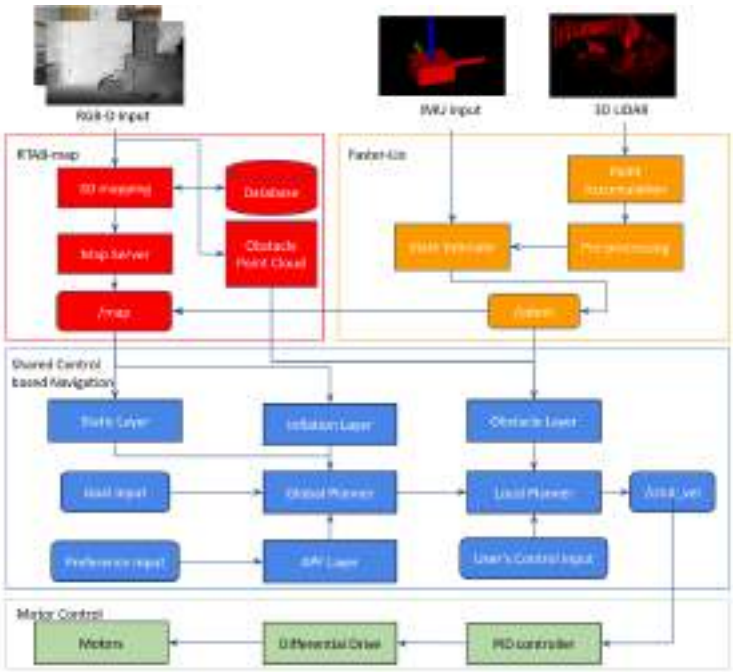


Figure 6. Software Structure of CoNav Chair

data undergoes pre-processing to filter out noise and structure the point cloud for efficient processing. These pre-processed point clouds are then accumulated to provide consistent spatial information. The state estimation module combines the IMU data with the processed LiDAR information, ensuring robust and precise localization even in dynamic environments. The resulting odometry data is published under the `/odom` topic, providing real-time position and orientation feedback to the wheelchair system.

4.2 Shared Control-based Navigation

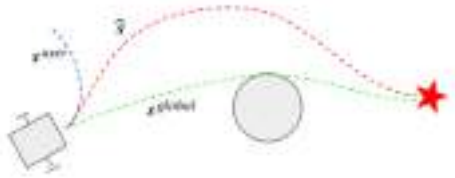


Figure 7. Shared Control-based MPC

Considering a wheelchair's role as an intimately user-integrated robotic system, it becomes essential to incorporate considerations of the user's control. Therefore, within our current framework, we have incorporated methods and concepts of shared control. In order to combine both user's control and autonomous navigation, we proposed a Shared Control-based Model Predictive Control local planner [12] to allow users to adaptively combine their control signals into the whole navigation process.

Here we utilized a non-linear model predictive control (MPC) [20] formulation for indoor environments navigation. The optimization problem is formulated below:

$$\begin{aligned} \boldsymbol{u}^* &= \arg\min_{\boldsymbol{u}\ \mathcal{U}} \mathcal{J}\ \boldsymbol{s}, \hat{\boldsymbol{s}}, \boldsymbol{u} \\ s.t. s_{t\ 1} &= f\ s_t, u_t \\ \boldsymbol{s} &\quad \mathcal{S}_{free} \\ \boldsymbol{u} &\quad \mathcal{U} \end{aligned} \tag{1}$$

where $s = s_0, ..., s_T$ is the state sequence of the robot drawn from the feasible set S_{free} by passing a control sequence $u = u_0, ..., u_{T\in 1}$ drawn from a control space U through the robot dynamic f. $\hat{s} = \hat{s}_0, ..., \hat{s}_T$ is the combined reference states of robot by considering both the user's control signal sequence and the global path plan generated by the user-preference field. $\mathcal{J}$ is the cost function that expresses the considerations of efficiency, safety, and user control.

To enable the user to take control of the navigation process, we incorporate the predicted state generated by the user's control sequence into the cost function $\mathcal{J}$, as shown below.

$$\begin{aligned} \mathcal{J}\ \boldsymbol{s},\hat{\boldsymbol{s}},\boldsymbol{u}\ &= \mathcal{J}_s\ \boldsymbol{s},\hat{\boldsymbol{s}}\quad \mathcal{J}_u\ \boldsymbol{u} \\ \mathcal{J}_s\ \boldsymbol{s},\hat{\boldsymbol{s}}\ &= \ \boldsymbol{s}\in\hat{\boldsymbol{s}}\ ^T\boldsymbol{Q}_s\ \boldsymbol{s}\in\hat{\boldsymbol{s}} \\ \hat{\boldsymbol{s}} = \eta\ k\ \boldsymbol{s}^{user}\quad & 1\in\eta\ k\ \ \boldsymbol{s}^{global} \\ \mathcal{J}^u &= \boldsymbol{u}^T\boldsymbol{Q}_u\boldsymbol{u} \end{aligned} \tag{2}$$

where $\mathcal{J}_s$ represents the state cost and $\mathcal{J}_u$ represents the input cost. The combined cost, denoted as $\hat{s}$, takes into account both the global planning state sequence $\boldsymbol{s}_{global}$ and the user's control sequence $\boldsymbol{s}_{user}$. The parameter $\theta\ k\quad 0,1$ acts as a weight for the user's control sequence, determining the degree of control the user exerts. The variable k indicates the number of control signals the user provides within a given time frame. To adjust the levels of user control in our system, we devised a weight function based on the frequency of user control, reflecting their intent to steer the wheelchair. Specifically, each press of the user control interface (e.g., joystick or button) increases the weight of user input, indicating a stronger desire to steer the wheelchair. Conversely, if pressing frequency decreases, the system gradually shifts to rely more on autonomy. We capture this behavior with an exponential weighting function, as shown below, ensuring that urgent user interventions result in greater manual influence while non-urgent scenarios default to a higher degree of autonomous control. An exponential function is used to model the user's intent, as shown below:

$$\theta\ k\ = 1\in e^k \tag{3}$$

As shown in Figure 7, the reference path combined the user's control signal and the global path plan. The MPC local planner will then allow the robot to follow the combined path while avoiding obstacles.

5 Experiment Setup and Result

In this part, we set up a real-world experiment and tested our wheelchair within an indoor-built environment. The experiment setup and the result are shown below.

5.1 Experiment Setup

The CoNav Chair runs on Ubuntu 20.04 with ROS Noetic on board. The navigation task is set up in the corridor on the first floor of an academic building on our campus, shown in Figure 8. We created four corridor scenes, each within a 1.2m-wide hallway, where obstacles were distributed to test maneuverability. The CoNav Chair measures approximately 0.8m in width, leaving only 0.4 m of total clearance. The four scenes—Scene 1, Scene 2, Scene 3, and Scene 4—differ by the number of obstacles (3, 4, 5, and 6, respectively), introducing progressively tighter navigation constraints. One of these scenes is shown in Figure 8. This setup was used to evaluate the performance of our shared-control algorithm against both fully autonomous and fully manual control methods. We assessed the planner based on time, success rate, collision rate, user satisfaction, trajectory length, and angle difference summary representing smoothness [21] against fully manual and autonomous controls.



Figure 8. Experiment Environment and Navigation Route

5.2 Experiment Result

The quantitative results in Figure 9 show that the proposed shared control-based framework significantly outperforms fully manual and autonomous navigation [22] modes in both efficiency and reliability. The shared control planner achieves shorter completion times and trajectory lengths, reflecting faster and more direct navigation. Its smoother trajectories, indicated by lower cumulative absolute angle differences, ensure fewer abrupt turns and more natural directional changes—critical for applications like assistive robots where comfort and predictability are key. Additionally, the shared control planner demonstrates superior reliability, with higher success rates and fewer collisions across all scenarios.

From the qualitative results in Figure 10, we see that shared control yields fewer collisions compared to either autonomous or manual control. Under manual control, the trajectory is more curved, and users can lose control due to unfamiliarity with the system. In contrast, autonomous control tends to choose the most efficient path, sometimes compromising safety and resulting in collisions. With shared control, users can assume partial control as needed, allowing the system to safely avoid obstacles without risking a complete loss of control.

Manual control, while allowing users to choose direct routes, often suffers from increased collisions and lower success rates due to users' unfamiliarity with the robot's dynamics and excessive adjustments near obstacles. Autonomous navigation, on the other hand, excels at obstacle avoidance and path optimization but struggles in complex scenarios like sharp turns or centrally positioned obstacles, often leading to inefficiencies or 'freezing

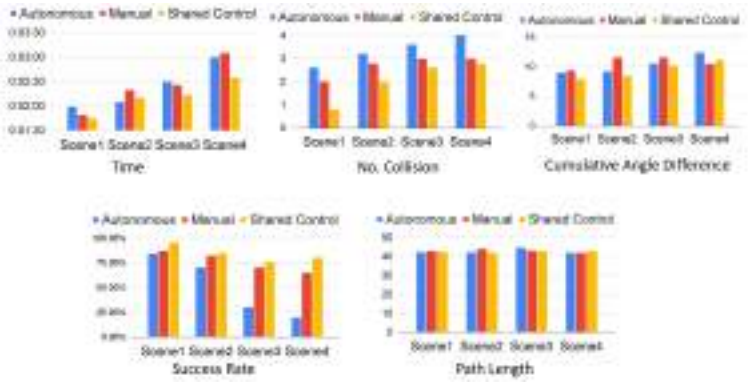


Figure 9. Quantitative Results of the CoNav Chair Evaluation Experiments

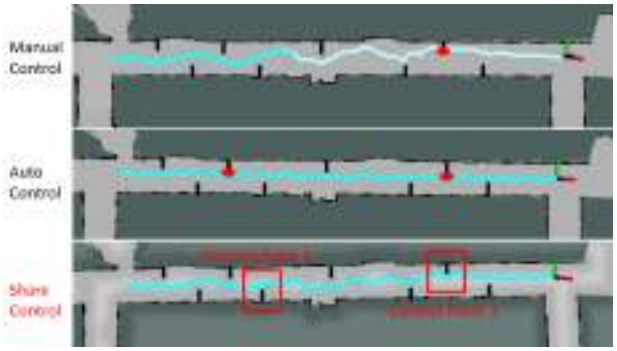


Figure 10. Qualitative Results of the CoNav Chair Evaluation Experiments. (Red explosion shape represents collision)

robot' issues [23]. The shared control framework combines the strengths of both approaches. It leverages autonomous navigation in simpler environments while enabling user intervention in complex scenarios, allowing for effective handling of challenging conditions. This hybrid strategy provides a balanced solution, enhancing efficiency, smoothness, and overall reliability.

6 Conclusion

In this work, we propose the CoNav Chair, a ROS-based smart wheelchair with shared control-based navigation and obstacle avoidance that addresses critical gaps in current PWC designs, by overcoming the limitations of fully autonomous or manual systems. Combining a multi-sensor hardware platform with LiDAR, a camera, an IMU, and encoders for real-time situational awareness and a ROS-based shared control software framework, our system integrates user inputs with autonomous navigation for smoother and more efficient mobility. Tested in real-world built environments, the CoNav Chair demonstrates superior performance in navigation effectiveness and user comfort, offering a reliable and intuitive solution to enhance independence for PWD.

7 Acknowledgement

The work presented in this paper was supported financially by the United States National Science Foundation (NSF) SCC-IRG 2124857. The support of the NSF is gratefully acknowledged.

References

[1] A. M. Koontz, D. Ding, Y. K. Jan, S. de Groot, and A. Hansen. Wheeled mobility. *BioMed Research International*, 2015:Article ID 138176, 2015. doi:10.1155/2015/138176.

[2] J. Flagg. Chapter 1: Wheeled mobility demographics. Technical report, Rehabilitation Engineering Research Center on Technology Transfer, University of New York at Buffalo, Buffalo, NY, 2009.

[3] R. A. Cooper. *Wheelchair Selection and Configuration*. Springer Publishing Company, 1998.

[4] L. Fehr, M. A. Langbein, and S. B. Skaar. Adequacy of power wheelchair control interfaces for persons with severe disabilities: A clinical survey. *Journal of Rehabilitation Research and Development*, 37(3): 353–360, 2000.

[5] Peter E McKenna, Marta Romeo, Jhielson Pimentel, Mohammed Diab, Meriam Moujahid, Helen Hastie, and Yiannis Demiris. Theory of mind and trust in human-robot navigation. In *Proceedings of the First International Symposium on Trustworthy Autonomous Systems*, TAS '23, page 1–5. ACM, July 2023. doi:10.1145/3597512.3597514. URL `http://dx.doi.org/10.1145/3597512.3597514`.

[6] G. Bourhis and P. Pino. Mobile robotics and mobility assistance for people with motor impairments: Rational justification for the vahm project. *IEEE Transactions on Rehabilitation Engineering*, 4(1): 7–12, 1996.

[7] W. S. Gribble, R. L. Browning, M. Hewett, E. Remolina, and B. J. Kuipers. Integrating vision and spatial reasoning for assistive navigation. In *Assistive Technology and Artificial Intelligence*, volume 1458 of *Lecture Notes in Computer Science*, pages 179–193. Springer, 1998. doi:10.1007/BFb0055978.

[8] H. A. Yanco. Wheelesley: A robotic wheelchair system: Indoor navigation and user interface. In *Assistive Technology and Artificial Intelligence*, volume 1458 of *Lecture Notes in Computer Science*, pages 256–268. Springer, 1998. doi:10.1007/BFb0055992.

[9] S. Levine, D. Bell, L. Jaros, R. Simpson, Y. Koren, and J. Borenstein. The NavChair assistive wheelchair navigation system. *IEEE Transactions on Rehabilitation Engineering*, 7(4):443–451, 1999.

[10] J. J. Park, S. Lee, and B. Kuipers. Discrete-time dynamic modeling and calibration of differential-drive mobile robots with friction. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6510–6517, 2017.

[11] F. Kopun. Toronto researchers develop technology for self-driving wheelchairs. *The Toronto Star*, 2023.

[12] Yifan Xu, Qianwei Wang, Vineet Kamat, and Carol Menassa. Socially-aware shared control navigation for assistive mobile robots in the built environment, 2024. URL `https://arxiv.org/abs/2405.17279`.

[13] Kiam Heong Ang, G. Chong, and Yun Li. Pid control system analysis, design, and technology. *IEEE Transactions on Control Systems Technology*, 13(4):559–576, July 2005. ISSN 1063-6536. doi:10.1109/tcst.2005.847331. URL `http://dx.doi.org/10.1109/TCST.2005.847331`.

[14] C. Bai, T. Xiao, Y. Chen, H. Wang, F. Zhang, and X. Gao. Faster-lio: Lightweight tightly coupled lidar-inertial odometry using parallel sparse incremental voxels. *IEEE Robotics and Automation Letters*, 7(2):4861–4868, 2022. doi:10.1109/LRA.2022.3152830.

[15] M. Labbé and F. Michaud. Appearance-based loop closure detection for online large-scale and long-term operation. *IEEE Transactions on Robotics*, 29(3): 734–745, 2013. doi:10.1109/TRO.2013.2242375.

[16] Jesse F. Leaman, Hung Manh La, and Luan Van Nguyen. Development of a smart wheelchair for people with disabilities. *2016 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, pages 279–284, 2016. URL `https://api.semanticscholar.org/CorpusID:5651927`.

[17] Sushil Kumar Sahoo and Bibhuti Bhusan Choudhury. Design and implementation of advanced sensor systems for smart robotic wheelchairs: A review. *Management Science Letters*, 14(3):151–168, 2024. ISSN 1923-9343. doi:10.5267/j.msl.2024.1.002. URL `http://dx.doi.org/10.5267/j.msl.2024.1.002`.

[18] Robin Amsters and Peter Slaets. Turtlebot 3 as a robotics education platform. In *Intelligent Robotics and Applications*, volume 11741 of *Lecture Notes in Computer Science*, pages 171–179. Springer, 2020. doi:10.1007/978-3-030-26945-6_16.

[19] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Ng. ROS: An open-source robot operating system. In *Proceedings of the ICRA Workshop on Open Source Software*, volume 3, 2009.

[20] A. Wang, C. I. Mavrogiannis, and A. Steinfeld. Group-based motion prediction for navigation in crowded environments. *arXiv preprint arXiv:2107.11637*, 2021. URL `https://arxiv.org/abs/2107.11637`.

[21] Silvia Guillén Ruiz, Luis V. Calderita, Alejandro Hidalgo-Paniagua, and Juan P. Bandera Rubio. Measuring smoothness as a factor for efficient and socially accepted robot motion. *Sensors*, 20(23), 2020. ISSN 1424-8220. doi:10.3390/s20236822.

[22] Amir Salimi Lafmejani and Spring Berman. Nonlinear mpc for collision-free and deadlock-free navigation of multiple nonholonomic mobile robots. *Robotics and Autonomous Systems*, 141:103774, July 2021. ISSN 0921-8890. doi:10.1016/j.robot.2021.103774. URL `http://dx.doi.org/10.1016/j.robot.2021.103774`.

[23] Peter Trautman and Andreas Krause. Unfreezing the robot: Navigation in dense, interacting crowds. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 797–803, 2010. doi:10.1109/IROS.2010.5654369.

Advancing Automated Building Inspection: A BIM and GSD Integrated Framework for UAV Applications

Emadaldeen Benshaaban[1], Po-Han Chen[2], Shahinuzzaman[3]

Gina Cody School of Engineering and Computer Science, Concordia University, Canada
Emadaldeen.Benshaaban@Concordia.ca, pohan.chen@concordia.ca, s_sujon@live.concordia.ca

Abstract –

Unmanned Aerial Vehicles (UAVs) have become indispensable for building inspections, offering safer, more efficient, and automated alternatives to traditional manual methods. However, current approaches often rely on predetermined inspection parameters, such as fixed distances, without adequately aligning with specific inspection goals or ensuring consistent data quality. This paper presents a novel methodology that addresses these challenges by integrating Building Information Modeling (BIM) and Ground Sampling Distance (GSD) into the inspection planning process. BIM is leveraged to automate the generation of Field of Views (FoVs) and viewpoints, significantly reducing human intervention and ensuring precise coverage of inspection areas. GSD is utilized to determine the optimal distance between the UAV and the surface, guaranteeing that collected data meets predefined objectives, such as detecting fine defects. To further enhance efficiency, a Genetic Algorithm (GA) is employed to optimize the flight path, minimizing travel distance while maintaining comprehensive coverage of all required areas. While the methodology is developed primarily for building inspections, it is scalable and adaptable to other infrastructure types, demonstrating its versatility in diverse scenarios. By combining BIM-based automation, GSD-driven planning, and advanced optimization techniques, this work provides a robust solution for automated, efficient, and high-quality UAV-enabled inspections for future innovations in automated inspection systems.

Keywords –

Building Inspection; Unmanned Aerial Vehicles (UAVs); Ground Sampling Distance (GSD); Field of View (FoV)

1 Introduction

Infrastructure inspections, particularly for high-rise buildings, are essential for ensuring public safety and maintaining structural integrity. Traditional visual inspection methods have been the primary approach for identifying surface defects such as cracks, spalling, and corrosion due to their simplicity and accessibility [1]. However, these methods are often subjective, time-intensive, and risky for inspectors, especially when accessing high-rise structures or geometrically complex components [2]. Moreover, manual inspections frequently rely on an inspector's expertise and judgment, leading to inconsistencies and potential inaccuracies [3]. These challenges highlight the urgent need for more reliable and efficient solutions. Unmanned Aerial Vehicles (UAVs) have emerged as a transformative tool in building inspections, offering advantages such as enhanced safety, efficiency, and accessibility. By utilizing UAVs, inspectors can reduce physical risks, streamline data collection, and minimize disruptions to building operations. Despite their potential, current UAV-enabled inspection methodologies face notable challenges, including reliance on fixed inspection parameters and limited integration of specific data quality goals [4]. Most existing approaches employ arbitrary distances for data collection without scientific justification, which can compromise the accuracy and resolution of inspection data. Furthermore, these approaches often lack automation in critical stages, such as generating inspection viewpoints and planning efficient flight paths [5]. To address these challenges, this study proposes an autonomous UAV-based framework specifically designed for façade defect inspection, focusing on the detection of surface cracks and

corrosion. Unlike general UAV-based reality capture for progress monitoring or 3D reconstruction, our approach ensures precise defect visibility by optimizing data collection based on Ground Sampling Distance (GSD). The framework integrates GSD calculations with UAV flight planning to guarantee that surface details are captured at the required resolution for defect identification. Automation and optimization techniques, including advanced algorithms for flight path planning, are integrated to streamline the process and reduce human intervention.

The remainder of this paper is organized as follows: Section 2 provides a comprehensive review of the existing literature on UAV-based inspections, emphasizing current advancements and challenges. Section 3 details the proposed methodology, outlining its key components and implementation. Lastly, the conclusion discusses the findings of this study and highlights directions for future work.

2 Literature Review

2.1 UAV-Based Building Inspections

The advent of UAVs has revolutionized the field of building inspections by offering a safer, faster, and more cost-effective alternative to traditional methods. UAVs are particularly advantageous in inspecting high-rise structures, complex geometries, and hazardous environments, where manual inspections pose significant risks. Equipped with advanced sensors and cameras, UAVs can capture high-resolution data for defect detection and structural analysis. Studies such as Rakha and Gorodetsky [6] emphasize the potential of UAVs in automating building inspections, particularly in capturing data for damage detection. They outline the benefits of UAVs in reducing human error and operational costs while enhancing the accuracy of data collection processes. Kerle et al. [7] highlight the integration of AI and machine learning algorithms with UAVs to process large datasets and automate the identification of structural damage. Their work underscores the transformative potential of UAVs when combined with intelligent systems to achieve fully autonomous inspection workflows. Recent studies have introduced optimization techniques to improve UAV-based reality capture plans. Ibrahim et al. [8] proposed a set of metrics to evaluate the effectiveness of UAV-based image capture for construction monitoring, focusing on visual coverage, resolution, and viewpoint orientation. Their later work [9] presented a multi-objective optimization model for improving reality capture plans, integrating flight path adjustments based on redundant observations and 3D reconstruction stability. While these studies introduced systematic UAV capture planning methods, they primarily optimize for construction progress monitoring, focusing on improving overall visual data collection efficiency.

2.2 Challenges in Current UAV-Based Inspection Methods

Despite the numerous advantages, UAV-based inspections face several challenges that limit their widespread adoption. A critical issue is the dependence on operator skills for flight path planning and data collection, which often leads to inconsistent coverage. Additionally, environmental factors such as wind, lighting, and weather conditions can significantly affect data quality. Feroz and Abu Dabous[10] identify key limitations in UAV-based inspections, such as the inefficiency of flight paths and the need for manual intervention in data collection. They argue that the lack of standardized protocols for UAV inspections contributes to inconsistencies across projects. Similarly, Mader and Westfeld [11] discuss the challenges of using UAVs in varying lighting and weather conditions, emphasizing the importance of robust data processing algorithms to mitigate these effects. Another significant challenge is the integration of UAV data into existing inspection workflows. The large volume of data collected by UAVs often requires extensive post-processing, which can delay decision-making. Addressing these challenges necessitates advancements in automation, optimization, and data integration.

2.3 Role of BIM in UAV-Driven Inspections

Building Information Modeling (BIM) has emerged as a powerful tool to enhance UAV inspections by providing a digital representation of structures. BIM facilitates automated planning of inspection paths and the overlay of inspection data onto structural models, enabling more efficient defect identification. Tan and Zhou [12]

demonstrate how BIM can optimize UAV flight paths to ensure complete coverage of building surfaces. By integrating UAV data into BIM, they propose a system that reduces human intervention and improves data accuracy. This integration is particularly valuable in large-scale projects, where manual inspection planning is impractical. Feroz and Abu Dabous [13] highlight the potential of BIM to streamline data analysis by integrating inspection results with structural models. They argue that BIM not only improves data visualization but also aids in predictive maintenance by identifying areas prone to future damage. Despite these advantages, the seamless integration of UAVs and BIM remains challenging, requiring advanced data fusion techniques.

2.4 Ground Sampling Distance (GSD) in Data Collection

Ground Sampling Distance (GSD) can play a critical role in UAV-based inspections by determining the spatial resolution of captured images. High-resolution data is essential for accurate defect detection, but achieving optimal GSD often requires balancing flight altitude, speed, and camera settings. Mader and Westfeld [14] explore the impact of GSD on the accuracy of structural analyses. Their study shows that lower GSD values result in higher-resolution images, which are critical for identifying small defects. However, achieving low GSD often necessitates slower flight speeds and lower altitudes, which can increase inspection time and costs. The study by Kerle et al [15] highlights the need for adaptive GSD settings to optimize data collection across varying structural features. They propose dynamic adjustments to UAV flight parameters based on the complexity of the structure being inspected, ensuring both efficiency and data quality.

2.5 Optimization Algorithms in UAV Inspections

Optimization algorithms are critical for enhancing the efficiency and effectiveness of UAV inspections. Techniques such as Genetic Algorithms (GAs), Particle Swarm Optimization (PSO), and neural networks have been widely applied to optimize flight paths and reduce redundant data collection. Liu et al [16] present a framework for automated UAV-based inspections using optimization algorithms. Their work demonstrates how GAs can minimize flight time while ensuring comprehensive coverage of building façades. The framework also incorporates collision avoidance mechanisms, enhancing operational safety. Tan and Zou [17] explore the integration of edge computing with optimization techniques to enhance data processing efficiency. By processing data in real-time during inspections, their approach reduces the need for extensive post-processing and accelerates decision-making. The reviewed literature underscores the growing reliance on UAVs for building inspections, driven by their ability to enhance safety, efficiency, and precision. Despite notable advancements in integrating UAVs with BIM and optimization techniques, challenges such as inconsistent data quality, limited automation, and better adaptability to structural complexities remain. Research into optimizing Ground Sampling Distance (GSD) and leveraging advanced algorithms, such as Genetic Algorithms and real-time edge computing, provides promising pathways to overcome these issues. This study seeks to address these gaps by developing a robust, automated framework that integrates UAVs, BIM, GSD, and optimization strategies, thereby advancing the state of the art in building inspection methodologies.

3 Methodology

This study proposes a framework for UAV-based building inspections. It is designed to automate and optimize the inspection process while considering external environmental constraints that may impact UAV operations. The framework integrates Building Information Modeling (BIM), Ground Sampling Distance (GSD) calculations, environmental condition assessment, and optimization techniques to ensure efficient, adaptable, and high-quality inspection coverage.

3.1 Framework for Automated UAV-Based Building Inspection

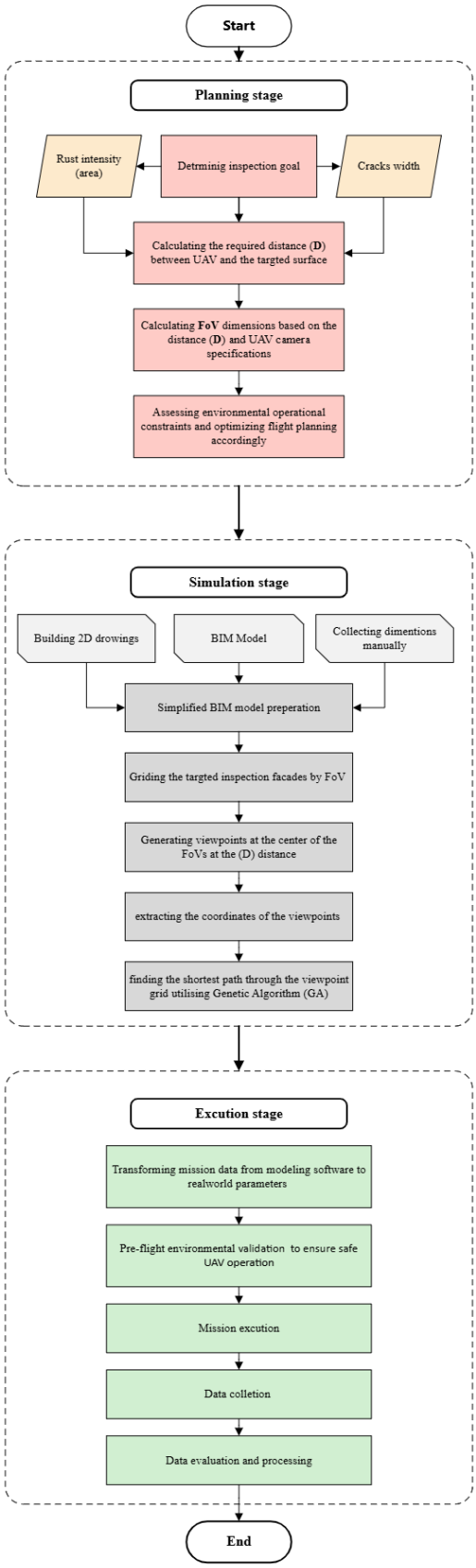
The framework is structured into three key stages: Planning, Simulation, and Execution, as shown in Figure 1. Each stage incorporates systematic processes to enhance accuracy and efficiency while adapting to real-world environmental conditions.

3.1.1 Planning Stage: The planning stage is crucial for defining inspection objectives and determining UAV flight parameters. The key components include:

a) *Defining Inspection Goals*: The UAV inspection is designed to detect surface defects such as cracks and corrosion while ensuring optimal data quality. *b) Calculating Required Distance (D):* to derive a generic equation for calculating the required distance between a UAV and a targeted surface based on inspection goals, we need to know all the parameters involved, understand how they affect the distance (D), and learn how to obtain them. The parameters influencing UAV distance can be classified into three categories, as detailed in Table 1. Deriving the final equation for D can be summarised in two simple steps. Step 1: *Relationship between GSD and sensor parameters:* Ground Sampling Distance (GSD) can be defined as the real-world size of one pixel in an image captured by a camera. It represents the spatial resolution of the image and determines the level of detail that can be captured from the ground [18].

Table 1. Parameters influence the distance between a UAV and the targeted surface

1- Inspection Requirements
GSD_r: Desired Ground Sampling Distance (mm/pixel). Figure 2
R_c: Crack width resolution requirement (mm).
FoV_w: Field of view width (mm). Figure 2
FoV_h: Field of view height (mm). Figure 2
2- Lens Parameters
f: Focal length of the camera (mm).
3- Sensor Parameters
S_w: Sensor width (mm).
S_h: Sensor height (mm).
P_x: Sensor resolution width (pixels).
P_y: Sensor resolution height (pixels).



As (GSD) is a function of the sensor size, resolution, and flight distance, it can be expressed mathematically:

$$\text{GSD} = \frac{Sw * D}{Px * f} \quad (1)$$

rearranging the equation (1) for D

$$\text{D} = \frac{GSD * Px * f}{Sw} \quad (2)$$

Step 2: *Incorporating Desired GSD and Crack Resolution:*

The desired GSD (GSD_r) is defined based on the inspection requirement for crack detection (R_c). To detect cracks of width (R_c), the resolution per pixel must satisfy:

$$GSD_r = R_c$$

Substituting GSD_r into the equation (2):

$$\boxed{\text{D} = \frac{Rc * Px * f}{Sw}} \quad (3)$$

c) Calculating FoV Dimensions: The Field of View (FoV) is calculated based on the camera's specifications and the desired distance (D). This ensures that the UAV captures an adequate area of the surface in each frame, balancing coverage with image resolution.

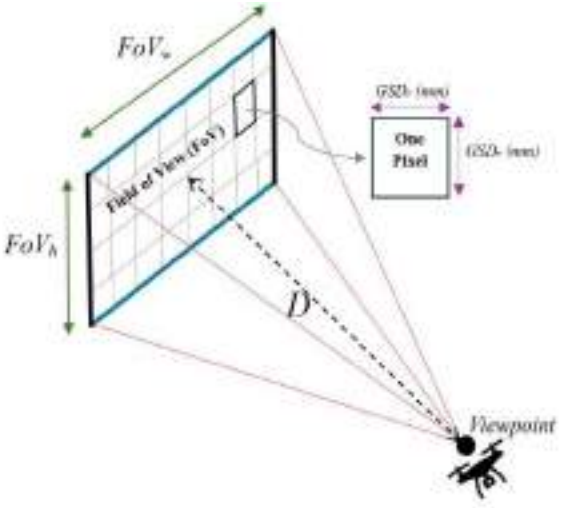


Figure 2. Field of view details, and GSD_r

d) Assessing Environmental Constraints: Environmental factors such as wind speed, temperature, lighting conditions, and GPS accuracy can significantly impact UAV stability and data quality; therefore, these variables are carefully considered during the planning phase. Wind speed constraints are analyzed to ensure stable flight, while lighting conditions are assessed to optimize visibility and defect detection. Temperature variations are considered, as extreme conditions can affect battery performance. Additionally, GPS accuracy is evaluated to minimize navigational errors.

*3.1.2 Simulation Stage***:** This stage involves virtually preparing the inspection environment and optimizing the UAV's flight path. It begins with creating a simplified BIM model of the targeted building by removing unnecessary architectural details from the facades included in the original BIM. In cases where a BIM model is unavailable for the building, the simplified model can be constructed using 2D building drawings or by manually collecting the required measurements. *Next Step:* Gridding the Inspection Facades: The targeted facades for inspection are segmented into a grid based on the calculated Field of View (FoV) dimensions. This segmentation ensures systematic coverage of the entire surface area, as illustrated in Figure 3. The FoVs can be distributed on the facades with or without overlapping, depending on dimensions and inspection requirements. Once the grid is established, viewpoints are generated at the center of each FoV, positioned at the calculated distance (D). These viewpoints act as reference points for precise UAV positioning during the inspection. It is important to note that both the grid creation for FoVs and the generation of viewpoints have been automated using Autodesk Dynamo software. The last step in the simulation stage is the flight path optimization. There are several methods and techniques for path planning; however, in this study, a Genetic Algorithm (GA) is employed to determine the shortest and most energy-efficient path through the grid of viewpoints. This optimization minimizes UAV flight time and energy consumption while ensuring full coverage of the

building façades. Figure 4 illustrates the optimal path obtained using GA.

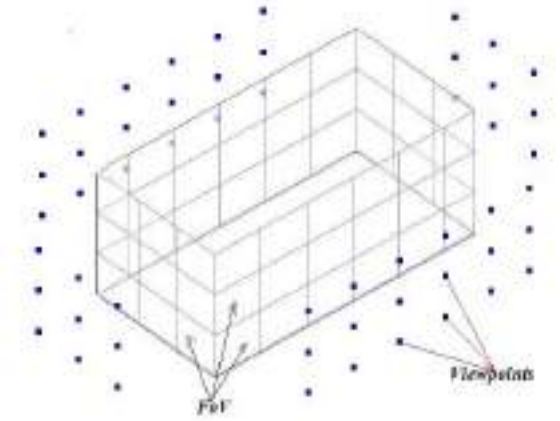


Figure 3. Griding the targeted facades and generating viewpoints at the desired distance

While the proposed framework ensures complete coverage in simulation, real-world conditions introduce uncontrollable variables such as UAV stability, environmental disturbances, and GPS accuracy, which may impact actual coverage. Therefore, a field experiment is planned to validate the methodology in real-world conditions and assess its effectiveness in achieving comprehensive façade inspection coverage.

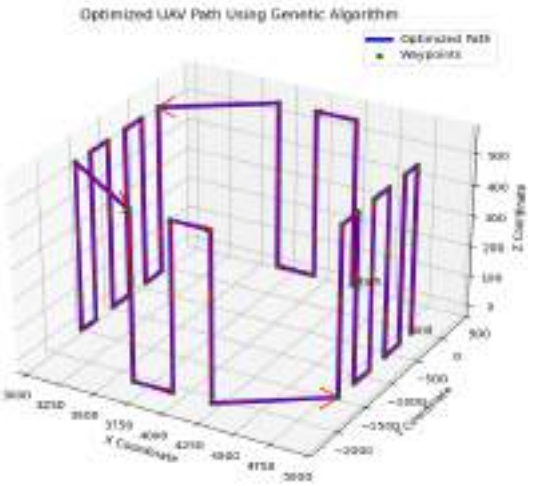


Figure 4. Optimized path planning using GA

3.1.3 Execution Stage: The execution stage translates the simulated flight plan into real-world UAV operations, ensuring reliable and environmentally adaptive data collection. This phase incorporates real-time validation and adjustments to maintain data accuracy and operational efficiency. The main steps are: (a) Mission Data Transformation: Data from the simulation stage, including waypoint (as same as viewpoints) coordinates and optimized flight paths, is transformed into real-world mission parameters compatible with UAV navigation systems. (b) Pre-Flight Environmental Validation and Adaptation: Before takeoff, real-time environmental conditions such as wind speed, lighting, and GPS stability are assessed. If any parameter exceeds operational thresholds, indicating a potential disruption, the operation is cancelled to ensure safety and data integrity. While such cancellations are rare, they are necessary to prevent compromised results and ensure mission reliability. (c) Mission Execution and Data Collection: The UAV executes the inspection mission based on the predefined flight path, capturing images and other required data for each FoV. (d) Data Evaluation and Processing: The collected data should be processed to extract actionable insights. This includes assessing rust intensity, crack width, and other structural defects.

4. Conclusion and Future Work

This study presents an automated framework for UAV-based building inspections, integrating Building Information Modeling (BIM), Ground Sampling Distance (GSD)-driven planning, and optimization algorithms. While the framework has yet to be experimentally validated, its design demonstrates significant potential to enhance inspection accuracy, operational efficiency, and scalability. The key contribution of this work is the development of a GSD-based equation to calculate the required distance (D) between the UAV and the targeted surfaces. This ensures high-resolution data collection tailored to specific defect detection needs, such as crack width and rust intensity. Additionally, the framework incorporates: (1) Automation of inspection parameters, including

Fields of View (FoVs) and viewpoints, to reduce human intervention and enable systematic and precise surface coverage. (2) The use of Genetic Algorithms (GA) to optimize UAV operations by minimizing travel time and energy consumption, enhancing inspection efficiency as part of the overall framework design. By combining these elements, the proposed framework provides a robust and adaptable solution for automated UAV-enabled building inspections, laying the groundwork for future innovations in infrastructure monitoring. While simulation results demonstrate full coverage of the inspection area, a future case study will be conducted to validate these findings in real-world conditions, accounting for potential operational constraints.

Future experiments will focus on validating these capabilities in real-world settings. Additionally, further research will explore the integration of real-time adaptive algorithms and AI-driven defect detection systems to enhance the framework's robustness and usability under varying environmental conditions.

5. References:

[1] H. Liang, S. C. Lee, W. Bae, J. Kim, and S. Seo, "Towards UAVs in Construction: Advancements, Challenges, and Future Directions for Monitoring and Inspection," Mar. 01, 2023, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/drones7030202.

[2] I. Jeelani and M. Gheisari, "Safety Challenges of UAV Integration in the Construction Industry: Focusing on Workers at Height," 2022. Accessed: Jan. 05, 2025. [Online]. Available: https://www.cpwr.com/wp-content/uploads/SS2022-UAV-safety-integration.pdf?utm_source=chatgpt.com

[3] H. Shin, J. Kim, K. Kim, and S. Lee, "Empirical Case Study on Applying Artificial Intelligence and Unmanned Aerial Vehicles for the Efficient Visual Inspection of Residential Buildings," *Buildings*, vol. 13, no. 11, Nov. 2023, doi: 10.3390/buildings13112754.

[4] V. Margolin, "Drone Inspections: A Comprehensive Guide," https://www.energy-robotics.com/post/drone-inspections-a-comprehensive-guide?utm_source=chatgpt.com.

[5] H. Hamledari, S. Davari, S. O. Sajedi, P. Zangeneh, B. Mccabe, and M. Fischer, "UAV-enabled Site-to-BIM Automation: Aerial Robotic-and Computer Vision-based Development of As-Built / As-is BIMs and Quality Control," 2018.

[6] T. Rakha and A. Gorodetsky, "Review of Unmanned Aerial System (UAS) applications in the built environment: Towards automated building inspection procedures using drones," Sep. 01, 2018, *Elsevier B.V.* doi: 10.1016/j.autcon.2018.05.002.

[7] N. Kerle, F. Nex, M. Gerke, D. Duarte, and A. Vetrivel, "UAV-based structural damage mapping: A review," Dec. 26, 2019, *MDPI AG*. doi: 10.3390/ijgi9010014.

[8] A. Ibrahim, M. Golparvar-Fard, and K. El-Rayes, "Metrics and methods for evaluating model-driven reality capture plans," *Computer-Aided Civil and Infrastructure Engineering*, vol. 37, no. 1, pp. 55–72, Jan. 2022, doi: 10.1111/mice.12693.

[9] A. Ibrahim, M. Golparvar-Fard, and K. El-Rayes, "Multiobjective Optimization of Reality Capture Plans for Computer Vision–Driven Construction Monitoring with Camera-Equipped UAVs," *Journal of Computing in Civil Engineering*, vol. 36,

no. 5, Sep. 2022, doi: 10.1061/(asce)cp.1943-5487.0001032.

[10] S. Feroz and S. A. Dabous, “Uav-based remote sensing applications for bridge condition assessment,” May 01, 2021, *MDPI AG*. doi: 10.3390/rs13091809.

[11] D. Mader, R. Blaskow, P. Westfeld, and C. Weller, “Potential of UAV-Based laser scanner and multispectral camera data in building inspection,” in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, International Society for Photogrammetry and Remote Sensing, 2016, pp. 1135–1142. doi: 10.5194/isprsarchives-XLI-B1-1135-2016.

[12] Y. Tan, S. Li, H. Liu, P. Chen, and Z. Zhou, “Automatic inspection data collection of building surface based on BIM and UAV,” *Autom Constr*, vol. 131, Nov. 2021, doi: 10.1016/j.autcon.2021.103881.

[13] N. Kerle, F. Nex, M. Gerke, D. Duarte, and A. Vetrivel, “UAV-based structural damage mapping: A review,” Dec. 26, 2019, *MDPI AG*. doi: 10.3390/ijgi9010014.

[14] D. Mader, R. Blaskow, P. Westfeld, and C. Weller, “Potential of UAV-Based laser scanner and multispectral camera data in building inspection,” in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, International Society for Photogrammetry and Remote Sensing, 2016, pp. 1135–1142. doi: 10.5194/isprsarchives-XLI-B1-1135-2016.

[15] N. Kerle, F. Nex, M. Gerke, D. Duarte, and A. Vetrivel, “UAV-based structural damage mapping: A review,” Dec. 26, 2019, *MDPI AG*. doi: 10.3390/ijgi9010014.

[16] B. T. Wang and C. M. Wang, “Advances in 21st Century Human Settlements Automating Cities Design, Construction, Operation and Future Impact.” [Online]. Available: http://www.springer.com/series/13196

[17] Y. Tan, W. Yi, P. Chen, and Y. Zou, “An adaptive crack inspection method for building surface based on BIM, UAV and edge computing,” *Autom Constr*, vol. 157, Jan. 2024, doi: 10.1016/j.autcon.2023.105161.

[18] B. Felipe-García, D. Hernández-López, and J. L. Lerma, “Analysis of the ground sample distance on large photogrammetric surveys,” *Applied Geomatics*, vol. 4, no. 4, pp. 231–244, Dec. 2012, doi: 10.1007/s12518-012-0084-2.

Towards Agent: LLM-based Framework for Robotic Construction by Leveraging IFC

Song Du[1,†], Yan Gao[1,†], Wei Tong, and Yiwei Weng *

[1]Department of Building and Real Estate, The Hong Kong Polytechnic University, Hong Kong
*Corresponding author yiwei.weng@polyu.edu.hk
[†]These authors contributed equally to this work

Abstract

This work explores the integration of Large Language Model (LLM)-powered agents into robotic construction workflows. An LLM-based agent framework by leveraging standardized IFC format is proposed to guide robotic tasks, such as bricklaying and 3D concrete printing. Through a combination of tailored tools, prompt templates, and a *CustomMemory* module, the framework enables efficient extraction of building information and execution of robotic tasks. Finally, experimental validation in a simulated environment demonstrates the framework's potential to streamline robotic construction workflows from textual or voice commands to physical construction activities, while reducing redundant operations and enhancing reliability through thoughtful design.

Keywords – LLM Agents, Robotic Construction, IFC, BIM, 3D Concrete Printing

1 Introduction

The construction industry is witnessing rapid advancements in automation, driven by the need to address challenges such as large-scale tasks, stringent precision requirements, and dynamic environments. Robotics has emerged as a promising solution, enabling efficient and precise execution of tasks like bricklaying and 3D concrete printing. Despite significant progress in robotic construction, the integration of large language models (LLMs) into these workflows remains limited, particularly for tasks requiring seamless communication between building information modeling (BIM) data and robotic operations. While LLMs have shown great potential in robotics operations through their emergent and reasoning capabilities, their applications have mainly focused on the design stage, where generative tools powered by LLMs are widely adopted. However, their use in robotic construction processes remains largely unexplored.

This work addresses this gap by proposing an IFC-enhanced LLM agent framework for robotic construction. The framework leverages the standardized Industry Foundation Classes (IFC) format to extract building information efficiently and accurately. Furthermore, by incorporating tailored tools, structured prompt templates, and a designed *CustomMemory* module, the framework effectively guides robots in performing tasks such as bricklaying and 3D concrete printing. Experimental validation in a simulated environment demonstrates the framework's effectiveness in streamlining workflows, avoiding redundant operations, and enhancing reliability.

2 Related Works

2.1 Robotic Construction and LLM

LLMs have shown significant potential in robotics operations, enabling tasks such as natural language and code generation, numerical prediction, motion planning, and fault detection [1]. Works like RoboCodeX [2] and Reasoning Tuning [3] integrate visual and physical data in multimodal models to generate executable numeric predictions, achieving cross-platform and cross-scenario generalization by decomposing high-level semantics into traceable low-level actions. Methods like ReKep [4] enhance skill planning by identifying critical points, allowing robots to handle complex environments effectively, while Code-as-Monitor [5] combines Vision Language Models (VLMs) with real-time monitoring to predict risks and detect errors proactively.

Robots are advancing automation in construction, particularly in bricklaying and 3D concrete printing. For instance, mobile robots are used for automated bricklaying, integrated within BIM [6]. 3D concrete printing allows robotic arms to fabricate components efficiently, reducing material waste [7]. However, challenges in efficiency and structural stability persist, especially in large-scale projects. While robotic construction progresses, integration with LLMs remains limited. In contrast, LLM-driven tools like text2BIM [8] and AIstructure-Copilot [9] are making strides in generating layouts and BIM models from textual

descriptions during the design phase.

2.2 IFC Standardization

The IFC was introduced by buildingSMART International, to address interoperability challenges within the Architecture, Engineering, and Construction (AEC) and facility management (FM) industries [10]. IFC provides a standardized, neutral, and non-proprietary data format that spans the entire lifecycle of construction projects [11], including design, construction, and maintenance. By promoting seamless communication and reducing reliance on proprietary formats, IFC has become a universal standard in domains such as buildings, bridges, airports, and railways [12]. Leading AEC software platforms, including Autodesk and Bentley Systems, have adopted IFC to facilitate interoperability and enable efficient data exchange across stakeholders.

IFC's structured representation of building elements and attributes, accessible via tools like IfcOpenShell [13], supports automated construction workflows. By extracting dimensions, material properties, and placement information from IFC elements such as IfcWall and IfcBeam [14], the format facilitates integration with downstream processes like robotic assembly. Technologies like ifcJSON [15] further enhance this integration by converting IFC data into structured formats compatible with modern workflows. Leveraging IFC data in robotic construction, combined with LLM, has significant potential to address key industry challenges like large-scale task automation and precision requirements. For instance, LLM agents can potentially guide robots to efficiently and accurately perform complex tasks based on IFC files, including wall masonry and rebar tying.

3 Methodology

This work builds on prior research in the Text2BIM framework [8], [16], which utilizes LLM-based multi-agent systems to convert natural language commands into editable 3D building models. Extending this concept to robotic construction, our approach transforms engineers' natural language instructions into actionable robotic tasks based on BIM. A cornerstone of the approach is the use of IFC, offering a standardized and structured format for building design data to enable accurate extraction and utilization.

Directly processing IFC data with LLMs poses challenges due to its complexity and domain specificity, often resulting in parsing inaccuracies. Moreover, token limitations further restrict handling large files. To address these issues, IfcOpenShell [13] is employed for precise data extraction, effectively avoiding token limitations associated with processing large IFC files and ensuring efficient and structured retrieval of BIM data.

Hence, a suite of tools based on IfcOpenShell is developed for LLM-based agents to handle necessary calculations and generate operational code in robotic construction. The workflow, as shown in Figure 1, demonstrates seamless integration of tool usage, memory modules, and task performance to transform BIM models into robotic construction operation.

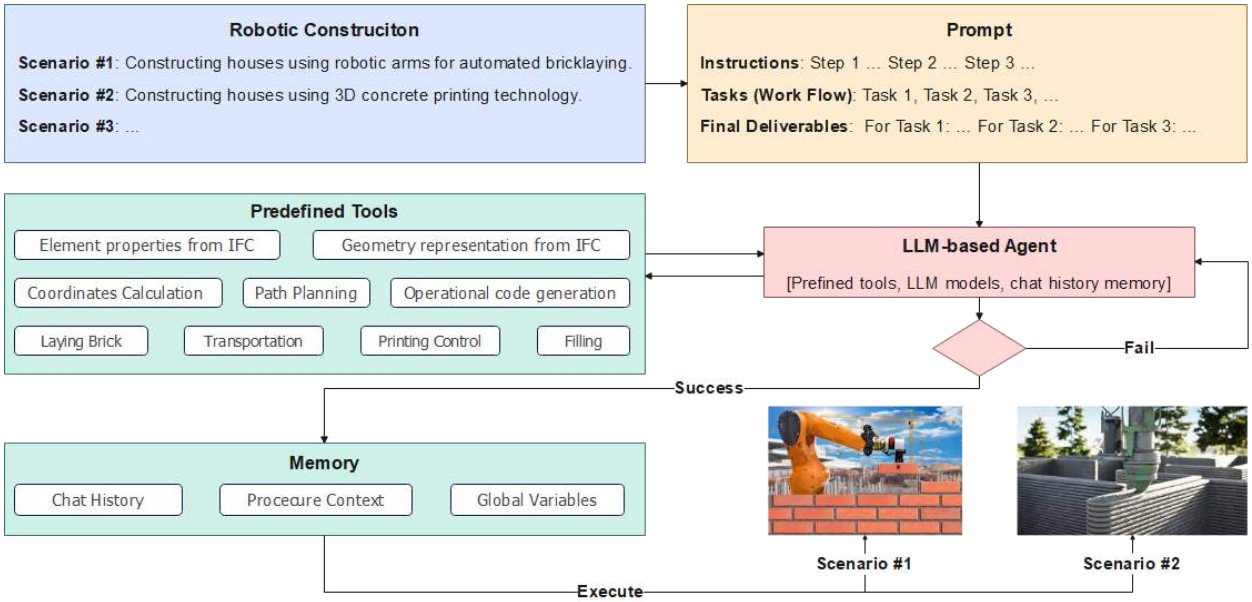


Figure 1. Proposed LLM agent-based Robotic Construction Framework

3.1 Toolset Design

Tools are designed from three key perspectives: (a) information extraction from IFC files, including both entity objects and geometric representations. Entity objects include building components such as walls, doors, windows, beams, columns, and roofs, while geometric representations cover attributes like shapes, placement, profiles, dimensions, extrusions, orientations, and spatial relationships. (b) essential calculations, such as brick coordinate calculation and printing path planning. (c) robotic operations, like bricklaying, printing control, material transportation, and filling.

Table 1 Examples of predefined tools

Name	Description	Outcome
LoadIFC File	Load an IFC file and return the ifcopenshell.file object.	The IFC file has been successfully loaded.
Extract Wall	Extract wall properties from an IFC file using ifcopenshell.	Wall properties are stored as variable *all_wall_props_global*
Calculate Bricks	Calculate brick coordinates based on wall dimensions and brick size. Input is a JSON format with keys: wall_properties, door_properties, window_properties, and brick_size.	Filtered brick coordinates are stored as variable filtered_brick_coordinates_global with {len(filtered_brick_coordinates_global)} records.
LayBricks	Lay bricks at the specified coordinates. Input is JSON format containing keys: filtered_brick_coordinates.	The system has successfully laid bricks at the specified coordinates.

The context of each tool is structured as three key components: 1) Name, to identify the tool during operation, 2) Description, to guide invocation and input format, and 3) Outcome, detailing the output variables and their storage locations, which are explicitly incorporated into memory module for LLM's contextual awareness. Table 1 presents examples of predefined tools and their descriptions. Prompt Template

The prompt template used in this work consists of three key components: Instructions, Tasks, and Final deliverables.

Instructions include directives to ensure: a) check for required data before each task; b) reuse available data from memory; c) use the provided tools and update results; d) transit between tasks with clear statements; 5) provide reasoning and results for each task before proceeding.

Task prompts are designed in two styles: a) open-ended prompt: providing a high-level instruction that allows the LLM to determine the appropriate tools and workflow based on its emergent and reasoning capabilities; b) tool-specific prompt: combining tool descriptions with input variables to guide precise tool invocation and task execution.

Final deliverables specify the expected output, which is stored in the memory module's chat history to ensure consistent progress tracking and provide context for subsequent tasks. Additionally, tools can be re-invoked at this stage to perform supplementary functions, such as saving output in a specific file format.

The tailored prompt template is shown in Figure 2.

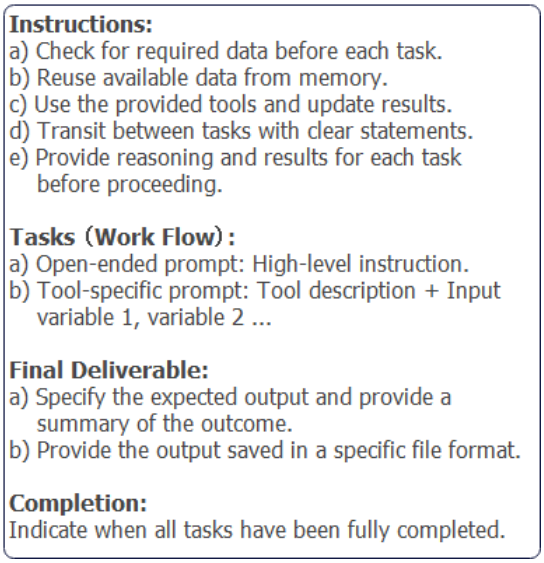


Figure 2. Structured prompt template

3.2 Memory Module

The memory module in this work consists of three components: 1) chat history, 2) output variables, and 3) procedure context. Using *ConversationBufferMemory* in LangChain [17], each prompt and the corresponding answers from the final deliverables section are recorded in the chat history. Meanwhile, the output variables during tool invocation are explicitly stored (e.g., as

global variables) and returned in an outcome summary, informing the agent to reuse these variables within the same session. However, once a session is completed, the outcomes generated during the process are discarded, leading to recalculations of previously generated variables in subsequent prompts.

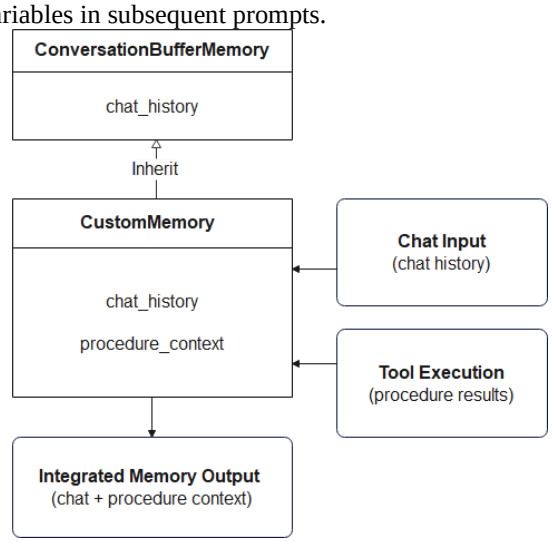


Figure 3. *CustomMemory* class and data flow

To address this issue, a *CustomMemory* class is designed, as shown in Figure 3, extending from *ConversationBufferMemory*, to retain chat history and store tool execution outcomes (i.e., procedure context). This integration enables the agent to review procedure results in any session, ensuring informed decisions and avoiding redundancy.

However, it is worth noting that memory content may continuously accumulate over multiple sessions, potentially leading to the scalability challenge. To mitigate this, efficient context management strategies, such as hierarchical summarization, may be required in future development to optimize memory usage while preserving relevant procedural data.

4 Proof of Concept

4.1 Preparation

To validate the effectiveness of the proposed LLM-based autonomous agent framework in robotic construction, experiments were conducted for two different scenarios: robotic bricklaying and 3D concrete printing. The experimental setup featured a simple structure comprising four walls, a door, and two windows, as shown in Figure 4. For simplicity, the bricks used in wall construction were cubic, with dimensions matching the wall thickness. Likewise, the layer height for 3D concrete printing was set equal to the wall thickness.

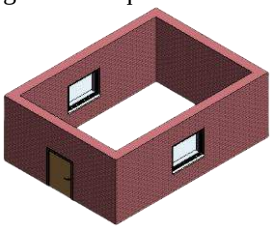


Figure 4. Experimental Structure

LangChain is adopted as the development foundation and a series of tools are developed for the experiment. The GPT-4 is utilized as the LLM engine, with the prompts shown in Figure 5.

Instructions:
0. Complete all tasks sequentially in the given order. Do not stop until all tasks are fully completed.
1. Before performing each task, explicitly use the QueryMemory tool to determine if the required information is already available, even if it is the first task.
2. If any part of the required information is already available in memory, reuse it instead of recalculating or re-extracting it.
3. Use the provided tools to complete the remaining parts of the tasks, storing results in the indicated global variables. Confirm that the variables are updated for subsequent use.
4. After completing each task, explicitly transition to the next task by stating, "Task [N] completed. Moving to Task [N+1]."
5. Provide clear reasoning and results for each task before transitioning to the next.

Tasks:
1. Load the project3.ifc file.
2. Extract all the elements from the loaded ifc object.
3. Calculate brick coordinates for the walls based on wall dimensions and brick size (300 mm cubic bricks), then filter out coordinates conflict with other structural elements or openings.
4. Generate 3D concrete printing paths for walls based on wall dimensions and layer height (300 mm), avoiding areas occupied by other structural elements or openings.
5. Lay bricks at the specified coordinates provided in the `filtered_brick_coordinates` input. Ensure all coordinates are processed, and return the result, including the number of bricks laid.
6. Print concrete along the 3D paths provided in the `layers_print_paths` input. Ensure all paths are processed, and return the result, including the number of layers printed.

Final Deliverable:
For Task 3: Save the filtered brick coordinates into a JSON file named 'filtered_brick_coordinates.josn'.
For Task 4: Save the 3D concrete printing paths into a JSON file named 'printing_paths.json'.
For Task 5: Summarize the brick-laying task, including the total number of bricks laid.
For Task 6: Summarize the 3D concrete printing task, including the total number of layers printed.

Completion:
Indicate when all tasks have been fully completed.

Figure 5. Prompts in the experiment

```
The first task is to read the project3.ifc file and initialize processing. However, before I do that, I need to check if the file has
already been loaded into memory.
Action: QueryMemory
Action Input: {}
Observation: Memory is empty. No prior actions or outputs are available.
Thought:Since the memory is empty, the project3.ifc file has not been loaded yet. I will now load the file using the LoadIFCFile tool.
Action: LoadIFCFile
Action Input: {"file_path": "project3.ifc"}{'file_path': 'project3.ifc'}
project3.ifc
<ifcopenshell.file.file object at 0x000001E5235BC610>

Observation: project3.ifc loaded successfully and stored as variable ifc_object.
Thought:The project3.ifc file has been successfully loaded and stored in the variable ifc_object. Task 1 is now completed. Moving to
Task 2, which is to identify and extract architectural elements from the IFC file.
Action: ExtractWallProperties
Action Input: {"ifc_object": "ifc_object"}
Observation: Wall properties stored as variable all_wall_props_global with 4 records.
Thought:The wall properties have been successfully extracted from the IFC file and stored in the variable all_wall_props_global. Now, I
will extract the door properties.
Action: ExtractDoorProperties
Action Input: {"ifc_object": "ifc_object"}
Observation: Door properties are stored as variable all_door_props_global with 1 records.
Thought:The door properties have been successfully extracted from the IFC file and stored in the variable all_door_props_global. Now, I
will extract the window properties.
Action: ExtractWindowProperties
Action Input: {"ifc_object": "ifc_object"}
Observation: Window properties are stored as variable all_window_props_global with 2 records.
Thought:The window properties have been successfully extracted from the IFC file and stored in the variable all_window_props_global.
Task 2 is now completed. Moving to Task 3, which is to calculate and filter bricks for potential 3D printing.
Action: CalculateAndFilterBricks
Action Input: {"wall_properties": "all_wall_props_global", "door_properties": "all_door_props_global", "window_properties":
"all_window_props_global", "brick_size": 300}
Observation: Filtered brick coordinates stored as variable filtered_brick_coordinates_global with 394 records.
Thought:The brick arrangement has been computed successfully, taking into account the size of the bricks and eliminating any conflicts
with doors and windows. The filtered brick coordinates are now ready for 3D printing path generation.
Action: Generate3DBrickPrintPath
Action Input: {"filtered_brick_coordinates": "filtered_brick_coordinates_global", "layer_height": 300}
Observation: 3D printing paths stored as the global variable layers_print_paths_global for 7 layers.
Thought:The 3D printing paths have been successfully generated and stored in the variable layers_print_paths_global. The paths are
structured into 7 layers, each with a height of 300 mm.
Action: QueryMemory
Action Input: {}
Observation:
### Procedure Context (Tools) ###
Tool: load_ifc_tool - Result: project3.ifc loaded successfully and stored as variable ifc_object.
Tool: extract_wall_tool - Result: Wall properties stored as variable all_wall_props_global with 4 records.
Tool: extract_doors_tool - Result: Door properties are stored as variable all_door_props_global with 1 records.
Tool: extract_windows_tool - Result: Window properties are stored as variable all_window_props_global with 2 records.
Tool: calculate_and_filter_brick_tool - Result: Filtered brick coordinates stored as variable filtered_brick_coordinates_global with
394 records.
Tool: calculate_and_filter_brick_tool - Result: 3D printing paths stored as the global variable layers_print_paths_global for 7 layers.

Action: ExecuteBrickPlacement
Action Input: {"brick_coordinates": filtered_brick_coordinates_global}
Action: {"StartBrickSupply", "RobotArmStart", "ArmMove from [supply_point] to [0.24,-0.36,0.15]",
 …,
"ArmMove from [supply_point] to [0.21,0,1.95]", "StopBrickSupply", "RobotArmStop"}
Thought:The brick laying task was completed, and a total of 394 bricks were laid.
Action: Execute3DPrinting
Action Input: {"print_paths": layers_print_paths_global}
Action: {"RobotArmStart", "MoveArm to layer_1_start", "StartExtrudeConcrete", "FollowPath layer_1_path", "FollowPath layer_2_path",
"FollowPath layer_3_path", "FollowPath layer_4_path", "FollowPath layer_5_path", "FollowPath layer_6_path", "FollowPath layer_7_path",
"StopExtrudeConcrete", "RobotArmStop"}
Thought:The 3D printing task is completed, with a total of 7 layers printed.
Final Answer: All tasks have been fully completed.
```

Figure 6. Tasks execution instructed by LLM agent

4.2 Experiment and Results

In the prompts, a robotic construction workflow is manually decomposed into a series of tasks. The experimental results demonstrate that the designed LLM-based agent can sequentially complete these tasks as specified, ultimately achieving robotic bricklaying and 3D concrete printing in a simulated environment.

The tasks execution instructed by the developed LLM agent is shown in Figure 6.

The tasks 5 and 6 execution instructed by the developed LLM agent is shown in Figure 7.

The experiments demonstrate that the proposed LLM-based framework can accurately invoke predefined tools according to the prompts. Moreover, the agent can consistently achieve reliable variable transmission through well-structured tool descriptions. For example, the building element properties are organized into *Action Input:{"wall_properties":all_wall_props_global,"door_properties":all_door_props_global,"window_properties":all_window_props_global,"brick_size":300}* based on the *CalculateBricks* tool description — "Input is a JSON format

with keys: wall_properties, door_properties, window_properties, and brick_size". Additionally, the agent can establish logical connections between tasks through well-designed outcome summaries in the results.

4.3 Efficiency and Reliability

The efficiency of *CustomMemory* is also validated in the experiment. For instance, when wall properties have already been extracted previously, and a task to extract properties of all elements is required, the agent can correctly identify that only the remained properties need to be extracted — *"The wall properties have already been extracted and stored in the variable all_wall_props_global. Now, I need to extract the properties of the doors and windows"*, demonstrating the ability to prevent the agent from redundant operations.

Similarly, both brick coordinates calculation and 3D printing path generation require extracting properties of all building elements. To quantify the efficiency improvement, we conducted controlled experiments to compare execution time with and without *CustomMemory* for the above tasks in a scenario where these properties had already been extracted in previous sessions.

As shown in Table 2 and Figure 8, the results indicate that *CustomMemory* reduces the execution time by over 60%. This is because, through the tailored procedure memory, the agent can recognize that the necessary data had already been extracted and stored, thereby avoiding unnecessary operations. In practice, the efficiency improvement is directly proportional to the number of previously generated and stored variables.

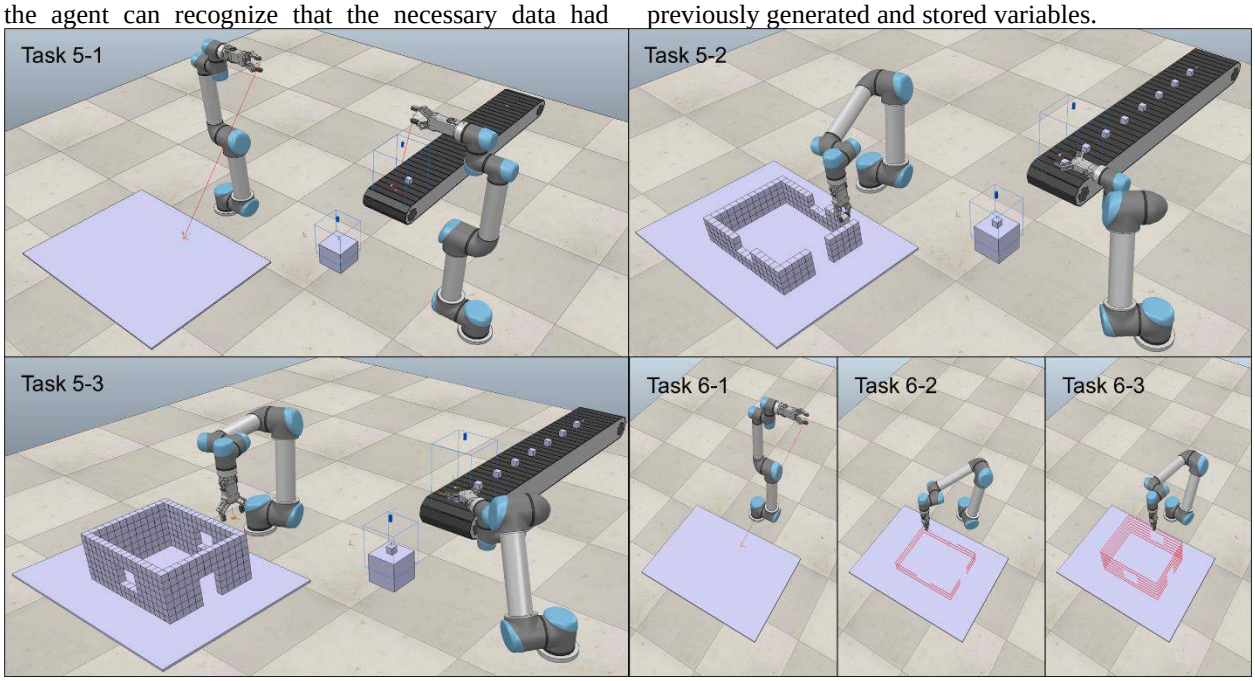


Figure 7. Tasks 5 and 6 executions instructed by LLM agent

Table 2 *CustomMemory* efficiency comparison

Tasks	Without CustomMemory	With CustomMemory
Brick Calculation	37.14s	13.67s
3D Path Generation	44.35s	15.87s

To assess the reliability of the proposed agentic framework, 10 similar prompts are generated using ChatGPT-4o to evaluate the execution success rate. Since the Instructions section is pre-designed as a fixed template, only the Tasks section is replaced in each prompt variation. Successful execution is measured based on how many tasks are correctly carried out.

The result shows that when the required parameters are explicitly specified (such as project3.ifc, 300-mm cubic brick, and 300-mm layer height), all 10 variations of the modified prompts successfully complete their respective tasks, demonstrating the excellent reliability of the proposed agentic framework and its robust adaptability to variations in input language.

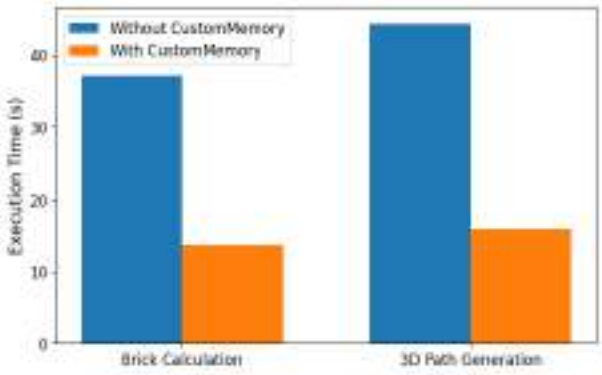


Figure 8. Execution time with vs without *CustomMemory*

However, it is worth noting that if these parameters are not explicitly provided (excluding variables generated dynamically by the agent during execution), the LLM tends to infer missing values, potentially leading to hallucinations. For example, it may assume a commonly used brick size, which could introduce unintended variations in execution.

5 Conclusion

This work presents an IFC-enhanced LLM agent framework for robotic construction, leveraging the standardized IFC format to enable efficient and accurate data extraction while avoiding computationally expensive approaches such as LLM fine-tuning or Retrieval-Augmented Generation (RAG). The tailored agent accurately invokes predefined tools through structured prompts, ensuring robust and reliable performance. Additionally, a *CustomMemory* module is developed to retain procedural context across sessions, significantly reducing execution time by eliminating redundant operations. The framework is validated in a simulation environment for robotic bricklaying and 3D concrete printing, demonstrating its effectiveness in automating construction workflows.

Furthermore, this work adopts an agentic workflow similar to Text2BIM, making it a natural extension of its approach. While Text2BIM primarily focuses on design-phase automation, this framework advances the concept by integrating LLM-driven robotic task execution, bridging the gap between BIM-based design generation and physical construction automation. Future research can explore the development of a full-process agentic pipeline, seamlessly connecting design-phase reasoning with robotic construction execution. This would enable an end-to-end intelligent workflow where LLMs autonomously generate, refine, and execute construction tasks based on BIM models, further enhancing efficiency, adaptability, customization, and automation in smart construction.

6 Discussion

Although the proposed framework demonstrates efficiency and reliability in simulation, transitioning to real-world robotic construction presents additional challenges. Integrating sensor feedback is crucial for real-time perception, as construction environments require adaptive decision-making based on LiDAR, cameras, and IMUs to handle unexpected obstacles. Error handling is essential for correcting deviations in robotic construction movements, such as bricklaying and 3D concrete printing, to ensure precise execution. Additionally, safety measures, including collision avoidance, emergency stop mechanisms, and regulation compliance, are required to ensure secure deployment in dynamic environments.

A promising solution to these challenges is the use of multiple agents to enhance coordination and execution. For example, instead of relying on a single agent, a multi-agent system could include: 1) a task planning agent, responsible for autonomous task decomposition and organization based on high-level construction goals; 2) an execution agent, handling low-level robotic control, ensuring precise movement, material placement, and process execution; 3) a supervisory agent, monitoring real-time sensor feedback, detecting errors, and triggering corrective actions when needed. By distributing responsibilities across specialized agents, this approach could improve reliability and scalability in real-world robotic construction.

Additionally, further investigation is needed to assess how the system handles ambiguous or conflicting instructions, particularly when multiple interpretations could lead to different robotic execution strategies. The risks of LLM hallucinations, such as generating incorrect or assumed parameters, require quantitative analysis to evaluate error rates and system robustness. Moreover, a deeper exploration of failure cases and error recovery mechanisms will be essential to ensuring the framework’s adaptability and robustness in real-world robotic construction.

Furthermore, while the framework has already demonstrated some chain-of-thought (CoT) reasoning in its current stage. For example, executing Tasks 1, 5, and 6 in Figure 5 (skipping Tasks 2, 3, and 4), the agent can still extract building elements and their properties, as well as perform necessary calculations based on its reasoning and tool descriptions. In future work, we will continue to explore leveraging LLM's CoT capabilities to further enhance autonomous task decomposition and organization, improving the framework’s adaptability in complex construction workflows.

References

[1] J. Wang *et al.*, 'Large Language Models for Robotics: Opportunities, Challenges, and Perspective', Jan. 2024, [Online]. Available: http://arxiv.org/abs/2401.04334

[2] Y. Mu *et al.*, 'RoboCodeX: Multimodal Code Generation for Robotic Behavior Synthesis', Feb. 2024, [Online]. Available: http://arxiv.org/abs/2402.16117

[3] J. Xu, S. Jin, Y. Lei, Y. Zhang, and L. Zhang, 'RT-Grasp: Reasoning Tuning Robotic Grasping via Multi-modal Large Language Model', Nov. 2024, [Online]. Available: http://arxiv.org/abs/2411.05212

[4] W. Huang, C. Wang, Y. Li, R. Zhang, and L. Fei-Fei, 'ReKep: Spatio-Temporal Reasoning of Relational Keypoint Constraints for Robotic Manipulation', Sep. 2024, [Online]. Available: http://arxiv.org/abs/2409.01652

[5] E. Zhou *et al.*, 'Code-as-Monitor: Constraint-aware Visual Programming for Reactive and Proactive Robotic Failure Detection', Dec. 2024, [Online]. Available: http://arxiv.org/abs/2412.04455

[6] B. Xiao, C. Chen, and X. Yin, 'Recent advancements of robotics in construction', Dec. 01, 2022, *Elsevier B.V.* doi: 10.1016/j.autcon.2022.104591.

[7] Y. Weng *et al.*, 'Comparative economic, environmental and productivity assessment of a concrete bathroom unit fabricated through 3D printing and a precast approach', *J Clean Prod*, vol. 261, Jul. 2020, doi: 10.1016/j.jclepro.2020.121245.

[8] C. Du, S. Esser, S. Nousias, and A. Borrmann, 'Text2BIM: Generating Building Models Using a Large Language Model-based Multi-Agent Framework', Aug. 2024, [Online]. Available: http://arxiv.org/abs/2408.08054

[9] S. Qin *et al.*, 'AIstructure-Copilot: Assistant for Generative AI-Driven Intelligent Design of Building Structures', *Smart Construction*, Mar. 2024, doi: 10.55092/sc20240001.

[10] 'buildingSMART International'. Accessed: Jan. 15, 2025. [Online]. Available: https://www.buildingsmart.org/

[11] S. T. Matarneh, M. Danso-Amoako, S. Al-Bizri, M. Gaterell, and R. Matarneh, 'Building information modeling for facilities management: A literature review and future research directions', Jul. 01, 2019, *Elsevier Ltd*. doi: 10.1016/j.jobe.2019.100755.

[12] 'Domains - buildingSMART International'. Accessed: Jan. 15, 2025. [Online]. Available: https://www.buildingsmart.org/standards/domains/

[13] 'IfcOpenShell - The open source IFC toolkit and geometry engine'. Accessed: Jan. 15, 2025. [Online]. Available: https://ifcopenshell.org/

[14] L. Zhang and N. M. El-Gohary, 'Automated IFC-based building information modelling and extraction for supporting value analysis of buildings', *International Journal of Construction Management*, vol. 20, no. 4, pp. 269–288, Jul. 2020, doi: 10.1080/15623599.2018.1484850.

[15] 'buildingsmart-community/ifcJSON: Repository containing the specification for IFC.JSON'. Accessed: Jan. 15, 2025. [Online]. Available: https://github.com/buildingsmart-community/ifcJSON

[16] C. Du, S. Nousias, and A. Borrmann, 'Towards a copilot in BIM authoring tool using a large language model-based agent for intelligent human-machine interaction'. [Online]. Available: https://huggingface.co/docs/transformers/transformers_agents

[17] 'LangChain'. Accessed: Jan. 15, 2025. [Online]. Available: https://www.langchain.com/

Advances in the Development of a Custom End Effector for Construction Automation

Dax K. Pimentel de la Rosa [1,2], Mario J. Cervantes Hernandez [1,2], Jiansong Zhang [1], Luis C. Félix-Herrán [2]*

[1] Polytechnic Institute, Purdue University, United States
[2] Tecnologico de Monterrey, School of Engineering and Sciences, Mexico
* corresponding author
daxpimentel@outlook.com, mxcervantes7@gmail.com, zhan3062@purdue.edu, lcfelix@tec.mx

Abstract –

Beyond increasing productivity, improving quality, performing hazardous tasks and reducing costs, robotics also offers sustainable solutions by optimizing the use of materials and improving energy efficiency. In addition, modern robots are highly adaptable platforms that result in more intelligent systems capable of performing complex tasks. The following article presents the advances in the development of an end effector that promises to be a more economical and efficient option, which is used together with a single robotic arm to allow for the gripping, positioning and fixing of construction panels. The focus of this research effort has been the programming of a digital twin in the Robot Operating System (ROS) framework which allows to analyze the kinematics of the robot and the end effector working together; in addition, the redesign of the physical connectors between the robot and the end effector is also discussed. It is known that nowadays digitization, simulation and the use of digital twins in robotic systems are essential to maximize efficiency and accuracy in the modern construction industry; besides, they have become competitive factors that give added value to construction processes.

Keywords –

End Effector; Robotic Arm; Digital Twin; Construction Robotics; Automation in Construction

1 Introduction

Population growth and urbanization have increased the demand for housing and buildings. However, construction faces a shortage of qualified personnel due to the risks and efforts involved, discouraging young people as the current workforce emigrate, age, or can no longer continue their work due to accidents or other complications. The construction industry is a sector of the global economy exposed to significant risks of workers' health and safety. In general terms, this is a sector with high ergonomic risks in its execution due to lifting heavy weights, handling tools that can be dangerous in their use, and in general, because most of the activities are carried out in environments where there are constantly changing and dangerous situations. Consequently, falls and injuries related to the use of tools and construction materials are relatively frequent.

The high physical demands of construction work, coupled with the prevalence of injuries, underscore the urgent need for innovative solutions to improve safety and efficiency. For instance, the repetitive task of handling plywood sheets, often weighing fifty pounds or more, exemplifies the ergonomic and safety challenges faced by workers. Manual handling exposes workers to risks of back injuries and fatigue, which could also contribute to workflow inefficiencies. Additionally, according to the International Labor Organization (ILO), construction workers experience some of the highest rates of occupational accidents and injuries, the leading causes being falls, being struck by objects, and repetitive strain injuries [1]. Furthermore, just in 2022, the United States Occupational Safety and Health Administration (OSHA) reported over 5,000 fatalities in the construction industry, accounting for nearly 20% of all workplace deaths [2]. Furthermore, non-fatal injuries, which include musculoskeletal disorders and cuts, contribute to billions of dollars in compensation claims, lost productivity, and medical expenses annually, according to the Workplace Safety Index, in 2024, $58.07 billion dollars were invested in compensation for worker injuries, the main cause being excessive physical effort involving external sources, which represents 21.5% of the total [3].

The use of robots and automation technologies in construction is increasingly heard, such is the example of the work conducted by T. Hachijo. and S. Igarashi [4] where they describe the development of an autonomous mobile system of two robots for construction called “Robo-BC”. In addition, incredible advances are seen in the way these types of robotic systems are trained, such as in the case of the work of L. Ci-Jyun and collaborators [5] where they teach a robotic arm to place panels on the

ceiling through videos, along with which they call "Learning from Demonstration".

The creation of digital twins continues to revolutionize all industries, A. Ali created a digital twin that identifies dangerous areas and tracks worker movements to ensure safety [6]. Another interesting way to take advantage of the digitalization of cyber-physical systems is the remote control and monitoring of applications where robotic elements are employed [7]. In this context, robotics has proven to be a fundamental ally in addressing the challenges of a society that demands more and more infrastructure solutions that are faster to build, affordable, and with fewer risks for those who build them.

To address these challenges, this research effort focuses on the improvement and continuation of an end effector conceived from the patent of Zhang J. et al [8], which gave rise to the work by Borgo et al. [9], where details of the design and construction of the end effector are provided. This first approach was designed for a commercial robotic arm available on the market, capable of grasping and placing plywood sheets, as shown in Figure 1, as well as manipulating them and nailing them in the context of residential construction projects. By leveraging robotics, the system aims to improve safety, precision and efficiency, thereby transforming conventional construction methods. The overall goal of this work is to create more economically accessible automation tools that improve the safety and efficiency of building construction.

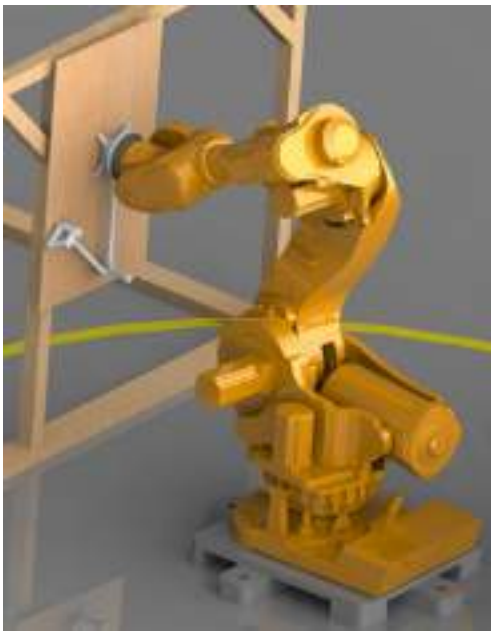


Figure 1. Render in SolidWorks of a robotic arm and end-effector placing a plywood panel on a wooden frame.

2 Methodology

Two significant advances were made in this work, the first advance was to analyze and make updates to the original design of the end effector to improve the robustness of the physical connection to the robot flange. [9]; the second was to use the ROS framework to simulate the motion of the end effector together with a robotic arm to test algorithms safely and inexpensively.

2.1 Observation

A comprehensive review of the lineage project documentation was undertaken to gain a thorough understanding of the target end effector's functionality and goals [9]. In addition, a detailed inventory of the materials, components, and resources available for the project was compiled. With this knowledge, a comprehensive literature review was conducted using SCOPUS as the primary database.

2.2 Proposal

Based on the gathered information regarding the project's current state and available resources, some proposals were formulated to enhance the development of the end effector, culminating in the objectives outlined at the beginning of this article. Besides, a workflow diagram was designed as shown in Figure 2, for the period of August to December 2024 as the proposed allocated timeframe.

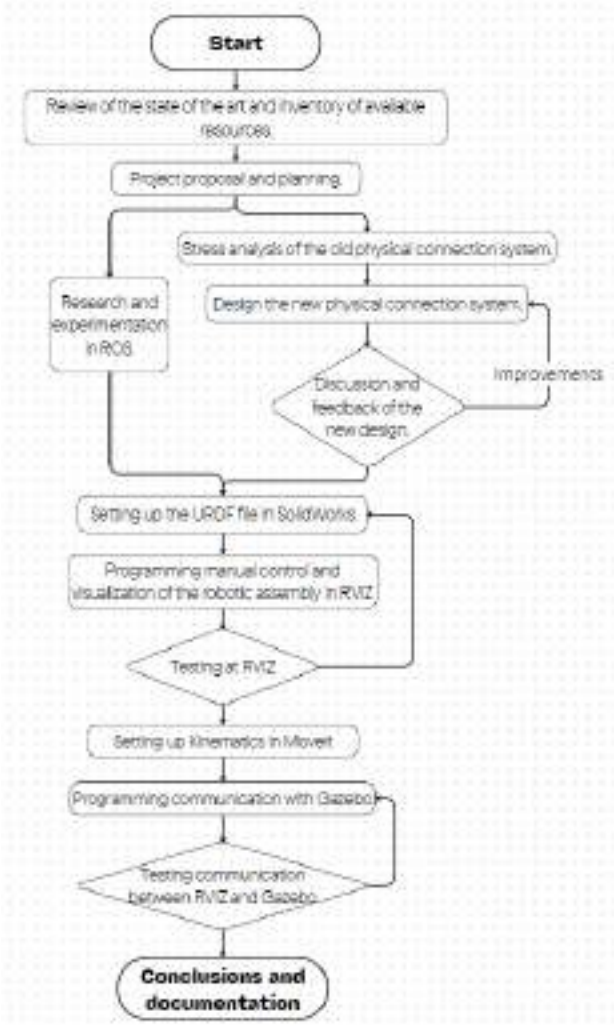


Figure 2. Workflow diagram showing the flow of work performed.

2.3 Experimentation and Results

Research was conducted to learn about the use and functionality of ROS, Gazebo and RVIZ for implementing project digitalization. Also, adjustments were made to the original design to attach the end effector to a new IRB7600 robotic arm, with heavier load capacity and simulating the real-world application better than the previous arm used, as well as certain improvements to strengthen the new component that connects these two elements. Furthermore, several programs were developed within the ROS framework to simulate the joint movement of an IRB7600 robot and the end effector, and some opportunity areas were found to carry out updates on the original design of the end effector. The modifications are aimed at improving performance, expanding the operation and improving the robustness of the end effector.

3 End effector Design

This robotic tool, as shown in Figure 3, possesses two degrees of freedom. The primary degree of freedom involves a rotary vacuum system equipped with suction valves to rotate a cross-shaped piece. The secondary degree of freedom allows the elbow of the end effector arm to rotate, enabling the nail gun to move in two dimensions within a circular area of approximately 4.15 square meters.

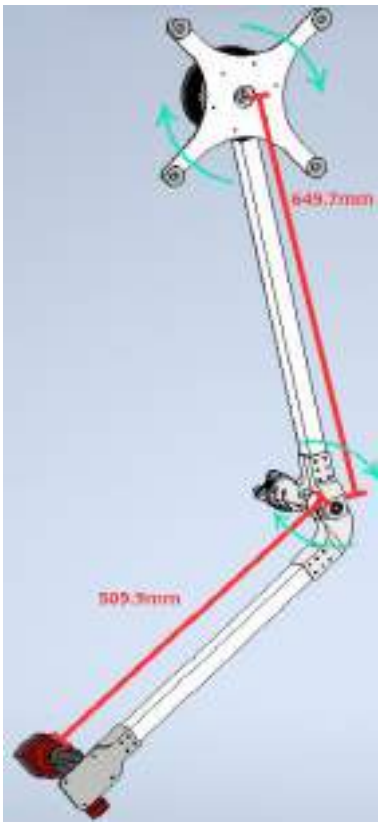


Figure 3. End effector assembly in Inventor.

3.1 Connector Design

The two components that physically connect the robot and the end effector have been redesigned [9]. The initial designs were intended for a KUKA KR16 L6, as shown in Figure 4. This component is attached to the robot's end and features small teeth that engage with the end effector. This initial design highlighted the importance of a simple coupling system that allows for easy separation and attachment of the end effector to the robot. In addition, it offers a cost-effective solution for attachment to other robots, as only this component needs to be redesigned and manufactured to suit the specific dimensions of the robot flange. However, a stress analysis, as shown in Figure 5, where a distributed load of 14 kg was applied, approximately equivalent to the weight of the end

effector and an attached wooden panel, shows a critical concentration of stresses in regions susceptible to shear failure. Furthermore, this component is made of an ABS plastic material, which is not flexible, and the two protruding teeth are too thin.

For practical and prototyping purposes, all components were analyzed using ABS plastic, as it is suitable for 3D printing and, unlike conventional PLA, offers greater strength and better temperature resistance. Additionally, ABS is more accessible, and less abrasive compared to materials like carbon fiber. With this approach, the aim is to create a low-cost prototype to validate the design and functionality before manufacturing it in a more robust and expensive material like aluminium.

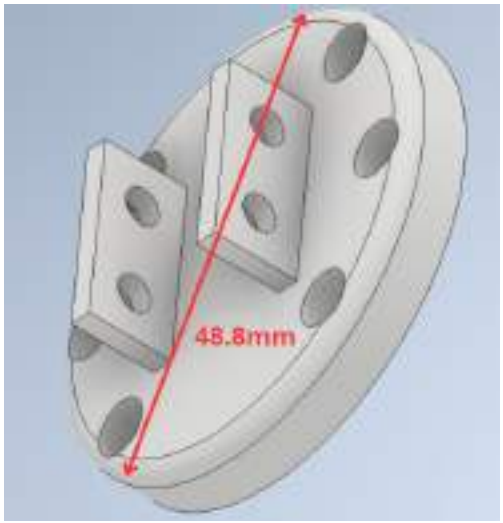


Figure 4. Old connector for the flange of a KUKA KR16 robot.

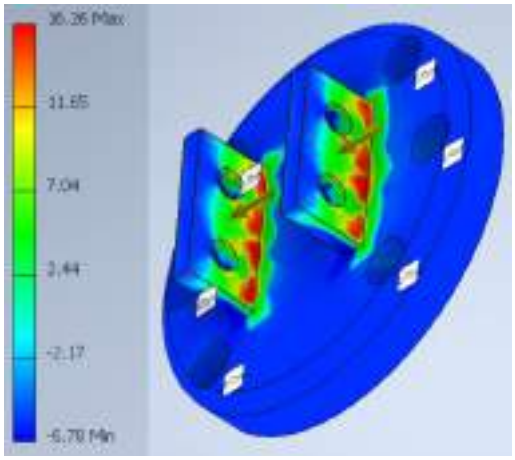


Figure 5. Stress analysis of the old connector. A total distributed load of 14 kg was applied. The color bar indicates the stress levels in MPa.

Due to the above, a new connection system, as shown in Figure 6, was designed to attach the end effector to an IRB7600 robot and incorporate physical features that improve its structural integrity by better distributing the stress caused by the loads, thus reducing the risk of shear failure in this component.



Figure 6. The assembly of the new connection system between the robot and the end effector. The white piece corresponds to the connection with the robot's flange, while the black piece is the base of the end effector.

In this new design, as shown in Figure 7, two additional teeth were added, and the design was smoothed with chamfers and fillets, resulting in a larger contact area and fewer stress concentrators. The original design had stress concentrations of up to 16.26 MPa in critical areas, as shown in Figure 5, while the new connector has only small stress concentrations in non-critical points, reaching a maximum of only 0.705 MPa, as shown in Figure 8. The new design also improves stress redirection during rotation thanks to the circular symmetry of the four teeth. Another notable aspect is that assembly only requires four screws.

Additionally, in this same stress analysis, it was found that the old connector exhibited a displacement of 0.2318 mm in the two connector teeth, compared to the new design, which showed a maximum displacement of only 0.0089 mm. This provides a clearer indication of the improved resistance to deformation and stress accumulation in the new connector. By reducing these factors, the new design enhances durability, especially considering that ABS material has very low flexibility, and the layer-by-layer nature of 3D printing makes it more susceptible to breaking even under slight deformations.



Figure 7. The new connector for the flange of an IRB7600 robot.

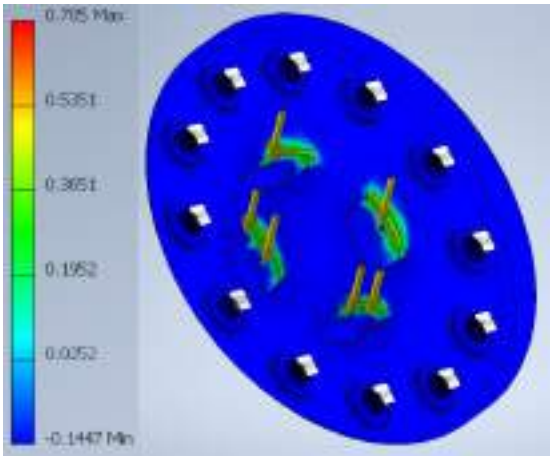


Figure 8. Stress analysis of the new connector for the flange of an IRB7600 robot. A total distributed load of 14 kg was applied. The color bar indicates the stress levels in MPa.

Due to the new robot connector, the end effector connector had to be adjusted, as shown in Figure 9. As with the previous parts, a stress analysis was performed, which shows that there are no stress concentrations at cut points, as shown in Figure 10.

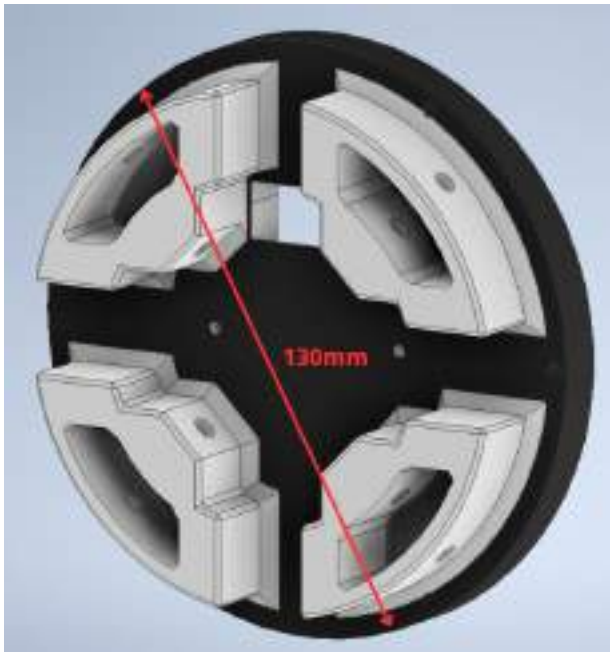


Figure 9. The new connector for the end effector.

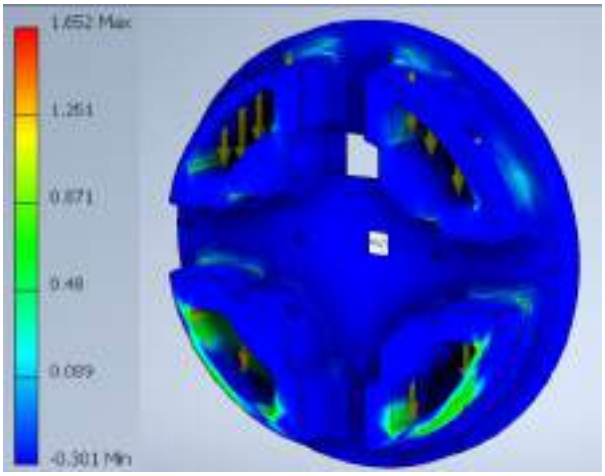


Figure 10. Stress analysis of the new connector for the end effector. A total distributed load of 14 kg was applied. The color bar indicates the stress levels in MPa.

4 ROS Simulation

The ROS structure allows decomposing a robotic system into nodes that communicate with each other through messages. This facilitates the creation of complex simulations and the integration of different sensors and actuators. With this approach, more functions and components are added to the representation towards improving the capabilities of the herein end-effector, all integrated into a single robust and efficient operating system.

After several tests to program and configure the simulation of a robotic arm and the end effector in ROS,

three significant advances were reached for this digitization. The first breakthrough was to visualize the robotic assembly in RVIZ, the second was to generate the inverse and direct kinematics of the robot with Moveit, finally, communication to a digital twin in Gazebo was achieved.

4.1 Configuration of the URDF file in SolidWorks 2021

Specialized Software for computer aided design was employed to model and assemble an IRB7600 robot and the proposed end effector. SolidWorks has an extension that allows the use of URDF (Unified Robot Description Format) file where the details of the robot are described, such as 3D models, joints, hierarchies, type of movements, masses of components, inertial properties, collision meshes, among other factors. The IRB7600 robotic arm, as shown in Figure 11, has six degrees of freedom and with the end effector they total eight degrees of freedom.

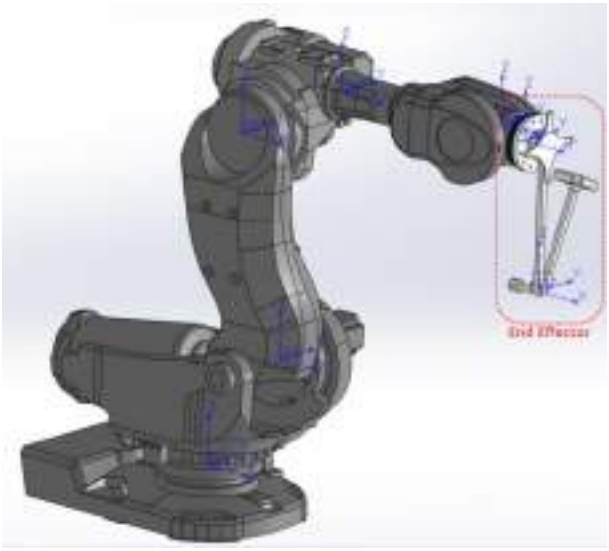


Figure 11. SolidWorks assembly of the IRB7600 robot with the end effector and in blue the coordinates of the joints of both systems.

4.2 Visualization of the URDF file in RVIZ

RVIZ is a tool that allows 3D visualization, as shown in Figure 12, within the ROS ecosystem. Thanks to this, a program, as shown in Figure 13, was designed and coded. In his algorithm, the IRB7600 robot and the end effector are displayed together in the RVIZ environment. In this program it is possible to manually control each of the degrees of freedom of the system independently. To obtain this result it was necessary to generate the URDF file.

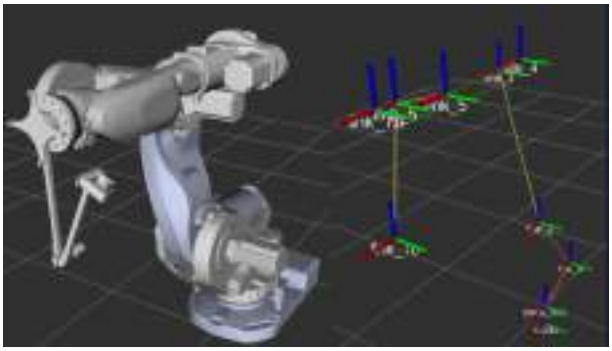


Figure 12. Visualization of the robotic assembly and its kinematic chain in RVIZ.

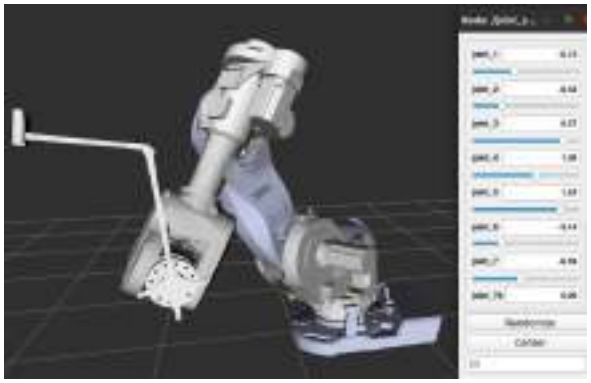


Figure 13. Program visualized in RVIZ that allows to manipulate each degree of freedom of the robotic system.

4.3 Robot Kinematics with Moveit software

The “Moveit” application was used as a framework to create the kinematics and control of the robot and to define some poses to test the movements of the assembly. As can be seen in Figure 14, with this tool the movements and groups of the robotic assembly are planned. In addition, it allows generating YAML (YAML Ain't Markup Language) files that were used to configure the controllers and startup parameters for the nodes.

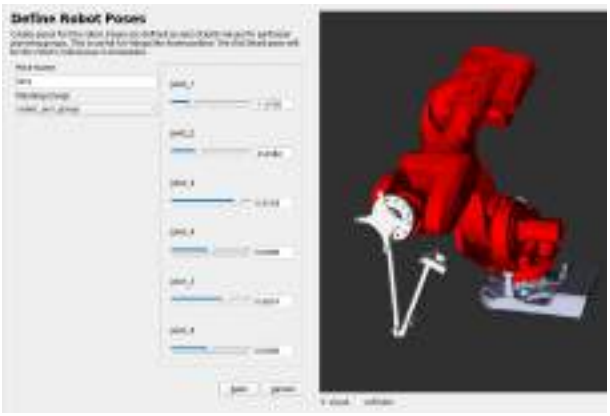


Figure 14. Moveit assistant to define positions.

4.4 Twin communication in Gazebo

This program displays two windows, the first one shown in Figure 15, depicts the robotic system in RVIZ with the displacement and rotation axes of the robot flank. The program calculates the inverse kinematics of the robot, performs several iterations from the original point to the target point, generating a linear displacement between both points.

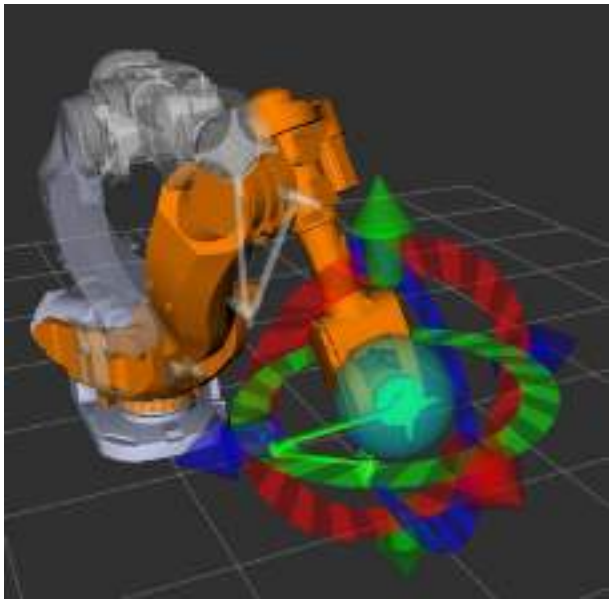


Figure 15. Inverse kinematics of the robotic system in RVIZ.

After the program calculates the robot's displacement path, it performs the movement and communicates it to the twin in Gazebo. This is included in the second window, as shown in Figure 16, and replicates the previous movements.

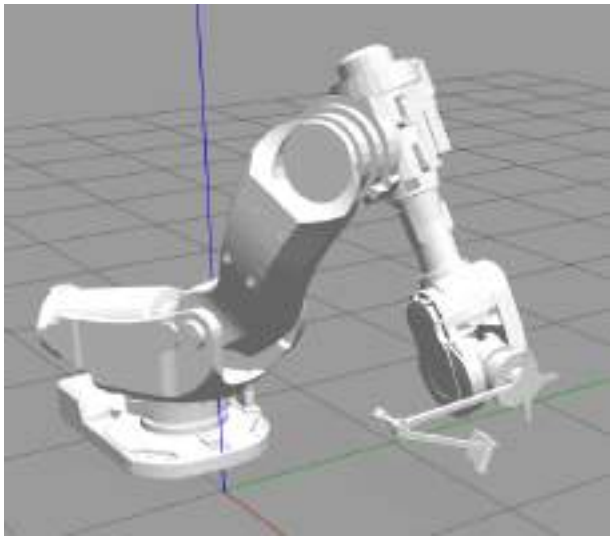


Figure 16. Visualization of the twin robotic system in Gazebo.

Unlike traditional ROS-based simulations, the herein work develops and implements the fundaments of a new frame that integrates ROS for Digital Twin. This integration provides enhanced real-time simulation and visualization capabilities that are not typically available in standard ROS-based digital twin frameworks. In another hand, the proposed approach aims to provide technical capabilities to the ARM and End Effector System for prediction. By leveraging ROS simulation environment, the ability to model and anticipate robotic interactions is improved by enhancing decision-making processes and reducing errors in real-world deployment. Furthermore, although digital twins are broadly used, this work intends to apply this technology to the construction engineering field. To create a training environment for construction workers, allowing them to interact with robotic systems learning process before real-world implementation. This addresses a critical gap in robotics education and deployment in construction.

5 Results and Conclusions

Progress in the design of an end-effector for a robotic arm, oriented to automated construction tasks has been presented. This manuscript introduces the development, analysis, and refinement of a patented robotic end effector for timber construction. Improvements in the design of the connection between the robot and the end effector provide greater robustness due to a larger contact area and better distribution of stress. Furthermore, the connector's teeth allow for easier coupling and decoupling and offer the advantage of physical

compatibility with other robots, as it is easier to design a single piece that does not modify the function or structure of the end effector.

During the programming of the end effector and robotic arm in ROS, an inconvenience was found with the current design of the end effector, which prevented it from acting independently of the robotic arm. This represents a problem as it complicates the algorithm necessary to synchronize the robotic arm and the end effector and prevents the same programming from working for a different robot, meaning that the robot's kinematics would have to be recalculated and reprogrammed each time it is attempted to use on a different robotic arm. The implementation in ROS of the proposed solution digitalization brings a new perspective on the project's scope, as it allows for a faster development pace by eliminating the need for a physical robot. This enables safe testing of algorithms, making changes, and adding more features to the simulation.

The time required to recalculate the robot's kinematics may vary depending on its design and degrees of freedom. However, this can be avoided by implementing a design in which the robotic arm's end-effector can be isolated to function independently of the robot. This would also provide the advantage of using the end-effector for testing or other tasks without the need for a physical connection to the robot.

This improved end-effector design will increase the lifespan of the robotic system. Stress concentrations in the critical areas where the end-effector connector and the robotic arm met, were reduced, as well as the deformations caused by the load from the actuator and the plywood. This enhancement will improve task safety and the precision with which the plywood is placed. Furthermore, the automated system will reduce construction times for handling, placing, and securing the plywood, thereby reducing construction costs.

6 Future Work

Thanks to the use of cutting-edge, the kinematic behavior of the robotic assembly has been understood in greater detail, allowing the identification of areas for improvement in the end effector design. As a result, it is planned to make improvements to the original design, leading to a more robust and functional version. It is also planned to continue programming and simulation in ROS, aiming to create a digital twin that integrates other functions such as fault anticipation, maintenance support, and other aspects and characteristics of a digital twin.

As part of the future work, an effort to mitigate the effects of latency and noise in the proposed system could be implemented. Techniques such as Extended Kalman filtering (EKF) and fusion sensor techniques could be applied to eliminate or minimize these issues and enhance system accuracy. Additionally, it is planned to implement robust communication protocols, including I2C, UART, and Serial, to minimize latency and noise in data transmission between the robotic arm and the end effector. These improvements will contribute to a more realistic and reliable digital twin representation.

References

[1] International Labour Organization (ILO). Safety and health at work. On-line: https://www.ilo.org. Accessed: 02/12/2024.

[2] United States Occupational Safety and Health Administration (OSHA). Commonly Used Statistics. On-line: https://www.osha.gov Accessed: 01/03/2025.

[3] Liberty Mutual Insurance. 2024 Workplace Safety Index. Liberty Mutual Business Insurance. On-line: https://business.libertymutual.com/insights/2024-workplace-safety-index/ Accessed: 01/03/2025.

[4] T. Hachijo and S. Igarashi. Autonomous Robot for Ceiling Board Construction Work “Robo-Buddy Ceiling”. *2023 IEEE 19th International Conference on Automation Science and Engineering (CASE)*, pages 1-7, Auckland, New Zealand, 2023. doi: 10.1109/CASE56687.2023.10260608.

[5] L. Ci-Jyun, R. Vineet Kamat, C. Carol Menassa. Teaching robots to perform quasi-repetitive construction tasks. 2020, *Automation in Construction*. Volume 120. Article 103370, 2020.

[6] A. Ali Khairadeen Digital Twin Framework for Real-time Monitoring of Robot-Arm in Construction Site. *Passer Journal of Basic and Applied Sciences*, 2024, pages 173-199. doi: 10.24271/psr.2024.188549.

[7] Q. Tan, M. Denojean-Mairet, H. Wang, et al. Toward a telepresence robot empowered smart lab. *Smart Learn. Environ. 2019.* doi: 10.1186/s40561-019-0084-3.

[8] J. Zhang, C. M Lacny, and N. Reardonet. Autonomous robotic system for placing and fastening paneling material for building construction operations, September 05, 2023. U.S. Patent Number: 11745356.

[9] A. Borgo, A. Ruiz, J. Zhang and L. C. Félix-Herrán Implementation of a Robotic Manipulator End Effector for Construction Automation. In *Proceedings of the 41st International Symposium on Automation and Robotics in Construction (ISARC)*, pages 177-184, Lille, France, 2024.

Exploration of Sustainable Beam Design Using Voronoi Patterns in Concrete 3D Printing

Gokul Santhosh S[1], Benny Raphael[2], Manu Santhanam[2]

[1]Ph.D. Research Scholar, Building Technology and Construction Management Division, Dept. of Civil Engineering, IIT Madras, Chennai 600036, India
[2]Professor, Building Technology and Construction Management Division, Dept. of Civil Engineering, IIT Madras, Chennai 600036, India
ce22d002@smail.iitm.ac.in, benny@civil.iitm.ac.in, manus@civil.iitm.ac.in

Abstract –

The emergence of concrete 3D printing presents a transformative opportunity for decarbonizing the construction industry by enabling innovative design solutions that prioritize material efficiency and sustainability. While 3D-printed lattice beams have demonstrated significant material reductions using conventional, repetitive patterns, nature-inspired shapes have not been explored. This study investigates the potential of nature-inspired Voronoi patterns for material-efficient concrete beams. Their performance is compared with traditional Truss patterns for symmetrical and unsymmetrical loading conditions. A parametric design system integrated with a genetic algorithm is used to generate various infill patterns for 3D-printable beams. Beam dimensions, orientations, and configurations are optimized while ensuring adequate structural performance. The findings confirm the feasibility of using nature-inspired Voronoi patterns in concrete 3D printing and demonstrate substantial material savings. This approach not only reduces embodied carbon in construction but also enables innovative architectural designs.

Keywords –

Concrete 3D Printing; Voronoi; Shape Optimization; Nature-inspired patterns

1 Introduction

The construction industry is one of the major contributors to carbon emissions, with CO_2 from cement production reaching a staggering 2.9 billion tons in 2021 [1], [2]. Traditional construction methods, which have long dominated the sector, are increasingly unable to meet the growing economies' demands and satisfy sustainability goals at the same time. As a result, optimizing construction processes and materials is crucial for advancing sustainable practices and meeting modern-day demands within the industry.

Additive manufacturing, particularly concrete 3D printing, is emerging as a transformative technology with the potential to address these challenges. By eliminating the need for formwork and enabling direct use of CAD files for construction [3], concrete 3D printing offers significant cost savings, particularly in formwork materials, which typically account for 25-30% of total structural construction costs [4]. Beyond economic benefits, this technology also provides architects and engineers with unprecedented design flexibility, enabling the creation of complex and unconventional geometries that would be difficult or impossible to achieve using traditional methods.

The technology is rapidly advancing, fueled by ongoing research aimed at improving efficiency and broadening adoption. Early studies in concrete 3D printing focused on developing printing technologies [5], [6], optimizing printable mix designs with specific rheological properties [7], [8], integrating supplementary cementitious materials (SCMs) for sustainability [9], and investigating the impact of print parameters on material properties. However, most research has centered on vertical structural elements, such as columns and wall panels, with little attention given to horizontal members [10]. These horizontal elements, such as beams, represent around 43% of the material used in reinforced concrete construction [11]. To further advance concrete 3D printing, research must shift towards optimizing these horizontal, flexural members for improved material efficiency.

Some studies have explored innovative approaches to 3D printing beams. For instance, one method involved assembling smaller units with rebars to print beams, like the process for printing columns [12]. However, this modular approach doesn't fully capitalize on the advantages of 3D printing. Other research focused on optimizing beam structures by printing hollow boundaries and filling them with concrete conventionally [13], but this also fails to maximize the benefits of 3D

concrete printing. Recent studies have tested the bending performance of 3D-printed lattice beams with various patterns, showing that triangular lattice beams offer superior performance [14]. Yet, these studies typically use fixed lattice configurations and do not incorporate shape-optimized geometry, which would have further enhanced material efficiency and load-bearing performance.

One promising approach to optimizing 3D-printed beams involves using shape optimization techniques to minimize material usage while maximizing flexural strength. Prior studies have demonstrated the potential of shape-optimized 3D-printed reinforced concrete beams [15], but scalability remains a challenge, particularly as load complexity increases. To make concrete 3D printing a viable solution for large-scale construction, further research is needed to develop optimized, scalable beam designs with a high strength-to-weight ratio.

Structural optimization plays a key role in improving material efficiency, but the complexity of optimized geometries can pose challenges for traditional construction methods, especially when it comes to fabricating formwork. Concrete 3D printing, with its ability to produce complex geometry, aligns well with structural optimization, enabling more sustainable and efficient architectural designs. Nature, with its inherent resource efficiency, has long been a source of inspiration for optimization strategies in materials and structures [16]. However, applying nature-inspired designs in conventional construction has been limited by practical challenges. Concrete 3D printing, on the other hand, is uniquely suited to implement nature-inspired principles, offering the potential to enhance structural efficiency, reduce material usage, and expand the variety of possible constructions [17].

One example of a nature-inspired pattern suitable for 3D-printed lattice beams is the Voronoi pattern, a tessellation often found in nature (e.g., giraffe skin, dragonfly wings, soap bubbles). Previous research has explored the use of Voronoi patterns to optimize the tensile strength of porous structures through genetic algorithms, suggesting potential applications in structural materials [18]. Some parametric studies on 2D rectangular beams with Voronoi tessellations have examined the effects of cell sizes, uniformity, and support conditions on load-carrying capacity [19].

However, the application of Voronoi patterns in concrete 3D printing has not yet been explored. While fabricating Voronoi lattice beams using traditional construction methods would present challenges, concrete 3D printing offers a more feasible and promising approach. Traditionally, Truss patterns have been favoured for their efficiency in load distribution and ease of mass production due to their modularity, whereas the potential of Voronoi patterns remains untapped, mainly in this context.

2 Objectives

The primary objective of this study is to explore the material savings potential of shape-optimized lattice beams featuring Voronoi cell patterns and compare them against the conventional Truss pattern and a regular solid beam for both symmetrical and unsymmetrical loading scenarios.

3 Methodology

The potential for material savings is assessed by comparing various geometric configurations of lattice beams, each individually optimized for shape. 3D-printable lattice beams with different infill patterns are created using a parametric design system and optimized via a genetic algorithm. The system adjusts the geometric parameters within defined limits, and each design is analyzed using finite element software. The sizes of the beam members are determined based on the forces from the analysis, ensuring that all relevant constraints are met. Subsequent sections provide further details on the methodology adopted.

3.1 Geometric Modelling

The study compares beams with two distinct lattice patterns—Truss and Voronoi—against a conventional rectangular beam for both symmetrical and unsymmetrical loading scenarios to assess material efficiency under similar constraints. The beams are modelled as assemblies of line elements representing the edges of the lattice structures. Each lattice type includes multiple variants generated through a combination of parameters. Adopting a parametric modelling approach eliminates the need for manual creation of each variant, enabling efficient exploration of a large design space. The parametric model is implemented using Rhinoceros3D, Grasshopper3D, and Python.

A 6 m span beam is used as a case study to perform the comparative evaluation, with the length-to-depth (l/d) ratio varied between 2 and 20. In addition to the self-weight of the beam, for the symmetrical loading condition, a uniformly distributed load (UDL) of 5 kN/m is applied on the beam's top surface (Figure 1), while for the unsymmetrical loading condition, a point load of 10 kN at 1.5 m from the left support along with a UDL of 5 kN/m running across the full span is applied (Figure 2). The support conditions taken for the study are fixity at both ends. The beam's bottom profile can take either a convex or concave shape, depending on "arc – midpoint offset" parameter. The beam bottom is modelled as a three-point arc with the arc-midpoint offset defining the

offset of the midpoint on the arc. The minimum offset is restricted to 10 cm from the midpoint of the beam to facilitate printability. The first and last points of the three-point arc are dependent on the l/d ratio adopted. The depth of the beam at the center changes depending on the offset position of the arc's midpoint, which can range from 0 to 7 meters in steps of 0.1 meters. The thickness and width of all elements are kept constant for a given configuration, where an element represents the connectivity between two nodes. The element dimensions are calculated within specified ranges of the respective parameters, provided in Table 1 and Table 2 for the two lattice beam configurations, ensuring the internal forces remain within specified stress limits. To facilitate thorough design exploration, the starting and ending values of each parameter are deliberately set far apart. These ranges are later verified to confirm that they do not constrain the optimization process.

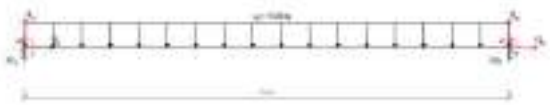


Figure 1. Symmetrical Loading Scenario

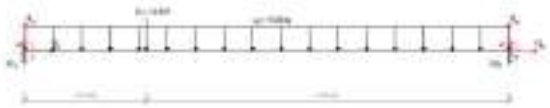


Figure 2. Unsymmetrical Loading Scenario

Table 1. Truss Beam

Sl no	Input parameters	Start Value	End Value	Total Possible Values
1	l/d (ratio)	2	20	19
2	Arc - midpoint offset (m)	-7.0	0.0	71
3	Truss Divisions	5	50	46
4	Thickness (cm)	1	20	20
5	Width (cm)	10	100	91

Table 2.Voronoi Beam

Sl no	Input parameters	Start Value	End Value	Total Possible Values
1	l/d (ratio)	2	20	19
2	Arc - midpoint offset (m)	-7.0	0.0	71
3	Voronoi Count	12	100	89
4	Voronoi Seed	0	1000	1001
5	Thickness (cm)	1	20	20
6	Width (cm)	10	100	91

3.2 Numerical Study

3.2.1 Optimization

Due to the impracticality of manually exploring all possible parameter combinations, a genetic algorithm was employed to optimize the beam configurations for each case, enabling a more efficient identification of the optimal design. Wallacei software, an evolutionary engine for Grasshopper3D, was used for this purpose [20]. Wallacei utilizes the NSGA-2 algorithm [21] as its primary evolutionary algorithm and applies the K-means clustering method. To determine suitable values for population size and the number of generations, multiple trials were conducted with varying parameters, and combinations leading to convergence were identified.

Practical constraints, such as the minimum and maximum lengths of Voronoi edges and cell sizes, were accounted for by defining a range for Voronoi numbers. These constraints were enforced using a penalty function that combines step and ramp components (Figure 3). Penalty functions assign penalty values to solutions that violate constraints, with the penalty magnitude proportional to the degree of violation. The ramp function helps steer the optimization process by gradually reducing constraint violations, while the step function sharply differentiates the feasible design space by assigning high fitness values to feasible regions and penalizing boundary regions with violations. This approach discourages solutions with significant constraint violations, ultimately guiding the optimization toward feasible solutions. Over time, this ensures convergence to designs that fully satisfy all constraints.

To prevent premature convergence to local optima, the genetic algorithm maintains variation within the population. This diversity is introduced using random seed points, with the Voronoi beam utilizing up to 1000 seed points. A higher number of seed points broadens the range of potential configurations, enhancing spatial adaptation under the applied loading conditions. However, increasing the seed points also significantly raises computational power requirements.

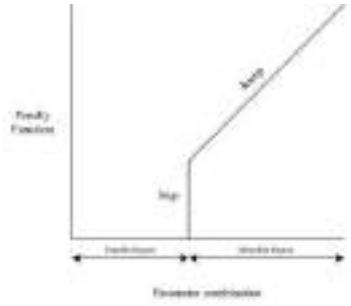


Figure 3. Step and Ramp Function

3.2.2 Implementation Details

The Voronoi pattern was generated using the Voronoi module in Grasshopper3D, based on specified values for the parameters Voronoi seed and Voronoi count. The generated Voronoi was trimmed to fit within the beam's defined boundaries. All parametrically generated line elements were converted into 2D beam elements for structural analysis. This analysis was carried out using Karamba3D [22], a structural analysis plugin for Grasshopper that integrates seamlessly with the parametric modelling approach adopted in the study.

The fixity conditions at the beam's supports were enforced by applying constraints to the nodal points at both ends. A custom Python script was developed to identify the node numbers at these locations and impose the necessary boundary conditions whenever the parametric model was updated during the genetic algorithm's execution.

The beam elements were assigned Material C30/37, compliant with the EN 206-1:2000 standard [23], as its compressive strength closely matched the concrete print mix used in the study. Strength-based constraints were applied, with a maximum allowable tensile stress of 2 MPa and compressive stress of 20 MPa. These constraints guided the optimization process, as no reinforcement is incorporated in this study.

The genetic algorithm varied the cross-sectional dimensions and orientations of the beam elements to ensure compliance with the stress limits. The primary objective was to determine the most efficient configuration for each lattice type by minimizing weight while satisfying criteria for allowable tensile stress, compressive stress, and displacement.

The large design space, resulting from numerous parameters and their wide range of values, facilitated an extensive and systematic search for optimal solutions.

3.2.3 Optimization Results

The masses of beam configurations for the same span were compared to assess potential material savings. Table 3 and Table 4 present the overall mass and percentage material savings relative to the conventional rectangular beam for each configuration for the symmetrical and unsymmetrical loading scenarios respectively.

Table 3. Overall mass and material savings (%) for symmetrical loading

Sl no	Beam Geometry	Mass (kg)	Material Savings (%)
1	Regular Beam	1350	-
2	Truss Beam	875	35.19
3	Voronoi Beam	761	43.64

Table 4. Overall mass and material savings (%) for unsymmetrical loading

Sl. no	Beam Geometry	Mass (kg)	Material Savings (%)
1	Regular Beam	1892	-
2	Truss Beam	1182	37.53
3	Voronoi Beam	833	55.95

The optimized variants of the Truss and Voronoi configurations are detailed in Table 5 and Table 6, respectively. Figure 4 and Figure 5 illustrate the Voronoi and Truss configurations for symmetrical loading case respectively and Figure 6 and Figure 7 represent the same for unsymmetrical loading respectively.



Figure 4. Optimized Voronoi Configuration for Symmetrical Loading



Figure 5. Optimized Truss Configuration for Symmetrical Loading



Figure 6. Optimized Voronoi Configuration for Unsymmetrical Loading



Figure 7. Optimized Truss Configuration for Unsymmetrical Loading

Table 5. Selected variants of Truss beam configuration

Sl no	Input parameters	Symmetrical Loading	Unsymmetrical Loading
1	l/d (ratio)	5	4
2	Arc - midpoint offset (m)	0	0
3	Truss Divisions	13	16
4	Thickness (cm)	12	7
5	Width (cm)	10	19
6	Mass (kg)	875	1182
7	Displacement (cm)	0.0468	0.0470
8	Maximum Tensile Stress (MPa)	0.192	0.198
9	Maximum Compressive Stress (MPa)	0.319	0.392

Table 6. Selected variants of Voronoi beam configuration

Sl no	Input parameters	Symmetrical Loading	Unsymmetrical Loading
1	l/d (ratio)	4	6
2	Arc - midpoint offset	0	0
3	Voronoi Count	25	18
4	Voronoi Seed	227	885
4	Thickness (cm)	13	19
5	Width (cm)	11	10
6	Mass (kg)	761	833
7	Displacement (cm)	0.0407	0.0498
8	Maximum Tensile Stress (MPa)	0.199	0.190
9	Maximum Compressive Stress (MPa)	0.267	0.495

After optimization, the total mass for the Truss and Voronoi configurations in the symmetrical load case was 875 kg and 761 kg, respectively. This resulted in material savings of 35.19% and 43.64% compared to the regular solid beam. As for the unsymmetrical load case, the total mass for the Truss and Voronoi configurations was 1182 kg and 833 kg, respectively. These values correspond to material savings of 37.53% and 55.95% compared to the regular solid beam.

3.3 3D Concrete Printing

To illustrate the printability of Voronoi patterned lattices, a small-scale Voronoi patterned beam is printed. The mix design used for printing the Voronoi beam incorporates a 45 % partial replacement of cement with a blend of limestone and calcined clay [9], and the corresponding mix design is as detailed in Table 7. Superplasticizers and viscosity-modifying admixtures were added at dosages of 0.6% and 0.1% of the binder, respectively, to achieve a printable consistency. This mix had a flow value of 180 mm, measured using the mortar flow table test with 25 drops, and retained 90% of its workability for 60–90 minutes (open time). The mix achieved a density of 2184 kg/m³ and a compressive strength of 30 MPa. The Voronoi lattice pattern generated was converted to G-code and fed to the 3D printer. The mix was fluid enough to bond well at the joints. Figure 8 shows the small-scale printed Voronoi beam.

Table 7. Mix Design

Material	Quantity (kg/m^3)
Quartz Sand – Grade B, 30-80 Mesh	1139
Ordinary Portland Cement (OPC) – 53 Grade	418
LC2 (Limestone + Calcined Clay in 1:2 ratio)	342
Water	304



Figure 8. 3D printed Voronoi Lattice Beam

4 Discussion

The optimization results reveal that the lattice patterns adopted an arch-shaped bottom soffit for both load cases, which helps improve load redistribution and reduce tensile stress in the concrete. Among the patterns analyzed, the Voronoi configuration shows a higher potential for material savings in both load cases, with much more saving potential for the unsymmetrical loading case. This superior performance can be attributed to the unstructured nature of Voronoi patterns. This illustrates the pattern's flexibility in efficiently distributing material where it is most needed and adapting its geometry to the specific characteristics of the design problem.

Of the six optimization variables in the Voronoi beam, the Voronoi seed variable is inherently random. This randomness introduces a degree of disorder along that axis, which limits the genetic algorithm's ability to optimize the seed value globally. However, the algorithm effectively determines the optimal combination of the remaining variables for a fixed seed value. Thus, in the case of Voronoi beams, increasing the number of iterations can help the algorithm uncover even better solutions, although this improvement comes at the cost of additional computational time.

This study thus adopted a parametric-driven approach to shape and topology optimization to develop efficient beam geometries based on load and support conditions. The approach presented is scalable and can be readily applied to larger spans and more complex loading conditions.

5 Conclusion

This study explored the potential of nature-inspired Voronoi patterns for shape optimization in 3D-printed concrete beams. It compares two modular configurations—Truss and Voronoi patterns—against a standard solid beam for both symmetrical and unsymmetrical loading conditions. The key findings include:

- For the symmetrical loading case of the 6m span beam, the percentage savings in the material by the Truss and Voronoi configurations were 35.19% and 43.64% respectively.
- As for the unsymmetrical loading case, the material savings for the Truss and Voronoi configurations were 37.53% and 55.95% respectively.
- The superior performance of the Voronoi configuration for the unsymmetrical loading case is attributed to its non-uniform, unstructured nature-inspired design, which enables optimal material distribution and geometric orientation tailored to specific design requirements.
- This research provides proof of concept that Concrete 3D Printing can successfully fabricate complex Voronoi patterns—challenging to achieve with traditional formwork—while significantly minimizing material usage.

This study highlights how nature-inspired patterns, made feasible through additive manufacturing methods like Concrete 3D Printing, can advance sustainability goals and address the evolving demands of modern architecture. Due to the limitation of testing, we have conducted only numerical study which will be a limitation; however, it is assumed that the theoretical model will be applicable for structures of any scale. Future research will focus primarily on strategies for optimal reinforcement to further enhance load-bearing performance and incorporate ductile failure mechanisms.

6 Acknowledgements

This project is funded by the Science and Engineering Research Board (SERB), India, through the Grant IMP/2018/000224. The Institute of Eminence Research Initiative Project on Technologies for Low-Carbon Lean-Construction (TLC2) and SBI Foundation also provided financial support for this research. Tvasta Manufacturing Solutions Pvt. Ltd., India is thanked by the authors for their support in providing 3D printers for the study.

7 References

[1] B. Sizirici, Y. Fseha, C. S. Cho, I. Yildiz, and Y. J. Byon, "A Review of Carbon Footprint Reduction in Construction Industry, from Design to Operation," Materials 2021, Vol. 14, Page 6094, vol. 14, no. 20, p. 6094, Oct. 2021, doi: 10.3390/MA14206094.

[2] R. Chaudhury, U. Sharma, P. Thapliyal, and L. Singh, "Available online 4," J Clean Prod, vol. 417, pp. 959–6526, 2023, doi: 10.1016/j.jclepro.2023.137466.

[3] T. Wangler, N. Roussel, F. P. Bos, T. A. M. Salet, and R. J. Flatt, "Digital Concrete: A Review," Cem Concr Res, vol. 123, p. 105780, Sep. 2019, doi: 10.1016/J.CEMCONRES.2019.105780.

[4] Z. Huszár and É. Lublóy, "Examination of the cost ratio of the formwork," Acta Technica Jaurinensis, vol. 14, no. 2, pp. 155–170, 2021, doi: 10.14513/actatechjaur.00598.

[5] J. Chung, G. Lee, and J.-H. Kim, "Framework for technical specifications of 3D concrete printers," Autom Constr, vol. 127, p. 103732, 2021, doi: https://doi.org/10.1016/j.autcon.2021.103732.

[6] F. Lyu, D. Zhao, X. Hou, L. Sun, and Q. Zhang, "Overview of the Development of 3D-Printing

Concrete: A Review," Applied Sciences, vol. 11, no. 21, 2021, doi: 10.3390/app11219822.

[7] A. V Rahul, M. K. Mohan, G. De Schutter, and K. Van Tittelboom, "3D printable concrete with natural and recycled coarse aggregates: Rheological, mechanical and shrinkage behaviour," Cem Concr Compos, vol. 125, p. 104311, 2022, doi: https://doi.org/10.1016/j.cemconcomp.2021.104311.

[8] J. Zhang, J. Wang, S. Dong, X. Yu, and B. Han, "A review of the current progress and application of 3D printed concrete," Compos Part A Appl Sci Manuf, vol. 125, p. 105533, Oct. 2019, doi: 10.1016/J.COMPOSITESA.2019.105533.

[9] B. Shantanu, J. Smrati, and S. Manu, "Criticality of binder-aggregate interaction for buildability of 3D printed concrete containing limestone calcined clay," Cem Concr Compos, vol. 136, p. 104853, Feb. 2023, doi: 10.1016/J.CEMCONCOMP.2022.104853.

[10] B. Raphael, S. Senthilnathan, A. Patel, and S. Bhat, "A review of concrete 3D printed structural members," Front Built Environ, vol. 8, p. 1034020, Jan. 2023, doi: 10.3389/FBUIL.2022.1034020/BIBTEX.

[11] H.-W. Cho, S.-G. Roh, Y.-M. Byun, and K.-S. Yom, "Song-Geun Roh; Young-Min Byun; Kyong-Soo Yom ctbuh.org/papers 662 CTBUH," 2004.

[12] J. J. Assaad, A. A. Yassin, F. Alsakka, and F. Hamzeh, "A Modular Approach for Steel Reinforcing of 3D Printed Concrete—Preliminary Study," Sustainability 2020, Vol. 12, Page 4062, vol. 12, no. 10, p. 4062, May 2020, doi: 10.3390/SU12104062.

[13] S. Maitenaz, R. Mesnil, P. Onfroy, N. Metge, and J. F. Caron, "Sustainable Reinforced Concrete Beams: Mechanical Optimisation and 3D-Printed Formwork," RILEM Bookseries, vol. 28, pp. 1164–1173, 2020, doi: 10.1007/978-3-030-49916-7_110/FIGURES/8.

[14] D. Dey, V. N. Van, H. N. Xuan, D. Srinivas, B. Panda, and P. Tran, "Flexural performance of 3D printed concrete structure with lattice infills," Developments in the Built Environment, vol. 16, p. 100297, Dec. 2023, doi: 10.1016/J.DIBE.2023.100297.

[15] L. Breseghello, H. Hajikarimian, H. B. Jørgensen, and R. Naboni, "3DLightBeam+. Design, simulation, and testing of carbon-efficient reinforced 3D concrete printed beams," Eng Struct, vol. 292, p. 116511, Oct. 2023, doi: 10.1016/J.ENGSTRUCT.2023.116511.

[16] N. S. Ramzy, "Beyond Sustainability, Design for Well-Being: Gaudí's Monument to Nature, Biomimetic Functions with Biophilic Morphology," Journal of Architectural Engineering, vol. 28, no. 1, p. 05021016, Mar. 2022, doi: 10.1061/(ASCE)AE.1943-5568.0000515/ASSET/DAB279F9-EFE2-4656-ABB1-3254D2EB97D0/ASSETS/IMAGES/LARGE/FIGURE18.JPG.

[17] A. Du Plessis, A. J. Babafemi, S. Chandra Paul, B. Panda, J. P. Tran, and C. Broeckhoven, "Biomimicry for 3D concrete printing: A review and perspective," 2020, doi: 10.1016/j.addma.2020.101823.

[18] C.-C. Tung, Y.-Y. Lai, Y.-Z. Chen, C.-C. Lin, and P.-Y. Chen, "Optimization of mechanical properties of bio-inspired Voronoi structures by genetic algorithm," 2023, doi: 10.1016/j.jmrt.2023.08.210.

[19] S. Sehat and A. Khatami, "Parametric study on application of Voronoi patterns in design of 2-D beams," Advances in Structural Engineering, vol. 24, no. 11, pp. 2375–2389, Aug. 2021, doi: 10.1177/1369433221999740/FORMAT/EPUB.

[20] "Evolutionary Engine for Grasshopper3D | Wallacei." Accessed: Feb. 02, 2024. [Online]. Available: https://www.wallacei.com/

[21] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II," IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, vol. 6, no. 2, 2002.

[22] "Karamba3D." Accessed: Jun. 27, 2024. [Online]. Available: https://karamba3d.com/

[23] "EN 206-1:2000 - Concrete - Part 1: Specification, performance, production and conformity." Accessed: Jan. 17, 2024. [Online]. Available: https://standards.iteh.ai/catalog/standards/cen/af40b93d-23eb-43ae-8880-d6b6c9088325/en-206-1-2000

Design and Development of a Remote User Interface for Multi-Robot On-site Component Inspection

Juan Carlos Cruz Rivera[1], Mohsen Navazani[2], Ibukun Awolusi, Ph.D.[3], Ao Du, Ph.D.[4], and Jiannan Cai, Ph.D.(corresponding author)[5]

[1]Department of Mechanical Engineering, The University of Texas at San Antonio, United States
[2]School of Civil Environmental Engineering, and Construction Management, The University of Texas at San Antonio, United States
[3]School of Civil Environmental Engineering, and Construction Management, The University of Texas at San Antonio, United States
[4]School of Civil Environmental Engineering, and Construction Management, The University of Texas at San Antonio, United States
[5]School of Civil Environmental Engineering, and Construction Management, The University of Texas at San Antonio, United States

juan.cruzrivera@utsa.edu, mohsen.navazani@utsa.edu, ibukun.awolusi.utsa.edu, ao.du@utsa.edu, jiannan.cai@utsa.edu

Abstract -

This paper presents the design and development of an open-source browser-based interface for coordinating a multi-robot system for on-site construction inspections. The developed system enables remote human operators to effectively oversee and direct automated inspection tasks using both aerial and ground-based robotic platforms. The interface incorporates gaze-tracking technology for intuitive target selection and implements a containerized microservice architecture to ensure reliable deployment across different operating environments. The complete system architecture, implementation methodology, and a real-world inspection scenario are presented. The application has been released as open-source software to encourage adoption and further development by the construction automation community.

Keywords -

Human-machine interface; Construction inspection; Teleoperation; Human-robot interaction

1 Introduction

Recent trends in infrastructure construction show a rising usage of off-site construction of components [1]. While off-site manufacturing facilities maintain rigorous quality control, the components can still suffer damage during transportation and on-site storage, requiring additional on-site inspection to ensure structural integrity before final assembly. On-site inspections, however, can be time-consuming, labor-intensive, and may not be as accurate as desired due to human error and limitations in inspection equipment [2].

Automated inspection systems have the potential to address these challenges by providing consistent, accurate, and efficient inspection capabilities. Robotic systems can be used to conduct the inspections in an efficient and automated manner [3]. Moreover, Artificial Intelligence (AI) enabled systems can leverage machine learning algorithms to better identify defects and anomalies [4], as well as navigate the challenging environments of construction sites [5].

In this paper, we present a human-machine interface for remote multi-robot on-site component inspection which capitalizes on the ongoing advancements in robotics and AI. Our application allows for the integration of rapidly acquired contextual data by one robot (e.g., drone) into an inspection mission for another robot (e.g., ground robot), all while allowing a human operator to monitor and direct the inspection process and provide additional context as needed.

We have also designed the application to allow for gaze-enabled inputs, providing an intuitive method for an operator to communicate intention and interest in a particular object.

In the following sections, we will provide further context on the use of robots in construction and infrastructure inspection, the need for human operators in automated systems, the use of gaze tracking for human input, and how the developed application fits into this context. We will then describe the design and development of the application, as well as an example use case. Performance evaluations and real-world case studies are planned for later work.

The interface's software has been released as open-source software under the MIT license on GitHub: `https://github.com/AIConLab/gaze_based_remote_user_interface`.

Our contributions to this paper are as follows:

- Development of a novel containerized architecture specifically designed for multi-robot construction inspection that ensures flexibility and portability.

- Integration of gaze-tracking as an intuitive human input method for remote inspections, which addresses challenges in traditional input methods on construction sites.
- Creation of a platform-agnostic interface human interface for coordinating aerial and ground robot inspection.
- Release of a complete open-source implementation that can serve as a foundation for further research and development in construction automation.

2 Background

GPS-based navigation systems have been a staple in outdoor robotic applications, allowing for precise navigation based upon a common coordinate system. Importantly, the common coordinate system allows the absolute positioning of a robot or multiple robots [6].

As reported in [7], ground robots have been more popular in the past for general civil infrastructure tasks, being able to do both contact and non-contact tasks, while aerial robots have been more popular for contactless tasks, such as visual inspection tasks of bridges, buildings, and other structures. Aerial robots, however, face fundamental challenges when it comes to extended usage and large payloads, limiting the amount of sensing and inspection equipment that can be carried and the time that the drone can be in the air [8].

Ground-based mobile robots, in contrast, can carry larger sensor payloads and operate for extended periods of time [9], but lack the mobility of aerial robots.

Combined approaches using both aerial and ground robots provide a more promising solution for infrastructure inspection tasks as these approaches can take advantage of the strengths of both systems as shown in [10].

In both fully autonomous and human-in-the-loop systems, human decision-making is still required at some stage of the system process, such as for providing the robots with an initial plan or additional context at the time of inspection.

Even in AI-enabled systems there exist limitations in their ability to generalize to new and unseen scenarios. In such cases, providing large amounts of training data could improve the performance of a model for a specific scenario, but this is not always possible in real-world applications as high-quality data can be difficult to collect [11]. Thus, human operators are often needed to provide additional context and guidance to the system to ensure that the inspection is conducted correctly and efficiently.

To provide human input into the system, gaze tracking has been shown to be one promising solution [12]. The incorporation of gaze tracking has the benefit of allowing the operator to provide input without the need for physical interaction, which can be important in situations where the operator is wearing gloves or is in a hazardous environment.

A user interface or visualization tool then allows the human user to manage how their input is processed and how the robot(s) respond to the input. One such visualization tool is Rviz [13], an open-source software package that allows the user to see navigation plans, robot transform frames, and more. Mission planning plugins for Rviz such as [14], create more sophisticated planning interfaces within Rviz allowing the human operator to view and plan for multi-robot systems.

Notably, these tools are designed to be used within the Robot Operating System (ROS) ecosystem, which is a popular framework for developing robotic applications. Our application differs in that it is designed to provide a standalone, browser-based interface that has the ability to communicate with ROS-based systems - but importantly does not require them to run.

An ongoing challenge in modern robotics research has been software replicability [15]. Our application provides a working example of one approach to providing software replicability through the use of Docker containers.

Docker is a platform for creating containerized applications. Containers contain everything needed to run a specific application, are lightweight, and are loosely isolated from the host machine allowing a developer to access host files, hardware, and resources when needed. A container will run an image instance, which is a snapshot of a certain environment, for example a specific version of Ubuntu with a specific version of Python. Docker Compose provides a tool for orchestrating multiple containers. A Docker Compose file and Docker image are extremely portable text-based files that allow an application to be built and run on almost any host machine.

3 Methodology

This section provides a high-level overview of the application's architecture, data handling, and software module descriptions. In the Implementation section, we will provide an example implementation of the application based on trials done with real-world robots. The application, however, is designed to be platform agnostic and can be used with any robot platform that provides the required input data and can receive the output data. These inputs and outputs are further described in the following section.

3.1 System Overview

The developed application allows for the input of GPS-based waypoint coordinates of structural inspection targets, along with the reference images of those targets. Video streams from a variety of sources can be passed

to the application for viewing by the user. Gaze and fixation data is also streamed to the application to allow for the gaze-based segmentation mask generation output when remotely inspecting a structure. The application outputs are directed at the ground robot which will be performing the remote inspection. The outputs include formatted Robot Operating System (ROS) 1 messages including mission commands, target reference images, the GPS waypoints, and the structure segmentation mask. Handling and formatting of the input and output data will be further discussed in the following sections. Figure 1 provides a visualization of the system's inputs and outputs, where the inputs are the incoming arrows and the outputs are the outgoing ones.



Figure 1. System inputs and outputs.

To allow for wide compatibility and guarantee software functionality across different systems, we containerized the application with Docker and separate module concerns to different Docker compose services. By separating services and images used, we have the flexibility to configure the software environment as needed to run specific packages. Figure 2 visualizes the described architecture, where the *App* contains the services defined in our Docker Compose file - namely *Mission Manager*, *Web Host*, *Segmentation*, *Module Processor*, and *Pupil Dev* services. Each service then contains modules that run certain functionality, for example, the *Frontend* and *Backend* modules within the *Web Host* service provide the live updating and rendering of the browser-based user interface. Arrows between each service represent data that is transferred asynchronously between containers. The modules within the application are further explained in the following sections.

3.2 Data Handling

Our architecture for the asynchronous messaging is inspired by that of ROS 1 [16]. Inter-container and inter-module message passing is achieved through an asynchronous ZeroMQ (ZMQ) publish-subscribe system using the ZMQ proxy pattern, wherein a central message broker facilitates the distribution of published messages to subscribers. The message broker acts as an intermediary that decouples publishers from subscribers, allowing for flexible and scalable communication between modules regardless of whether they are in the same container or not. For serialization and deserialization of messages, we use the MessagePack protocol.

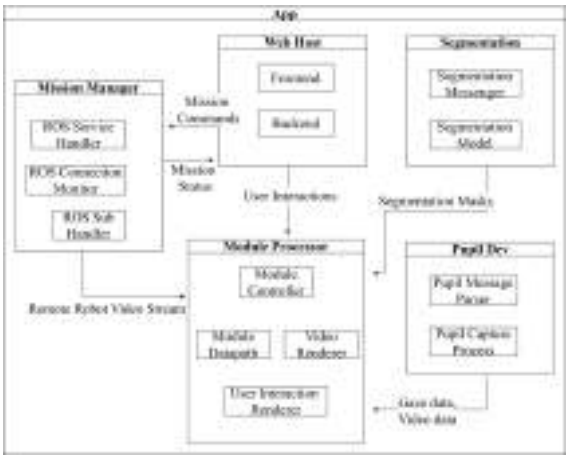


Figure 2. Docker compose services and module data flow within the application. The containers represent the services, and the modules are represented by the contained rectangles.

3.3 Module Descriptions

3.3.1 Mission Manager

The GPS locations of targets to be inspected, along with their respective reference images, are inputted into the application through its local file system. Once loaded, the files are processed within the Mission Manager service which provides modules for interacting with the remote ROS robot. The Mission Manager service also contains modules for connection to a remote ROS Master URI, subscribing to published ROS data from the robot, and publishing data to the robot. The receiving robot and Mission Manager service require ROS service files for structuring the mission command formatting, remote file transfer, and mission state updates. These are included within the applications repository.

3.3.2 Module Processor

Most other data is routed through the Module Processor service which centralizes multi-container data flow. In particular, the Module Datapath serves as the central subscriber for video stream inputs and gaze data. Once data is received, it publishes the data to any subscribed modules.

Video topics from a remote robot or from a local camera would be published to the Video Render module for frame-by-frame processing depending on the application's processing state and the current video topic selection. The application processing state is managed by the Module Controller module, which receives commands generated by user interactions (button clicks, video topic selection,

etc.) from the Frontend. Based on the application's processing state, the Frontend will either render a live stream of a selected video topic via the Video Render module or provide a still image of the latest frame and rendered fixation point via the User Interaction Render module.

3.3.3 Web Host

The Web Host service uses quart – a python based asynchronous web microservice based on Flask. The Frontend module handles the User Interface (UI) layout and user input capture from the web app, while the Backend module handles the data publishing and subscribing between different services.

3.3.4 Pupil Dev

Gaze data refers to normal gaze detection from eye activity, while fixations refer to gaze in a localized region for a certain period of time. We use the fixation data as a point trigger for identifying an operator's point of interest from a screen. Gaze data and gaze-related video stream are handled in the Pupil Development service, in this case, the service is named specifically for the hardware used to capture the gaze data. This is discussed further in the Implementation section.

3.3.5 Segmentation

The Segmentation services host the segmentation model and messaging modules. It is used to generate segmentation masks over images when a fixation point prompt is passed. The generated mask can then be used by the ground robot for target localization and specific area structural inspections.

3.4 User Interface

The User Interface (UI) is the primary means of interaction with the application. Figure 3 provides a skeleton of the interface design and the terms used to describe the regions in the UI. These regions include the *Mission Controls*, *Workspace*, and *Banner*.

The Workspace comprises the majority of the application display space, as it handles visualization of rendered video streams and display of fixation and segmentation outputs.

The Mission Controls section allows for the display of any mission state information, remote robot network status, and mission related utility functions.

The Banner offers state-driven button rendering based on the application's state as well as a persistent video topic selection drop down and a persistent button to toggle the gaze related functionality. The three states considered in this design are shown in Figure 4, where *Normal* is the normal operating state, *Fixation Captured* is the state when the Pupil Capture application detects a fixation on the screen, and the *Segmentation Results* state which activates once the segmentation masks have been generated, allowing a user to select which mask to send to robot during inspection time.

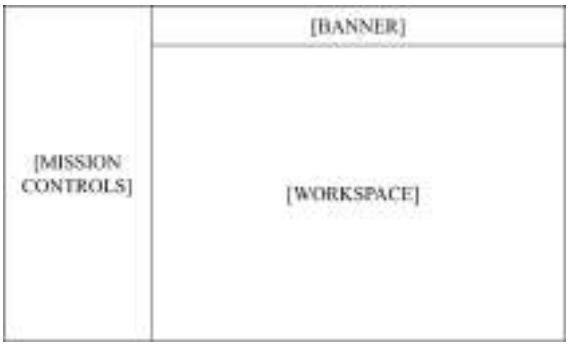


Figure 3. Skeleton of the UI layout.

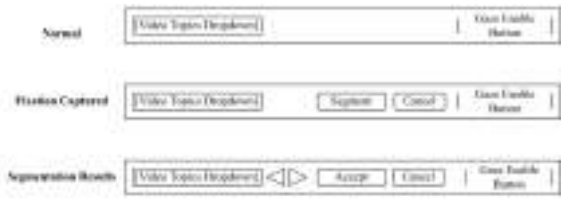


Figure 4. UI components display.

4 Implementation

As of the writing of this paper, the latest release of the application is V1.0.1. In the release, we have designed the application's implementation to function for the system shown in Figure 5. The arrows in Figure 5 describe the data flow between hardware as mentioned in Figure 1. The Ground Robot, Drone, and Gaze tracking hardware used in this test are as follows:

- Drone: DJI Matrice 30T
- Ground Robot: Clearpath Husky A200
- Gaze Tracking Hardware: Pupil Labs Core eye-tracking glasses

The User Computer block in Figure 5 defines the computer hardware used by the Remote Operator during testing. In this test, a Lenovo ThinkPad T490s with an Intel Core i7-8565U quad-core processor and 8 GiB of RAM is used. The laptop is running Ubuntu 24.04.2 LTS as the Operating System.

Further details of the implementation scenario and module implementation are described in the following sections.

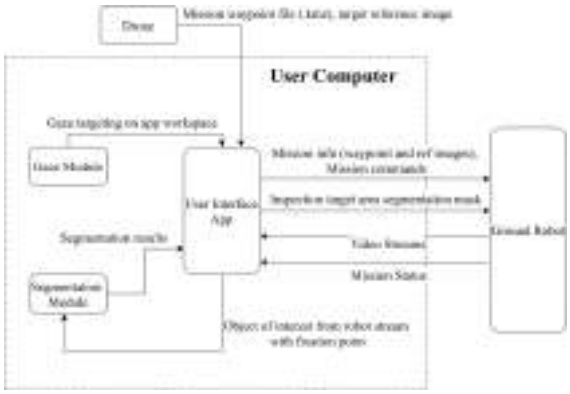


Figure 5. Overview of the system used in the implementation.

4.1 Task Scenario

In this inspection scenario, a remote operator is tasked with inspecting a mock bridge structure using a drone and a ground robot. We set up the inspection environment at the University of Texas at San Antonio's (UTSA) drone enclosure area, a 15000 square-foot and 60-foot-tall netted facility for drone testing. The wooden mock bridge was placed inside of the drone cage to act as our inspection object along with construction cones to add noise to the environment.

The remote operator is stationed outside of the cage with a laptop and the Pupil Labs Core eye-tracking glasses. An EAP225-Outdoor access point was set up to create a local wireless network between the ground robot and the operator's laptop. Images of this setup are shown in Figure 6. The entire inspection sequence is described in the following section.



Figure 6. (a) Outdoor testing environment showing the bridge structure and cones used for the scenario. (b) An image captured by the drone during the inspection mission. (c) The UGV and drone's starting position within the caged testing area.

4.2 Application Setup

Prior to starting the mission, the user must calibrate the Pupil Labs Core eye-tracking glasses and create a Surface in the Pupil Capture software [17]. The Surface is used to define the boundary of the gaze input area restricting it to a certain area which allows us to then extract exact pixel coordinates corresponding to that area. A Surface is defined by four Apriltags [18] which are rendered in the application's Workspace area and captured by the Pupil Core glasses' front-facing camera. These Apriltags are visualized in Figure 7(a)



Figure 7. (a) Apriltag rendered in the corner of the Workspace area of the application. (b) Teleoperation interface for manual UGV operation. (c) Mission control buttons and file viewer. (d) Mission and connection state.

In addition, the user must place a .env file in the .docker directory of the application. This .env file is used to define the IP address and hostname of the remote robot to connect to the ROS Master URI.

4.3 Mission Inspection Sequence

The operator must create a GPS-based waypoint of the bridge structure and capture a reference image using the drone. In our implementation, we use the DJI pilot app in order to save the waypoints as a KMZ file. The KMZ file contains a KML file from which we extract waypoint IDs and their respective latitude and longitude.

The reference images are stored under a directory named for their respective waypoint ID. The gathered drone data can then be placed in a directory within the application for formatting and processing.

When the operator selects "Process Mission Files" in the application, the Mission Manager service processes the GPS coordinates and reference images to prepare them for transmission to the UGV. The files are then wirelessly transferred over the local network when the "Mission Start" button is pressed. These buttons are shown in Figure 7 (c).

Upon receiving the files, the UGV converts the GPS coordinates into a local map point and creates a global and local navigation plan to autonomously navigate to the area. The operator can monitor the UGV's state (Figure 7 (d)) and video feed for any issues or anomalies while the robot navigates to the target location. If manual control of the robot is required at any point, the remote operator can teleoperate the ground robot's navigation base through the interface. The controls for this are shown in Figure 7 (b).

When the inspection target is reached, the operator directs the UGV to locate the inspection object of interest using the gaze-interface in the workspace. The UGV will wait until an inspection target area is specified by the remote operator.

The inspection target area is generated by the gaze-enabled input feature of the application. When the gaze is enabled, Apriltags are rendered in the workspace area of the application and overlaid on the video stream. The Apriltags create a bounded area to filter gaze/fixation points so that they correspond directly to pixel coordinates in the video stream frame. An example is shown in Figure 7 (a).

When a fixation is detected, the application stops rendering new video frames and displays the latest frame and respective fixation. The user can then accept the fixation as an input to the segmentation module or attempt the fixation process again.

An image and fixation point sent to the segmentation module are passed to the Segment Anything 2 model (SAM2) [19] which outputs three segmentation mask predictions. These are then displayed to the user in the application's Workspace area, where they can select a mask or reattempt the fixation process. An example of a segmentation mask is displayed in Figure 8. Here the mask is shown as an orange overlay over a pillar of the mock bridge the remote operator wishes to inspect. Within the mask, a red dot shows the specific point the user's fixation was captured at.



Figure 8. An example of a segmentation mask generated after fixation. The orange area depicts the target area, while the red dot depicts the fixation point.

With a target surface area mask received, the UGV could begin the required inspection and data-capturing process. When the inspection of the previous mask is complete the operator could provide another inspection target area or terminate the inspection and move on to the next waypoint. If no more waypoints were available, then the mission would be completed.

The entire mission sequence is visualized in Figure 9, showing a sequence chart with the Drone, Operator, and the UGV as actors. In the chart, the 3 phases of the implemented inspection process are the *Data Gathering Phase*, *Mission Execution Phase*, and *Inspection Phase*. The arrows show actions each actor takes from start to finish. Loops are shown as dashed rectangles that continue until an exit condition.

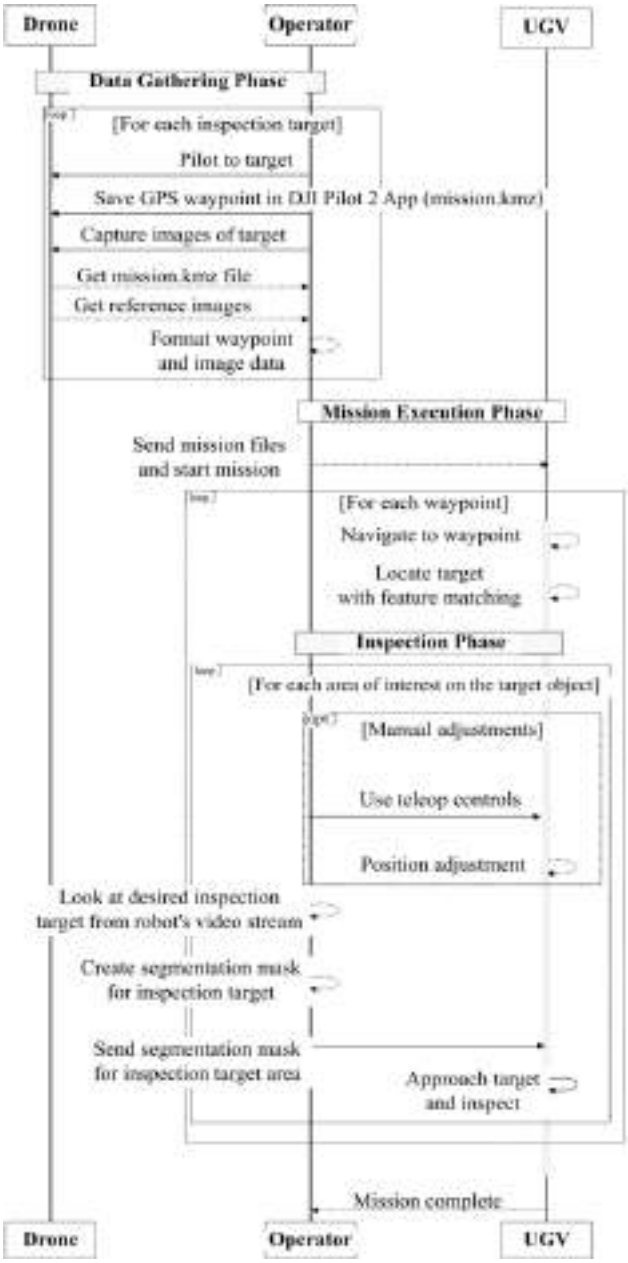


Figure 9. Mission Sequence

5 Limitations

The current approach assumes the drone can safely gather images and a GPS coordinate for the inspection

targets and that the ground robot can navigate to those targets. A mission may still be possible if image data can be captured of the objects, and the GPS coordinate for waypoints can be formatted to a KMZ file, but this approach requires additional effort by the operator and is not the intended approach for our implementation.

Image data captured by the first robot (the drone in our example) may lose relevancy over time if, for example, the images were captured at a different time of day or there was a significant scene change from when the initial images were captured.

6 Conclusions

On-site construction inspection is critical to ensure the quality and durability of infrastructure structures. On-site inspections, however, can be time-consuming, labor-intensive, and may not be as accurate as desired due to human error and limitations in inspection equipment. Automated inspection systems have the potential to address these challenges by providing consistent, accurate, and efficient inspection capabilities. Robotic systems can be used to conduct the inspections in an efficient and automated manner. However, due to the complexity of construction sites, human operators are often needed to provide additional context and guidance to the system to ensure that the inspection is conducted correctly and efficiently.

In this work, we provided an architecture overview and dataflow description of the developed human-machine interface application for multi-robot on-site component inspection. Our application allows for the integration of rapidly acquired contextual data by one robot (e.g., drone) into an inspection mission for another robot (e.g., ground robot), all while allowing a human operator to monitor and direct the inspection process and provide additional context as needed. An implementation of the application was demonstrated in a real-world scenario, showing the application's functionality and user interface design.

In future studies, we plan to run more inspection case studies to evaluate the application's usability and impact on productivity in more realistic inspection scenarios.

The current implementation serves as a functional base for use in novel inspection applications where the software can be easily modified to extend functionality and hardware compatibility. We encourage any interested parties to incorporate the applications into their projects.

7 Acknowledgement

This research was funded by the Transportation Infrastructure Precast Innovation Center (TRANS-IPIC), Tier 1 University Transportation Center (UTC) via Project UT-23- RP-02, and the U.S. National Science Foundation (NSF) via Grants 2138514 and 2222670. The authors gratefully acknowledge the support. Any opinions, findings, recommendations, and conclusions in this paper are those of the authors and do not necessarily reflect the views of TRANS-IPIC, NSF, and The University of Texas at San Antonio.

References

[1] R. H. Assaad, I. H. El-adaway, M. Hastak, and K. LaScola Needy. Quantification of the state of practice of offsite construction and related technologies: Current trends and future prospects. *Journal of Construction Engineering and Management*, 148 (7):04022055, 2022. doi:10.1061/(ASCE)CO.1943-7862.0002302.

[2] B. M. Phares, G. A. Washer, D. D. Rolander, B. A. Graybeal, and M. Moore. Routine highway bridge inspection condition documentation accuracy and reliability. *Journal of Bridge Engineering*, 9(4):403–413, 2004. doi:10.1061/(ASCE)1084-0702(2004)9:4(403). URL https://ascelibrary.org/doi/abs/10.1061/%28ASCE%291084-0702%282004%299%3A4%28403%29.

[3] J. Seo, L. Duque, and J. P. Wacker. Field application of uas-based bridge inspection. *Transportation Research Record*, 2672(12):72–81, 2018. doi:10.1177/0361198118780825.

[4] S. Lee, S. Kwon, M. Jeong, S. Hasan, and A. Kim. Automated on-site quality inspection and reporting technology for off-site construction(osc)-based precast concrete members. In *Proceedings of the 37th International Symposium on Automation and Robotics in Construction (ISARC)*, pages 1152–1159, Kitakyushu, Japan, October 2020. International Association for Automation and Robotics in Construction (IAARC). ISBN 978-952-94-3634-7. doi:10.22260/ISARC2020/0158.

[5] D. Shah, A. Sridhar, N. Dashora, K. Stachowicz, K. Black, N. Hirose, and S. Levine. Vint: A foundation model for visual navigation, 2023. URL https://arxiv.org/abs/2306.14846.

[6] X. Gao, J. Li, L. Fan, Q. Zhou, K. Yin, J. Wang, C. Song, L. Huang, and Z. Wang. Review of wheeled mobile robots' navigation problems and application prospects in agriculture. *IEEE Access*, 6:49248–49268, 2018. doi:10.1109/ACCESS.2018.2868848.

[7] X. Hu and R. H. Assaad. The use of unmanned ground vehicles (mobile robots) and unmanned

aerial vehicles (drones) in the civil infrastructure asset management sector: Applications, robotic platforms, sensors, and algorithms. *Expert Systems with Applications*, 232:120897, 2023. ISSN 0957-4174. doi:https://doi.org/10.1016/j.eswa.2023.120897. URL https://www.sciencedirect.com/science/article/pii/S0957417423013994.

[8] R. Hadidi, B. Asgari, S. Jijina, A. Amyette, N. Shoghi, and H. Kim. Quantifying the design-space tradeoffs in autonomous drones. In *Proceedings of the 26th ACM International Conference on Architectural Support for Programming Languages and Operating Systems*, ASPLOS '21, pages 661–673, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450383172. doi:10.1145/3445814.3446721. URL https://doi.org/10.1145/3445814.3446721.

[9] Q. Liu, Z. Li, S. Yuan, Y. Zhu, and X. Li. Review on vehicle detection technology for unmanned ground vehicles. *Sensors*, 21(4), 2021. ISSN 1424-8220. doi:10.3390/s21041354. URL https://www.mdpi.com/1424-8220/21/4/1354.

[10] S. Halder and K. Afsari. Robots in inspection and monitoring of buildings and infrastructure: A systematic review. *Applied Sciences*, 13(4), 2023. ISSN 2076-3417. doi:10.3390/app13042304. URL https://www.mdpi.com/2076-3417/13/4/2304.

[11] G. Fobiri, I. Musonda, and F. Muleya. Reality capture in construction project management: A review of opportunities and challenges. *Buildings*, 12(9), 2022. ISSN 2075-5309. doi:10.3390/buildings12091381. URL https://www.mdpi.com/2075-5309/12/9/1381.

[12] F. Nenna, D. Zanardi, and L. Gamberini. Enhanced interactivity in vr-based telerobotics: An eye-tracking investigation of human performance and workload. *International Journal of Human-Computer Studies*, 177:103079, 2023. ISSN 1071-5819. doi:https://doi.org/10.1016/j.ijhcs.2023.103079.

[13] H. R. Kam, S.-H. Lee, T. Park, and C.-H. Kim. Rviz: a toolkit for real domain data visualization. *Telecommunication Systems*, 60:337–345, 2015.

[14] R. Sakagami, S. G. Brunner, A. Dömel, A. Wedler, and F. Stulp. Rosmc: A high-level mission operation framework for heterogeneous robotic teams. pages 5473–5479, 2023.

[15] F. Bonsignorio and A. P. del Pobil. Toward replicable and measurable robotics research [from the guest editors], 2015.

[16] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng. Ros: an open-source robot operating system. In *ICRA workshop on open source software*, number 3.2, page 5, 2009.

[17] M. Kassner, W. Patera, and A. Bulling. Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Adjunct Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '14 Adjunct, pages 1151–1160, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-3047-3. doi:10.1145/2638728.2641695. URL http://doi.acm.org/10.1145/2638728.2641695.

[18] E. Olson. Apriltag: A robust and flexible visual fiducial system. In *2011 IEEE International Conference on Robotics and Automation*, pages 3400–3407, 2011. doi:10.1109/ICRA.2011.5979561.

[19] N. Ravi, V. Gabeur, Y.-T. Hu, R. Hu, C. Ryali, T. Ma, H. Khedr, R. Rädle, C. Rolland, L. Gustafson, E. Mintun, J. Pan, K. V. Alwala, N. Carion, C.-Y. Wu, R. Girshick, P. Dollár, and C. Feichtenhofer. Sam 2: Segment anything in images and videos, 2024. URL https://arxiv.org/abs/2408.00714.

SatAOI: Delimitating Area of Interest for Swing-Arm Troweling Robot for Construction

Jia-Rui Lin[1,2], Shaojie Zhou[1,2], Peng Pan[1,2,*], Ruijia Cai[3] and Gang Chen[4]

[1] Department of Civil Engineering, Tsinghua University, China
[2] Key Laboratory of Digital Construction and Digital Twin, Ministry of Housing and Urban-Rural Development, China
[3] Institute of Architecture Design and Research, Chinese Academy of Sciences, China
[4] China Construction Third Engineering Bureau Beijing Co., Ltd., China
panpeng@tsinghua.edu.cn

Abstract

In concrete troweling for building construction, robots can significantly reduce workload and improve automation level. However, as a primary task of coverage path planning (CPP) for troweling, delimitating area of interest (AOI) in complex scenes is still challenging, especially for swing-arm robots with more complex working modes. Thus, this research proposes an algorithm to delimitate AOI for swing-arm troweling robot (SatAOI algorithm). By analyzing characteristics of the robot and obstacle maps, mathematical models and collision principles are established. On this basis, SatAOI algorithm achieves AOI delimitation by global search and collision detection. Experiments on different obstacle maps indicate that AOI can be effectively delimitated in scenes under different complexity, and the algorithm can fully consider the connectivity of obstacle maps. This research serves as a foundation for CPP algorithm and full process simulation of swing-arm troweling robots.

Keywords –

Construction robotics; troweling robot; swing-arm robot; area of interest; path planning

1 Introduction

With the development of automation technology, the application of robots in the construction industry is becoming increasingly widespread [1]. On the one hand, construction robots can replace workers to complete detrimental or monotonic tasks. And on the other hand, robots equipped with high-precision sensors can improve productivity and quality [2], which makes construction robots an important part of achieving intelligent construction. Currently, common on-site construction robots are qualified to perform tasks such as bricklaying, wall painting, building components assembly, and concrete construction [6]. Take concrete floor troweling as an example, to ensure sufficient flatness, floor troweling requires plenty of workers to draw trowels over large surfaces multiple times, which is time-consuming, labour-intensive, and the quality dependent on worker's skills [9]. However, if using a troweling robot, the workload will be greatly reduced, and the surface quality will be easily controlled and consistently ensured [10].

However, the environment at construction sites is often variable and unstructured, posing significant challenges to the automation of construction robots [11]. In order to adapt to different work environments, robots have adopted different designs for high efficiency. Figure 1 shows two different forms of concrete troweling robots. Swing-arm robot is a commonly used concrete troweling robot, and the design of the swing arm enhances the robot's working range while maintaining a compact size. Compared to other types of robots, swing-arm robots exhibit greater flexibility in working modes, enabling them to adapt to a wider range of construction scene. But with an additional degree of freedom for swinging arm, the operation of them becomes more complex. These issues make it difficult to operate troweling robots automatically, and human-robot collaboration [14] or complex pre-defined paths [10] are usually adopted. In addition, the automation of agricultural robotics also faces these issues, but irregular field shapes are often considered as the main constraint of agricultural robots [15]. There usually few obstacles in farmyard but many in construction sites. It will be more challenging when there are many obstacles or holes in the environments. A large number of efforts are needed to improve the automation level of troweling robots.



(a) (b)

Figure 1. (a) Ground troweling robot of Bright Dream Robotics [12]. (b) Crawler troweling robot of FangShi Technology [13].

The essence of troweling work is the coverage path planning (CPP) problem within a certain scene, that is, to find a collision-free trajectory to ensure the robot fully covers an area of interest (AOI) [16]. The most classical method for CPP is cellular decomposition [16]. Firstly, AOI needs to be effectively delimitated and then decomposed into a series of cells. The final coverage path can be solved through simple planning within the cells and connections between the cells. Therefore, the delimitating AOI is a primary and critical part for this method [17], which will affect all subsequent processes. Furthermore, since the troweling and movement of the swing-arm robot are controlled by different components, the passability and workability between obstacles are complex, which poses a huge challenge to the delimitation of AOI.

In order to improve the troweling efficiency, this research focuses on how to delimitate AOI of swing-arm troweling robot (Figure 1 b). By analyzing the characteristic of robot, **AOI** delimitation algorithm for **S**wing-**A**rm **T**roweling robot (**SatAOI** algorithm, a wish for robots working on Sat.) is constructed for. The remainder of this paper is organized as follows. Section 2 introduces methodology of overall research. Section 3 builds the models of obstacle map and swing-arm robot. Section 4 proposes the SatAOI algorithm. Section 5 presents a case study. Section 6 summarizes the research and discusses possible future investigations.

2 Methodology

The objective of this research is to delimitate the AOI of a swing-arm troweling robot. To achieve this objective, it is necessary to fully analyze the working characteristics of the swing-arm robot and establish accurate collision principles. On this basis, AOI is determined based on proposed SatAOI algorithm. The overall methodology is shown in Figure 2.

The first step is to establish the mathematical models of the obstacle map and the troweling robot, of which the most critical part is the definition of collision principles for the swing-arm robot. Then, SatAOI algorithm, composed of collision detection algorithm and search algorithm, is built for AOI delimitation. Finally, a case study is conducted to validate the feasibility of the proposed method. And the AOI results are discussed and visualized for further applying in other downstream tasks such as coverage path planning.

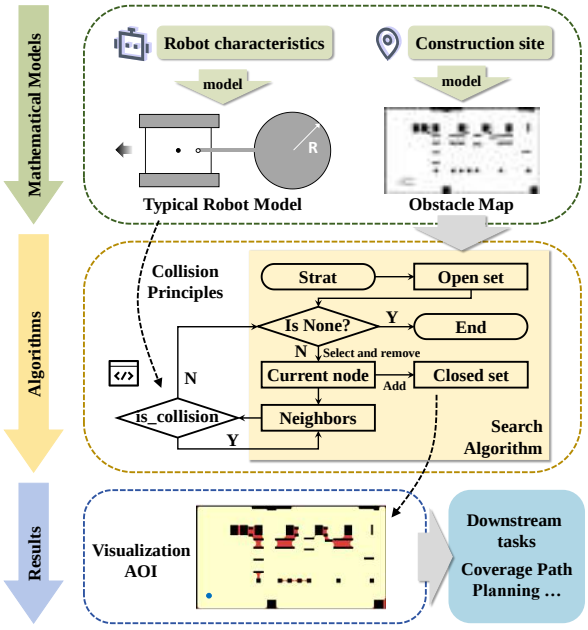


Figure 2. Overall Methodology

3 Model Definition

The challenge of delimitation AOI lies in the complexity of obstacle maps and troweling robots. Therefore, it is necessary to first simplify and define the mathematical model of these two research objects to ensure that the algorithm can accurately judge the robot's accessibility and workability according to certain collision principles.

3.1 Obstacle Map Model

In concrete troweling for building construction, obstacles refer to indoor building components that have already been constructed (such as walls, columns, etc.) or embedded parts and holes reserved on the ground for building function. This information is usually saved at CAD drawings or BIM models. But these structured data is not conducive for path planning. In continuous 2D environments, obstacle and map information is usually discretized into regular grids (such as equilateral polygons) for robot navigation [18]. These grids are easy to construct and clearly represent blocked and unblocked status. For convenience, squares with a side length of 10 mm are used to redraw the obstacle maps from CAD drawings in our research. As shown in Figure 3, the CAD

drawing is converted to a grid-based map. In grid-based map, the cells painted white indicate that areas are accessible. While the cells painted black indicate obstacles, corresponding to areas that robots cannot reach in the real environment, such as walls, columns, pre-embedded parts, holes, etc.

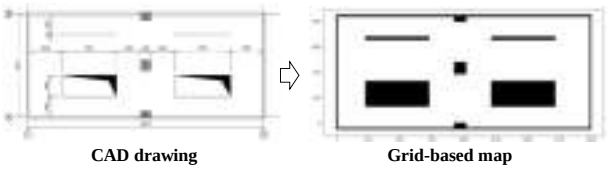


Figure 3. Obstacle map from CAD drawing

3.2 Robot Model

In this research, establishing an accurate and comprehensive mathematical model of swing-arm robot is the chief task. As shown in Figure 4, in our research, swing-arm robot is simplified into three parts: robot body (tracked vehicle), swinging arm and troweling disc. The body of robot is a racked vehicle, which controls the movement of robot. Tracked vehicle relies on the differential speed of its tracks for steering. Commonly, the geometric center of vehicle is used as the vehicle center and as well as control point of whole robot. The troweling disc is the main component for working, which can rotate around the swing-arm center within certain angle range. Notably, vehicle center and swing-arm center may not coincide. In order to ensure that the troweled areas are not disturbed, the robot's heading direction is the orientation of the tracked vehicle.

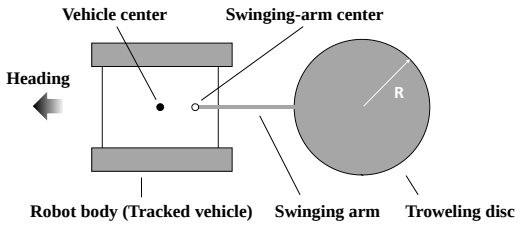


Figure 4. Swinging-arm robot model

As shown in Figure 5, the basic parameters of robot are defined: V_l and V_w represent the length and width of tracked vehicle; L_{vs} represents the offset between vehicle center and swing arm center; L_s represents the length of the swing arm; R represents the radius of troweling disc. Furthermore, if size of robot is determined, accurately defining the position of a swing-arm robot requires 4 parameters: x, y coordinate, yaw, and swing-arm angle (x, y, yaw, s). the x and y coordinates represent the accurate location of vehicle center ($\boldsymbol{V_c}$). The yaw is the angle between robot heading vector and positive x-axis direction, ranging from 0 to 2π (counterclockwise is positive). The swing-arm angle represents the angle deviation of the swing arm from its original position (counterclockwise is positive), and the range depends on the design parameters of robots. For convenience, this angle is specified as $[-\pi/2, \pi/2]$. The above four parameters are sufficient to definite any position of the robot. In a certain position, the effective work area is the coverage area of troweling disc. For convenience, use $\boldsymbol{S_c}: (S_x, S_y)$ and $\boldsymbol{D_c}: (D_x, D_y)$ to represent the center of swinging arm and troweling disc, separately. Then the relationship between $\boldsymbol{V_c}$, $\boldsymbol{S_c}$ and $\boldsymbol{D_c}$ can be expressed as follows:

$$\boldsymbol{V_c} = \boldsymbol{S_c} + L_{vs} \cdot \boldsymbol{h} \quad (1)$$

$$\boldsymbol{S_c} = \boldsymbol{D_c} + L_s \cdot (\cos(yaw + s), \sin(yaw + s)) \quad (2)$$

in which $\boldsymbol{h}$ represent the unit vector of heading direction:

$$\boldsymbol{h} = (\cos(yaw), \sin(yaw)) \quad (3)$$

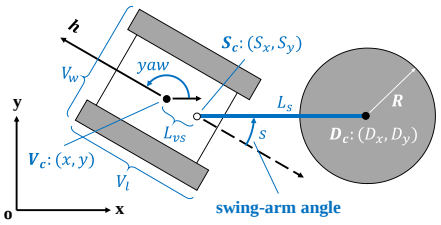


Figure 5. Four parameters for robot positioning: x, y, yaw, and swing-arm angle

On this basis, the collision principles of swing-arm robot are defined. Given that the tracked vehicle part and swinging arm part have different geometric shape, two kinds of collision principles are employed separately. As shown in Figure 6, if the position of an obstacle point is denoted as $\boldsymbol{O_c}: (O_x, O_y)$. In terms of tracked vehicle, collision detection is divided into two rounds. In the first round, the vehicle's circumscribed circle is used to exclude all obstacles outside the circle (obstacle No.1 in Figure 6, Equation 4).

$$\|\boldsymbol{O_c} - \boldsymbol{V_c}\|_2 \geq \frac{\sqrt{V_w^2 + V_l^2}}{2} \quad (4)$$

If any obstacles are inside the circle, proceed to the second round. By calculating the projections of the vector from the vehicle's center to the obstacle in both the heading direction and the perpendicular direction, the collision status can be determined. If the absolute values of these two projections are simultaneously less than half the vehicle's length and width, respectively, the obstacle is judged within the collision range (obstacle No.3 in Figure 6, Equation 5&6). Otherwise, it will not result in a collision (obstacle No.2 in Figure 6).

$$|(\boldsymbol{O_c} - \boldsymbol{V_c}) \cdot \boldsymbol{h}| \leq V_l/2 \quad (5)$$

$$\sqrt{\|\boldsymbol{O_c}-\boldsymbol{V_c}\|^2-|(\boldsymbol{O_c}-\boldsymbol{V_c})\cdot\boldsymbol{h}|^2}\le V_w/2 \quad (6)$$

As for swinging arm and disc, the collision cannot be judged by a simple circle or square. As shown in Figure 6, the avoidance area of this part is a capsule-like shape. This area can be solved by calculating the shortest distance from the obstacles to the line segment **Sc**-**Dc**. This process is also divided into 2 rounds. Firstly, eliminate most irrelevant obstacles through a circle externally connected to the avoidance area (obstacle No.4 in Figure 6, Equation 7).

$$\|\boldsymbol{O_c}-\boldsymbol{C}\|_2 \ge \frac{L_s+R}{2} \quad (7)$$

in where $\boldsymbol{C}$ is the center of avoidance area. As for the remaining obstacles, if they are within the perpendicular range of the endpoints of the line segment $\boldsymbol{D_c'}$-$\boldsymbol{D_c}$ ($\boldsymbol{D_c'}$ is the $\boldsymbol{D_c}$ about the symmetry point of $\boldsymbol{C}$), such as obstacle No.5 in Figure 6 (Equation 8).

$$|(\boldsymbol{O_c}-\boldsymbol{C})\cdot\boldsymbol{h'}| \le \frac{L_s-R}{2} \quad (8)$$

in where $\boldsymbol{h'}$ is the unit vector of $\boldsymbol{S_c}$ to $\boldsymbol{D_c}$.

$$\boldsymbol{h'}=\frac{\boldsymbol{D_c}-\boldsymbol{S_c}}{|\boldsymbol{D_c}-\boldsymbol{S_c}|} \quad (9)$$

In this case, their perpendicular distance to the line $\boldsymbol{D_c'}$-$\boldsymbol{D_c}$ needs to be greater than R to ensure no collision (Equation 10).

$$\sqrt{\|\boldsymbol{O_c}-\boldsymbol{C}\|^2-|(\boldsymbol{O_c}-\boldsymbol{C})\cdot\boldsymbol{h'}|^2} > R \quad (10)$$

If obstacles don't satisfy the Equation 8, such as No.6&7 in Figure 6. No collision means that the minimum distance between obstacle and $\boldsymbol{D_c'}/\boldsymbol{D_c}$ is greater than R (Equation 11).

$$\min(\|\boldsymbol{O_c}-\boldsymbol{D_c'}\|_2, \|\boldsymbol{O_c}-\boldsymbol{D_c}\|_2) > R \quad (11)$$

Notably, in the actual algorithm, all calculations can take redundancy into account to ensure sufficient safety.

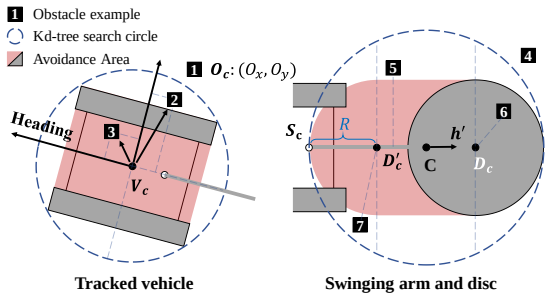


Figure 6. Collision principles of tracked vehicle, swinging arm and disc

4 SatAOI Algorithm for Delimitating AOI

In complex environments with a large number of pre-embedded parts and holes, delimitation of AOI will become more difficult and important. At the same time, since the working range of the troweling disc is separate from the main body of the robot, it further increases the challenges in the delimitation of AOI for the swing-arm robot. So, in this research, using the troweling disc as the main reference, whole map is searched to determine the AOI. The SatAOI algorithm is composed of two core parts: collision detection centered around disc (Algorithm 1: is_collision) and global AOI search (Algorithm 2: troweling_search).

Algorithm 1: is_collision. As shown in Figure 7, if troweling disc is put at certain coordinate, while swing-arm and vehicle are rotated to $angle1$ and $angle2$, separately, the yaw and swinging angle (s) can be calculated by Equation 12&13.

$$yaw = angle1 + angle2 \quad (12)$$

$$s = -angle2 \quad (13)$$

Then based on the robot model and collision principles in 3.2, the collision can be inferred by algorithm 1. If collision occurs, "True" is output. Otherwise, "False" is output.

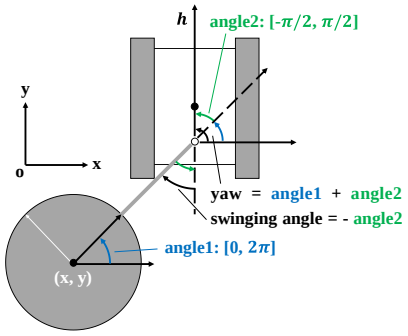


Figure 7. Collision detection algorithm using the troweling disc as the main reference

Algorithm 1: is_collision(D_c, *angle1*, *angle2*)

Input: Coordinate of disc center D_c: (x, y), rotate angles of swing-arm and vehicle *angle1*, *angle2*.
Output: If there is a collision, return True, else False.

```
// kdtree_o is specifically for storing obstacles
kdtree_o ← kd.KDTree(obstacle grids)
If kdtree_o.query_ball_point(D_c,R) then
    Return True
C ← center of capsule-like shape in Figure 6
// R is radius of disc
obstacle1 ← kdtree_o.query_ball_point(C, (L_S+R)/2)
If obstacle1 then
```

```
    h' ← unit vector of S_c to D_c. (Equation 9)
    For p in obstacle1 do
        vec ← p – C
        If |vec · h'| ≤ (L_s − R)/2 then
            If √(||vec||² − |vec · h'|²) ≤ R then
                Return True
        Else
            If min(||p − D'_c||₂, ||p − D_c||₂) ≤ R then
                Return True
    V_c ← Center of tracked vehicle
    r ← √(V_w² + V_l²) / 2
    obstacle2 ← kdtree_o.query_ball_point(V_c, r)
    If not obstacle2 then
        Return False
    Else
        For p in obstacle2 do
            vec ← p − V_c
            yaw ← angle1 + angle2
            h ← (cos(yaw), sin(yaw))
            proj ← vec · h
            proj_v ← ||vec||² − proj²
            If proj < V_l/2 and proj_v < V_w/2 then
                Continue
        Return False
```

Algorithm 2: troweling_search. Inspired by BFS (Breadth-first search) algorithm [19], the search process is a step-by-step exploring that spreads from one point to the surrounding areas. There are two sets: open set is points need to explore, and closed set is points already explored. Starting from a certain point (start node), the center of troweling disc is put at this point. Firstly, add the start node to the open set. Then, loop begins. While the open set is not empty, select the arbitrary node in the open set as the current node (also move it from the open set to the closed set). For each neighbor of the current node (other eight nodes centered around the current node in nine-square grid), if the neighbor is in the closed set, skip it as it has already been evaluated. Otherwise, stepping by $\pi/12$ and considering all discrete angle combinations ($24 \times 13 = 312$ pairs) of angle1&2 at this node, if there a at least a pair of certain angles that frees robot from collision, add this node to open set. Finally, all points that are connected to the start point and can be used to place a troweling disc (robot) are left in the closed set. So, this search algorithm can consider the connectivity of obstacle map.

Algorithm 2: troweling_search(*start_node*)

Input: The starting point of search *start_node*.
Output: area of interest *AOI*

```
// kdtree_v is specifically for storing points excluding obstacles
kdtree_v ← kd.KDTree(all grids excluding obstacle grids)
// Points need to explore open_set, points already explored closed_set
open_set ← {start_node}, closed_set ← {}
While open_set is not empty do
    current_node ← Arbitrary node in open_set
    Remove current_node from open_set
    For node in neighbors of current_node:
        If node in closed_set then
            Continue
        areso ← π/12  // angular resolution
        angle1 ← range(0, 2π, areso)
        angle2 ← range(-π/2, π/2, areso)
        For a1, a2 ← list(zip(angle1, angle2)) do
            If is_collision(node, a1, a2) then
                Continue
            Else
                Add node to open_set
                Break
For node in closed_set do
    For point in kdtree_v.query_ball_point(node, cR) do
        Add point to AOI
Return AOI
```

Based on SatAOI algorithm, all possible positions of the troweling disc have been determined, and the envelope region composed of all discs is the final theoretical limit AOI.

5 Case Study

In order to verify the feasibility of the proposed method, taking architectural floor plans of a substation under different complexity as research cases, the robustness of SatAOI algorithm for different scenes and its effectiveness for connected and non-connected areas are tested. Based on this, the importance of AOI delimitation for coverage path planning is discussed.

5.1 Troweling Robot Parameters

The troweling robot from MagicBIM is selected as the robot prototype for simulation. As shown in Figure 8, the tracked vehicle is simplified into a rectangle with a length (V_l) of 875mm and a width (V_w) of 830mm. The center of swing arm deviates from the center of vehicle (L_{vs}) by 142mm, the length of the swing arm (L_s) is 995mm, and the radius of the troweling disc (R) is 350mm. At the same time, the robot can achieve stationary turning by rotating the tracks on both sides in opposite directions. And the maximum swing-arm range is $[-\pi/2, \pi/2]$. These parameters are sufficient to characterize a typical swing-arm troweling robot, including the basic rules for motion, collision and troweling.

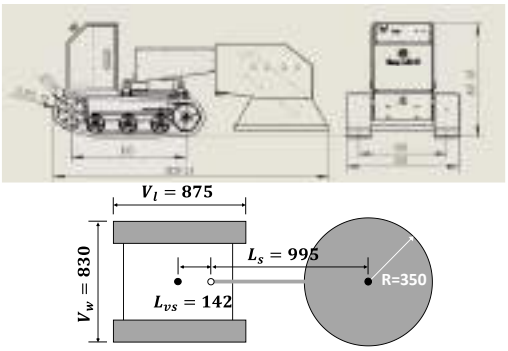


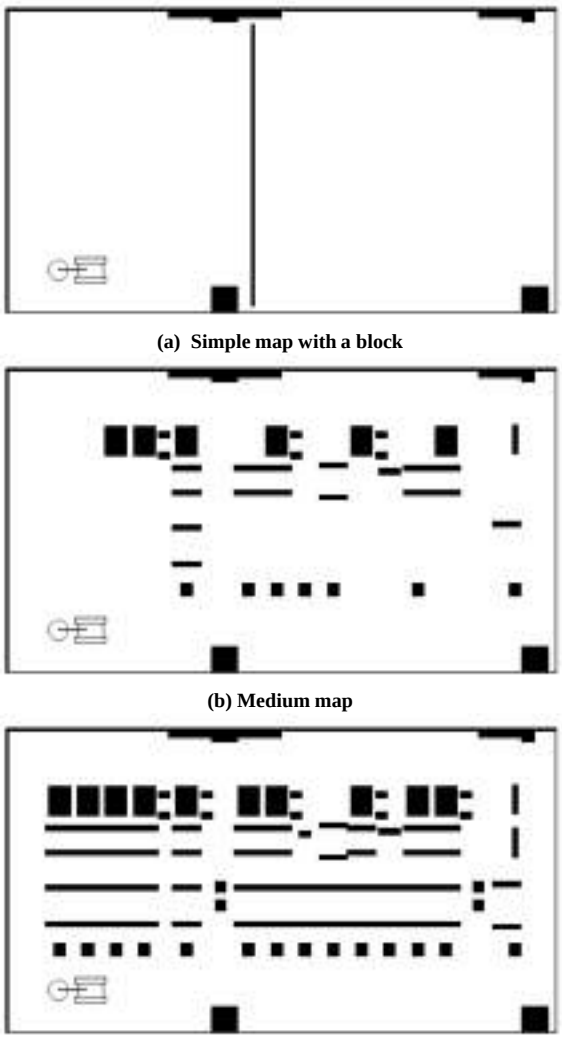
Figure 8. Prototype and key parameters of robot

5.2 Obstacle Maps Generation

The gas insulated switchgear (GIS) room is an important part in substations. However, a large number of electrical equipment requires a large number of embedded parts and holes to be reserved during the construction process, which poses a huge challenge for automation of troweling robots. So, using CAD drawings of different GIS rooms as the data source, obstacle maps under different complexity are generated for testing. As shown in Figure 9, floor plans are converted into grid-based maps using squares with a side length of 10mm. And there are three complexity levels of obstacle maps, denoted as "simple map", "medium map" and "complex map", separately. The overall size of three scenes is same of 10,450mm * 19,190mm. And size of Obstacles (embedded parts and holes) varies from 160mm to 2000mm. Notably, in order to test the ability of SatAOI algorithm to recognize map connectivity, a bottom-to-top block is added to the "simple map" and divides the map into left and right parts. According to statistics, the area proportions of obstacle in each map are shown in Table 1. The area of entire map reached 200.54 m^2, and the highest obstacle area proportion is 12.89%.

Table 1. Statistics of obstacles

Map type	Map area (m^2)	Obstacle area (m^2)	Proportion (%)
Simple		7.00	3.49
Medium	200.54	14.96	7.46
Complex		25.84	12.89



(a) Simple map with a block

(b) Medium map

(c) Complex map

Figure 9. Obstacle maps and robots. (a) (b) (c) represent maps under different complexity

In order to demonstrate challenges of obstacles to the swing-arm troweling robot in a more intuitive way, robots and obstacle maps are simultaneously visualized in Figure 9. It is evident that the complex obstacles in GIS room pose significant challenges to the swing-arm troweling robots, even if manual operations are also extremely difficult in this case. There is no doubt that the delimitation of AOI can assist in the planning and simulation of swing-arm troweling robots.

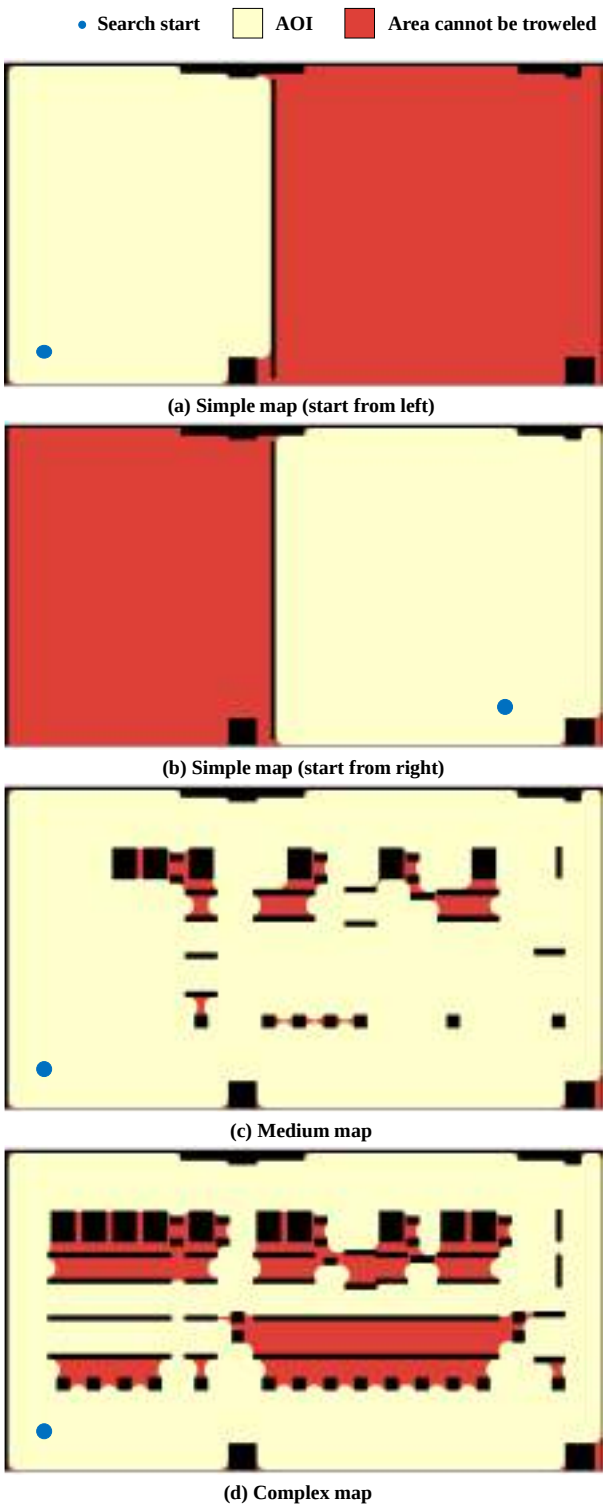


Figure 10. AOI results of different obstacle maps

5.3 AOI Delimitation

The results of SatAOI algorithm are shown in the Figure 10, in which the yellow areas represent the final AOI in all maps. Overall, the all results of AOI are reasonable and effective, only narrow gaps between obstacles and wall corners cannot be troweled by robots. Regardless of the complexity of maps, AOI results effectively represent the maximum troweling areas of the swing-arm robot. In complex maps (Figure 10 d), the algorithm can identify which areas between obstacles can be troweled and which cannot, bringing great convenience to robot simulation. Furthermore, in the simple map (Figure 10 a&b), AOI include left part and right part due to different search starts. This is because the left and right sides of the map are separated by a bottom-to-top block. And in reality, the robot cannot complete the troweling work of both parts at once. The SatAOI algorithm can identify the connectivity between different areas, ensuring that the AOI result can be troweling by the robot in a single task.

These analysis results can be applied to some downstream tasks for construction robots, such as assisting in the path planning and simulation of swing-arm troweling robot in complex construction sites. According to the AOI results, the robot can determine which areas can be reached and troweled, and which areas will not be accessed. On this basis, AOI can be further divided into small cells for CPP task. AOI is the goal of coverage path optimization, and will also participate in the evaluation process of path planning.

6 Conclusion

Although robots have improved the mechanization and automation of construction, swing-arm troweling robot still rely on manual operation in complex construction sites. In this research, the mathematical models of swing-arm robot and obstacle map are built. On this basis, collision detection principle and search method are proposed, together formed the SatAOI algorithm for the AOI delimitation of swing-arm troweling robot. A case study of GIS room from a real substation construction is conducted and the results show that SatAOI algorithm can effectively determine the AOI in maps under different complexity, and meanwhile can distinguish the connected and disconnected areas from a certain start.

The delimitation of AOI is only a first and primary task of troweling robot simulation. Further research is the downstream application of AOI. Cellular decomposition method or other CPP algorithm based on AOI results will greatly enhance the automation level of swing-arm troweling robots. AOI-based coverage path optimization and evaluation can further improve the work efficiency of the troweling robot. In addition, even if the robot cannot be fully automated, AOI can assist workers in achieving more efficient human-robot collaboration and improving construction productivity [20].

7 Acknowledgement

This research was funded by the National Key R&D Program of China (grant No. 2023YFC3805800), the project "Research on Key Technologies for Mechanized Construction of Substation Civil Engineering" of "Technology project funding from State Grid Corporation of China" (project No. 5200-202311481A-3-2-ZN) and the project "Constructing a theoretical system for sustainable development strategy of construction enterprises with digital platforms as the core" from China Construction Third Engineering Bureau Beijing Co., Ltd.

References

[1] Kamel S. Saidi, Thomas Bock, Christos Georgoulas, Thomas Bock, Christos Georgoulas, *Robotics in Construction*, pp. 1493–1520, Springer Handbook of Robotics, 2016.

[2] M. A. Faheem, N. Zafar, P. Kumar, M. Hassan, N. U. Prince, and M. A. A. Mamun, AI AND ROBOTIC: ABOUT THE TRANSFORMATION OF CONSTRUCTION INDUSTRY AUTOMATION AS WELL AS LABOR PRODUCTIVITY, *Remittances Review*, vol. 9, no. S 3, pp. 871–888, 2024.

[3] A. E. Oke, J. Aliu, P. Fadamiro, P. S. Jamir Singh, M. S. Samsurijan, and M. Yahaya, Robotics and automation for sustainable construction: microscoping the barriers to implementation, *SASBE*, vol. 13, no. 3, pp. 625–643, Apr. 2024.

[4] J. M. Davila Delgado *et al.*, Robotics and automated systems in construction: Understanding industry-specific challenges for adoption, *Journal of Building Engineering*, vol. 26, p. 100868, Nov. 2019.

[5] J. Brozovsky, N. Labonnote, and O. Vigren, Digital technologies in architecture, engineering, and construction, *Automation in Construction*, vol. 158, p. 105212, Feb. 2024.

[6] T. Bock, T. Linner, *Construction Robots: Elementary Technologies and Single-Task Construction Robots*, Cambridge University Press, 2017.

[7] A. Więckowski, "JA-WA" - A wall construction system using unilateral material application with a mobile robot, *Automation in Construction*, vol. 83, pp. 19–28, Nov. 2017.

[8] C. Kasperzyk, M.-K. Kim, and I. Brilakis, Automated re-prefabrication system for buildings using robotics, *Automation in Construction*, vol. 83, pp. 184–195, Nov. 20172.

[9] M. Hwang-Bo *et al.*, Development of an Unmanned Autonomous Concrete Floor Robotic Troweling System, presented at the 16th International Symposium on *Automation and Robotics in Construction*, Madrid, Spain, Sep. 1999.

[10] S. Zhao *et al.*, Application and Development of Autonomous Robots in Concrete Construction: Challenges and Opportunities, *Drones*, vol. 6, no. 12, p. 424, Dec. 2022.

[11] G. Carra, A. Argiolas, A. Bellissima et al., Robotics in the Construction Industry: State of the Art and Future Opportunities, In Proceedings of the 35th International Symposium on *Automation and Robotics in Construction (ISARC)*, Berlin, Germany, 2018.

[12] Bright Dream Robotics. Ground Troweling Robot. On-line: https://www.bzlrobot.com/contents/3/88.html, Accessed: 23/12/2024.

[13] FangShi Technology. Crawler Trowel Robot. On-line: https://www.fangshitech.com/page/57/7.html, Accessed: 23/12/2024.

[14] M.-H. Wu and J.-R. Lin, An Agent-based Approach for Modeling Human-robot Collaboration in Bricklaying, *37th International Symposium on Automation and Robotics in Construction*, Kitakyushu, Japan, Oct. 2020.

[15] Y. Li, W. Tao, Y. Xiao, L. Gong, and C. Liu, Path planning in continuous adjacent farmlands and robust path-tracking control of a rice-seeding robot in paddy field, *Computers and Electronics in Agriculture*, vol. 210, p. 107900, 2023.

[16] E. Galceran and M. Carreras, A survey on coverage path planning for robotics, *Robotics and Autonomous Systems*, vol. 61, no. 12, pp. 1258–1276, Dec. 2013.

[17] C. S. Tan, R. Mohd-Mokhtar, and M. R. Arshad, A Comprehensive Review of Coverage Path Planning in Robotics Using Classical and Heuristic Algorithms, *IEEE Access*, vol. 9, pp. 119310–119342, 2021.

[18] J. P. Bailey, A. Nash, C. A. Tovey, and S. Koenig, Path-length analysis for grid-based path planning, *Artificial Intelligence*, vol. 301, p. 103560, Dec. 2021.

[19] Cormen, T.H., Leiserson, C.E., Rivest, R.L. and Stein, C., Introduction to Algorithms, Second Edition. *MIT Press and McGraw-Hill*, Section 22.2: Breadth-first search, pp. 531–539, 2001.

[20] M. Wu, J.-R. Lin, and X.-H. Zhang, How human-robot collaboration impacts construction productivity: An agent-based multi-fidelity modeling approach, *Advanced Engineering Informatics*, vol. 52, p. 101589, Apr. 2022

Real-time 3D Perception System of Construction Equipment for Autonomous Mobile Robots

Duho Chung[a], Seunghun Im[a], Yohan Kim[a] and Hyoungkwan Kim[a]

[a]Department of Civil and Environmental Engineering, Yonsei University, Korea
E-mail jungduho1@yonsei.ac.kr, sunsetmorning1234@gmail.com, homez815@yonsei.ac.kr and hyoungkwan@yonsei.ac.kr

Abstract –

Automation using mobile robots at construction sites is gaining attention as a key technology to enhance productivity and safety. However, due to the unstructured nature of construction sites with dynamic elements like construction equipment, robots must perceive their positions and movements in real-time to adapt to workflow changes while performing tasks. To address this issue, this paper proposes a system that enables mobile robots to detect the spatial positions of construction equipment in real time. The system integrates a voxel-based deep learning detection model with a Simultaneous Localization and Mapping (SLAM) algorithm, allowing real-time detection of construction equipment within a consistent reference coordinate system using 3D point cloud data collected by the mobile LiDAR. Experimental results demonstrate that the system achieves an average detection accuracy of 74.51% (AP@R40, IoU 0.25) for excavators, dump trucks, graders, and rollers, demonstrating its capability to accurately recognize the spatial information of construction equipment. This approach highlights the potential for mobile robots to effectively perceive construction equipment in active construction sites, enabling safer and more efficient task execution.

Keywords –

Point Cloud Data; 3D Object Detection; Deep Learning; Vehicle Tracking; Autonomous Robot; Quadruped Robot.

1 Introduction

To address the persistent labor-intensive challenges in the construction industry, construction automation using mobile robots has emerged as a key technology to enhance productivity and safety on construction sites. Various robotic platforms, including drones and quadruped robots, are being explored for applications such as construction automation, 3D reconstruction, material assembly, and safety monitoring [1]. However, the low level of standardization and the dynamic nature of construction environments present significant challenges for robotic automation [2]. In active construction sites, mobile robots share space with dynamic obstacles like construction equipment. Each type of equipment follows distinct operational patterns, often working within a fixed radius or moving along predefined paths. Failure to recognize these machines can increase collision risks and lead to frequent, unnecessary path adjustments, disrupting workflow. To operate efficiently in dynamic construction environments, mobile robots must detect the location and type of construction equipment in real-time and track their movements over time.

Several studies have explored the automation of tasks using mobile robots on outdoor construction sites by employing sensors to identify and respond to obstacles. Kim et al. [3] automated laser scanning tasks with a mobile robot using initial map generated by a UAV (Unmanned Areial Vehicle) and addressed obstacles through a point cloud-based potential field. Asadi et al. [4] developed a mobile manipulator robot capable of grasping and transporting objects using vision-based environmental understanding, employing camera-generated point clouds for obstacle avoidance. Štibinger et al. [5] proposed a mobile manipulator robot for automated structural assembly, leveraging 3D point clouds to detect obstacles and perform navigation. Chung et al. [6] automated laser scanning of scaffold structures on outdoor construction sites without prior information, utilizing point clouds from a mobile LiDAR and IR camera to navigate around obstacles.

While these studies have demonstrated the automation of construction tasks using robots capable of avoiding dynamic obstacles in outdoor environments, there is a lack of research focused on accurately detecting construction entities and continuously recording their locations using point clouds to support robot operations. Studies such as [7] and [8] have explored situational awareness and interaction between mobile robots and construction workers, aiming to implement Human-

Robot Collaboration (HRC) systems. However, research on perception systems for mobile robots specifically designed to interact with construction equipment in 3D space remains significantly underexplored.

This study proposes a system that uses mobile LiDAR to detect the 3D positions of construction equipment in real-time and record them within a global coordinate frame. Figure 1 provides an overview of the proposed system, which integrates a voxel-based 3D object detection model with Simultaneous Localization and Mapping (SLAM). First, the system processes 3D point cloud data collected by the mobile LiDAR to detect various types of construction equipment, such as excavators, dump trucks, graders, and rollers, in real-time. The detected objects are then managed within a global coordinate frame through SLAM, enabling continuous tracking of construction equipment even in dynamic environments. The entire system is implemented on the Robot Operating System (ROS) framework, allowing seamless integration and real-time processing of sensor data acquired by the mobile robot. Using the proposed system, mobile robots can effectively perceive the locations of construction equipment on-site, providing a foundation for applications such as maintaining safe distances, identifying hazardous zones, and understanding work contexts.

Additionally, mobile LiDAR offers a significant advantage by covering a wide area while being less affected by occlusions, enabling the real-time acquisition of high-density 3D data [9]. Due to these characteristics, it is increasingly being integrated with SLAM for 3D reconstruction and safety management in construction sites [10]. The proposed system enables real-time perception of construction equipment movements using mobile LiDAR, which can be utilized to develop a comprehensive progress monitoring system for construction sites, including quality control of individual structures and process monitoring of construction equipment. Based on the potential of mobile LiDAR-based technology, this system supports mobile robot operations on construction sites and lays the foundation for establishing a comprehensive progress monitoring system for large-scale construction projects.

2 Methodology

The proposed system is designed to accurately identify the spatial positions of construction equipment in real time using 3D point cloud data collected by mobile LiDAR mounted on a robot. The mobile LiDAR continuously captures 3D point cloud data as the robot navigates the environment. This data is processed in real time by a voxel-based 3D object detection model, which detects construction equipment and generates bounding boxes. These detection results are then integrated with a SLAM algorithm, transforming them into a consistent global coordinate system, where they are recorded for further use.

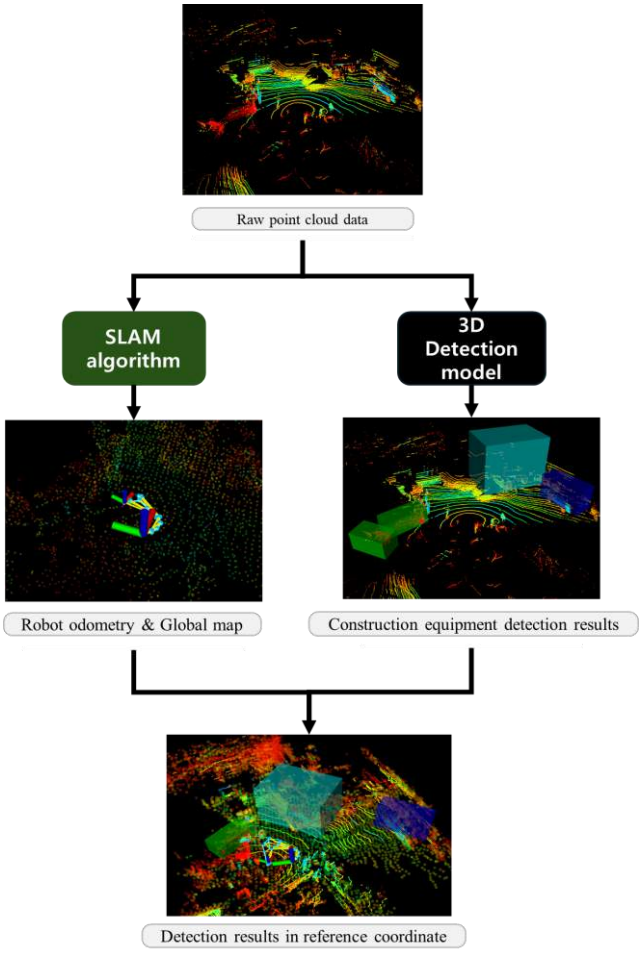


Figure 1. Overview of the proposed system

As shown in Figure 2, the mobile robot system comprises a quadruped robot (Unitree AlienGo), a mobile LiDAR (Velodyne VLP-32), an IMU sensor (Microstrain), and a mini-PC equipped with an NVIDIA GeForce RTX 2060 (Intel NUC 11 PHKi7C). This system was used to acquire the data necessary for model training and experimentation. The mini-PC was directly mounted on the quadruped robot along with a battery, enabling the entire proposed system to operate in a mobile environment. To construct a point cloud dataset for model training, the mobile robot was remotely operated to collect point cloud data at a construction site. During this process, the mobile LiDAR scanned excavators, dump trucks, graders, and rollers from various angles and distances, ensuring diversity in the training data.

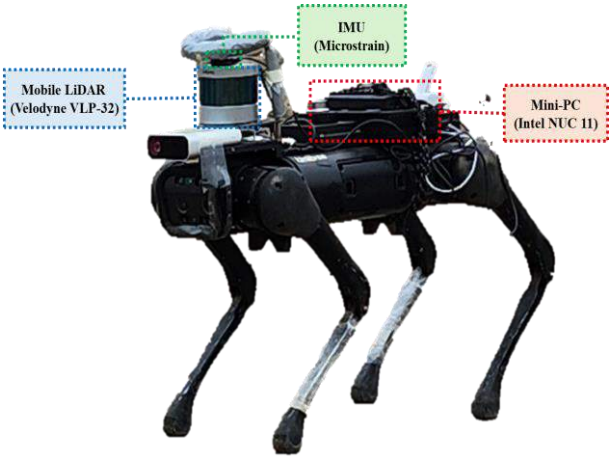


Figure 2. Hardware configuration.

For the mobile robot to understand and respond to the movements of construction equipment, it must rapidly and accurately identify their spatial positions despite limited computational resources. A voxel-based approach efficiently processes 3D point cloud data, significantly reducing computational complexity for large datasets. To achieve real-time detection of construction equipment in 3D point cloud data, we employed the voxel-based 3D object detection model VoxelNeXt [11]. VoxelNeXt utilizes a fully sparse convolutional network to directly predict 3D objects from voxel features, maximizing computational efficiency by eliminating the need for generating dense features after voxel transformation. This makes it particularly well-suited for mobile robot environments with limited computational resources. To train the model efficiently with limited data, VoxelNeXt was pre-trained on the KITTI dataset [12], and transfer learning was performed using this pre-trained model as a starting point. Figure 3 illustrates the architecture of VoxelNeXt. VoxelNeXt utilizes a six-stage Sparse CNN Backbone to process 3D point cloud data. Instead of generating a dense feature map, it applies Sparse Height Compression to create a BEV (Bird's Eye View) feature, preserving only the original data points in sparse format. This approach reduces unnecessary computations and ensures efficient object detection even with limited resources.

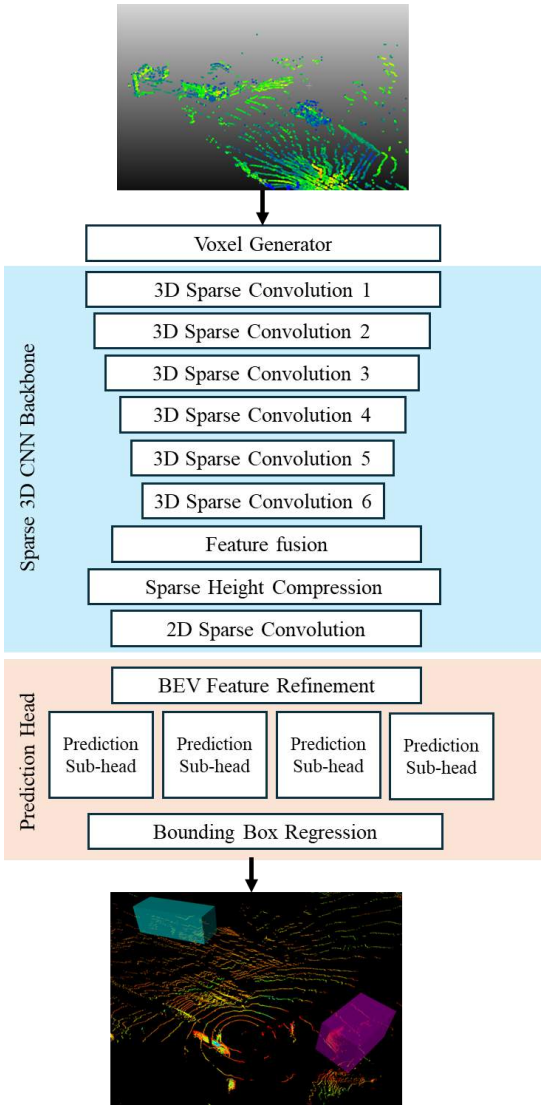


Figure 3. Structure of VoxelNeXt.

In environments where both mobile robots and construction equipment are in motion, integrating detection results into a consistent reference coordinate system is essential for long-term perception of construction equipment positions. To achieve this, the proposed method integrates the SLAM algorithm with VoxelNeXt to process detection results. SLAM is an algorithm that simultaneously maps the environment around a sensor and estimates the sensor's position using the sensor data. The proposed method utilizes the LIO-SAM [13] algorithm, which combines mobile LiDAR and IMU data to generate precise maps. By utilizing odometry information calculated by LIO-SAM, the detected bounding boxes are transformed into the reference coordinate system, enabling accurate localization of construction equipment on the global map

of the construction site. Figure 4 illustrates the process of registering detected equipment in the global coordinate frame of LIO-SAM using VoxelNeXt.

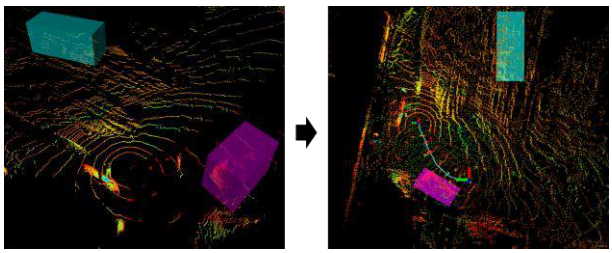


Figure 4. Transformation of detection results: from local coordinates (left) to global map using LIO-SAM (right)

3 Experiments and Results

In this study, a quadruped robot was remotely operated at four large-scale construction sites to collect mobile LiDAR data of construction equipment for VoxelNeXt model training. While navigating active construction sites with moving equipment, the quadruped robot recorded continuous point cloud data. The collected data was captured at one frame per three seconds and converted into PCD file format to construct the dataset.

During the data labelling process, only construction equipment within 40 meters of the robot was selected for annotation. Objects that were visually unidentifiable due to occlusion or motion blur were excluded from labelling. Figure 5 presents examples of the collected data from each site. A total of 1,246 frames of point cloud data were used for training and validation, while 261 frames of point cloud data were reserved for testing. This study focused on training the model to recognize four classes of construction equipment: excavators, dump trucks, graders, and rollers. Detailed class-wise data distribution is provided in Table 1.

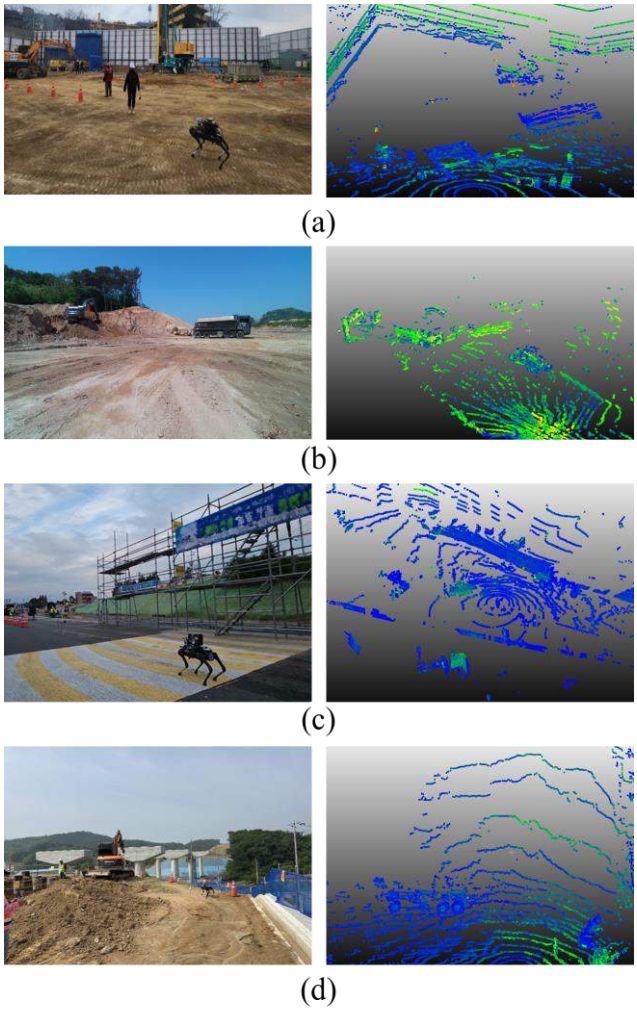


Figure 5. Picture of the site (left) and examples of point cloud data collected at each site(right); Site 1, 2 and 3 were used for training and validation, and Site 4 was used for testing; (a) Site 1, (b) Site 2, (c) Site 3, (d) Site 4.

Table 1. Class-wise dataset distribution

Class	Train	Validation	Test
Excavator	877	175	95
Dump Truck	564	121	71
Grader	199	72	117
Roller	125	32	118
PCD files	1000	246	261

The training of the construction equipment detection model was conducted on an Ubuntu 20.04 environment. Transfer learning was performed using pre-trained

weights from the KITTI dataset, with the model trained for 50 epochs using a batch size of 4. The training process employed an AdamW-based one-cycle learning schedule, where the learning rate increased to a maximum of 0.01 during the first 40% of training and decreased over the remaining 60%. These settings were based on the default VoxelNeXt configuration for training on the KITTI dataset.

During the transfer learning process, fine-tuning was conducted under four different configurations, based on the extent to which the 3D sparse convolution layers in the backbone were frozen. These layers correspond to "3D Sparse Convolution 1–6" in Figure 3. 1) Full fine-tuning was performed, with no layers frozen. 2) Only the first two 3D sparse convolution layers were frozen; the rest of the backbone and the detection head were trained. 3) Only the first four 3D sparse convolution layers were frozen; the remaining layers and the detection head were trained. 4) All 3D sparse convolution layers were frozen, while the detection head and subsequent 2D layers remained trainable. The results showed that full fine-tuning without freezing achieved the highest average class performance, which was used as the basis for selecting the final model configuration.

On the test set, the 3D bounding box accuracy based on AP@R40 IoU 0.25 was recorded as follows: dump truck (79.13%), excavator (70.81%), grader (68.74%), and roller (79.36%). Here, R40 (Recall 40) is an evaluation metric used in the KITTI dataset, where Average Precision (AP) is computed based on 40 recall points. The detailed performance evaluation results are presented in Table 2, while Figure 6 visually illustrates the detection results for each class.

To verify whether the proposed system can reliably perceive construction equipment in real-time environments using data acquired by an actual mobile robot, a live experiment was conducted using sequential point cloud data collected from Site 4. The results demonstrated that VoxelNeXt, implemented within the proposed ROS-based system, detected construction equipment at an average speed of 2.8 Hz on the mini-PC. The detection results were continuously integrated with LIO-SAM in real time, allowing the positions of construction equipment to be sequentially localized on the global map. Figure 7 illustrates the detection results of a continuously moving grader and roller over a 60-second period, recorded on the global map in the test environment. These results demonstrate that the proposed system effectively detects and records the spatial information of construction equipment even in real-time operational conditions.

Table 2. The performance of detection model

Class	AP (IoU > 0.5)		AP (IoU > 0.25)	
	BEV	3D	BEV	3D
All Layers Trainable (No Freezing)				
Excavator	66.59	41.97	70.81	70.81
Dump Truck	78.97	47.90	79.31	79.13
Grader	68.45	50.02	70.14	68.74
Roller	75.62	41.37	80.13	79.36
Freeze 3D Sparse Convolution 1-2				
Excavator	54.05	39.60	55.48	55.48
Dump Truck	82.60	45.21	83.59	81.08
Grader	69.85	53.66	75.99	73.92
Roller	73.14	39.01	77.76	76.41
Freeze 3D Sparse Convolution 1-4				
Excavator	42.38	38.32	54.82	53.12
Dump Truck	83.73	43.20	84.08	82.30
Grader	70.22	51.40	75.45	72.77
Roller	65.67	29.02	74.41	74.41
Freeze All 3D Sparse Convolution				
Excavator	20.63	12.45	28.94	28.27
Dump Truck	85.37	36.93	86.38	81.98
Grader	43.80	33.42	53.65	47.59
Roller	70.97	14.55	82.12	61.19

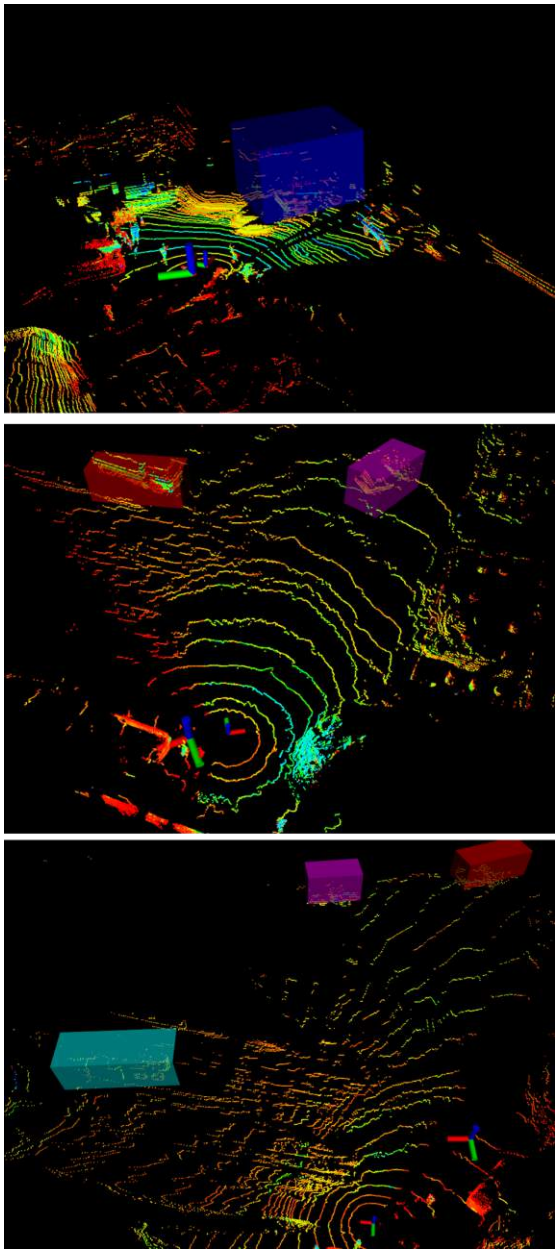


Figure 6. Visual examples of the detection outputs at test dataset: blue boxes represent excavator, red boxes represent truck; cyan boxes represent grader, and pink boxes represent roller.

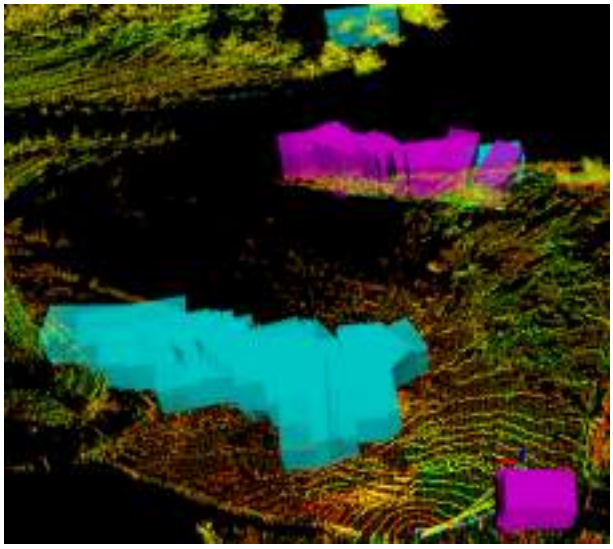


Figure 7. Visualization of grader(cyan) and roller(pink) bounding boxes recorded on the global map at Site 4.

4 Discussion

The experimental results confirmed that the proposed methodology can autonomously perceive construction equipment in real-world construction site data. By leveraging VoxelNeXt, the system demonstrated the ability to detect construction equipment in real time across a wide area and integrate the detection results with LIO-SAM, effectively mapping their positions within a consistent coordinate system. Furthermore, the system successfully recorded equipment positions continuously, enabling long-term perception of their movements and operational ranges. This capability supports mobile robot operations on construction sites and highlights the system's potential as foundational technology for comprehensive progress monitoring in large-scale construction projects.

However, the visually sparse nature of mobile LiDAR point clouds presents challenges in maintaining consistent detection performance, particularly in complex environments. In construction sites where structures and equipment are densely intermingled, achieving robust detection necessitates further research. These sparsity issues become more pronounced when applying technologies originally developed for autonomous driving to complex off-road environments, such as construction sites.

As shown in Table 2, when fine-tuning with more backbone layers frozen from the pretrained model (trained on the KITTI dataset), detection performance remained high for classes similar to those in KITTI, such as dump trucks. In contrast, performance significantly declined for equipment structurally different from KITTI

objects, such as excavators. This demonstrates a clear domain gap between autonomous driving datasets and construction site data, indicating that structural diversity of construction equipment and environmental complexity directly influence detection performance.

Therefore, acquiring construction-specific datasets is essential. Given the current lack of standardized 3D point cloud datasets for construction equipment, generating realistic, variability-inclusive virtual data is an effective alternative for dataset augmentation. This approach can significantly enhance data diversity, improve the generalization capability and robustness of detection models, and ultimately strengthen detection reliability in complex construction environments.

5 Conclusion

This paper proposed a system capable of perceiving the 3D positions of construction equipment in real time using a mobile LiDAR. The system integrates a voxel-based real-time object detection model with a SLAM algorithm to efficiently process 3D point cloud data collected by the mobile LiDAR and consistently map detected results in a global coordinate system. Experimental results demonstrated that the system achieves detection accuracies of 79.13% (dump truck), 70.81% (excavator), 68.74% (grader), and 79.36% (roller) based on AP@R40 with 3D IoU 0.25, proving its capability to reliably recognize the spatial locations of construction equipment. Additionally, experiments confirmed that mobile LiDAR-based detection can effectively identify and track a wide range of construction equipment in dynamic construction environments in real time. Future work will focus on enhancing detection performance and reliability by augmenting the dataset with realistic, simulation-based virtual data, thereby improving the generalization capability and robustness of the detection model. Through these advancements, the proposed system has the potential to support efficient mobile robot operations on construction sites and serve as a foundational technology for automated progress monitoring of large-scale construction projects using mobile LiDAR.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Ministry of Science and ICT (No.RS-2024-00457308) and the "National R&D Project for Smart Construction Technology (RS-2020-KA156488)" funded by the Korea Agency for Infrastructure Technology Advancement under the Ministry of Land, Infrastructure and Transport, and managed by the Korea Expressway Corporation.

References

[1] L. Zeng, S. Guo, J. Wu, & B. Markert. Autonomous mobile construction robots in built environment: A comprehensive review. Developments in the Built Environment, 19: 100484, 2024.

[2] B. Xiao, C. Chen, & X. Yin. Recent advancements of robotics in construction. Automation in Construction, 144: 104591, 2022.

[3] P. Kim, J. Park, Y. K. Cho, & J. Kang. UAV-assisted autonomous mobile robot navigation for as-is 3D data collection and registration in cluttered environments. Automation in Construction, 106: 102918, 2019

[4] K. Asadi., V. R. Haritsa, K. Han, & J. P. Ore. Automated object manipulation using vision-based mobile robotic system for construction applications. *Journal of Computing in Civil Engineering*, *35*(1): 04020058, 2021.

[5] P. Štibinger, G. Broughton, F. Majer, Z. Rozsypálek, A. Wang, K. Jindal, A. Zhou, & M. Saska, Mobile manipulator for autonomous localization, grasping and precise placement of construction material in a semi-structured environment. *IEEE Robotics and Automation Letters*, *6*(2): 2595-2602, 2021.

[6] D. Chung, J. Kim, S. Paik, S. Im, & H. Kim. Automated system of scaffold point cloud data acquisition using a robot dog. *Automation in Construction, 170*: 105944, 2025.

[7] Y. Liu, M. Habibnezhad, & H. Jebelli. Brainwave-driven human-robot collaboration in construction. Automation in Construction, 124: 103556, 2021.

[8] Y. Kim, S. Kim, Y. Chen, H. Yang, S. Kim, S. Ha, M. Gombolay, Y. Ahn, & Y. K. Cho. Understanding human-robot proxemic norms in construction: How do humans navigate around robots? *Automation in Construction*, *164*: 105455, 2024.

[9] D. Hauser, C. Glennie, & B. Brooks. Calibration and accuracy analysis of a low-cost mapping-grade mobile laser scanning system. Journal of Surveying Engineering, 142(4), 04016011, 2016

[10] J. Kim, J. Kim, S. Paik, & H. Kim. Point cloud registration considering safety nets during scaffold installation using sensor fusion and deep learning. *Automation in Construction*, *159*: 105277, 2024.

[11] Y. Chen, J. Liu, X. Zhang, X. Qi, & J. Jia. (2023). Voxelnext: Fully sparse voxelnet for 3d object detection and tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 21674-21683).

[12] A. Geiger, P. Lenz, C. Stiller, & R. Urtasun. Vision meets robotics: The kitti dataset. *The*

international journal of robotics research, *32*(11), 1231-1237, 2013.

[13] T. Shan, B. Englot, D. Meyers, W. Wang, C. Ratti, & D. Rus. “Lio-sam: Tightly-coupled lidar inertial odometry via smoothing and mapping,” in *2020 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, IEEE: pp. 5135–5142, 2020.

Dynamics Model of an Excavator in an Open Source Physical Simulator for Developing Autonomous Construction

Taro ABE[1], Yosuke Matsusaka[1], Daisuke Endo[1], Genki Yamauchi[1], Takeshi Hashimoto[1]

[1]1Public Works Research Institute, Japan

abe-t575co@pwri.go.jp, matsusaka@mid-japan.com, endou-d177cl@pwri.go.jp, yamauchi-g573bs@pwri.go.jp, t-hashimoto@pwri.go.jp

Abstract -

Japan's Public Works Research Institute (PWRI) has developed OPERA (Open Platform for Earthwork with Robotics and Autonomy), an open source-based research and development environment consisting of construction machinery and its physical simulator, to promote the development of automated construction technology. Constructing a dynamic model of construction machinery is key to reproducing its behavior in a physical simulation. However, manufacturers typically treat such information as proprietary, making it difficult for users to access and utilize. This paper shows a Dynamics model based on available information on the user side for a hydraulic excavator implemented in OPERA's physical simulator. The behavior of each joint in the simulator is compared with the actual behavior of each joint under the same joint angle commands. The results show that the actual excavator behavior can be simulated in each joint on the simulator excavator.

Keywords -

Autonomous Construction; Dynamic Model; Excavator;

1 Introduction

Construction sites, issues of labor shortages due to aging populations and population decline are becoming critical challenges. As a result, reducing the number of required workers has become an urgent priority. Efforts to achieve this through automation are gaining attention, such as replacing excavation and transportation tasks with automated construction machinery or automating entire construction operations.

Japan's Public Works Research Institute has developed OPERA (Open Platform for Earthwork with Robotics and Autonomy) for the development of autonomous construction machinery and automated construction technology [1]. OPERA is an open platform designed to facilitate collaborative development not only with construction machinery manufacturers and construction companies but also with diverse industries such as academic research institutions, software vendors, sensor manufacturers, and system integrators. OPERA comprises three main elements: hardware, including modified construction machinery such as hydraulic excavators and crawler dumpers to allow electronic control; a physical simulator that simulates the behavior of these construction machines; and middleware for controlling them. Developers of automation programs can conduct preliminary verification using the simulator, and if no issues arise, proceed to verification with actual machines. This enables model-based development of automation programs. To replicate the behavior of actual machines in simulations, it is crucial to have accurate dynamic models of the target construction machinery. Examples of model-based development utilizing simulators in construction machinery include works by Kawabata et al. [2] and Shintani et al. [3]

Here, we describe other cases of model-based development of construction machinery. Kawabata et al. [2] describe the use of simulations for dynamic behavior analysis of hydraulic excavators to shorten product development cycles. They evaluated behaviors during flatland and rough terrain traversal using dynamic models, achieving reductions in development time. Shintani et al. [3] proposed model-based development employing 1DCAE technology for performance development of construction machinery. Using 1DCAE technology, they improved design change efficiency and product quality, employing simulations during excavation, among other scenarios.

In model-based development by construction equipment manufacturers, the machine dynamics model is confidential information, so it will not normally be disclosed to all users. Thus, this is a major bottleneck for users such as construction companies. To address such issues, we propose a method for constructing dynamics models based on the following concepts.

1. *Using user-friendly information (e.g., product catalogs for construction machinery):*This study target is to enable users to develop automation systems even if they are unfamiliar with the dynamics model of construction machinery.
2. *Simulating each joint motion in a construction machine operated by hydraulic cylinders, etc., with a simple rotational drive.*This study aims to enable users who are not familiar with hydraulic equipment

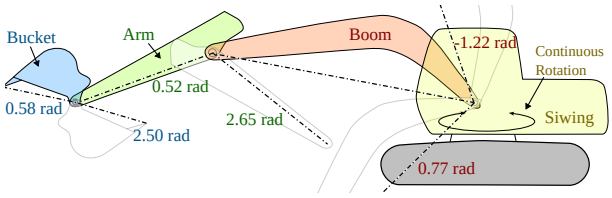


Figure 1. Definition of the joint angles of Boom, Arm, Bucket, and Upper Structure (Swing) in the hydraulic excavator ZX120 (Hitachi Construction Machinery Co.)

to participate in automation development.

This paper describes a method for constructing a dynamics model of a hydraulic excavator based on the above concepts.

2 The physical simulators in OPERA

The physical simulator in OPERA is designed with a focus on the extensibility of functionality on the user side [2]. Other cases of physical simulators have been developed to train construction machinery operators [3, 4]. These are designed for specific applications only, thus providing limited functionality. In contrast, the physical simulators of OPERA provide extensibility so that users can use them in a variety of applications, such as automation and teleoperation.

OPERA provides two physical simulators, OperaSim-AGX [5] and OperaSim-PhysX [6], depending on the user's development objectives and cost requirements. OperaSim-AGX is characterized by its high accuracy in simulating soil deformation. OperaSim-PhysX is inferior to OperaSim-AGX in terms of soil deformation simulation accuracy; however, it has the advantage of not requiring a license fee for using the physics engine, thus reducing the user's development cost. In this paper, we describe a method for simulating the behavior of a hydraulic excavator in OperaSim-PhysX.

3 Dynamics model of the hydraulic excavator

3.1 The system configuration for the hydraulic excavator

Each joint of a hydraulic excavator is controlled by position control. We used the hydraulic excavator ZX120 (Hitachi Construction Machinery Co.). We define the joint angles of Boom, Arm, Bucket, and Upper Structure (hereinafter referred to as Swing) as shown in Figure 1. The system components of the excavator to control the joint angles are shown in Figure 2. The excavator is controlled by repeating the following section 3.1.1 ˜3.2.3

3.1.1 Getting the angle of each joint

The bucket, arm, boom and swing joint angles are measured with absolute encoders. Each joint angle information is sent to the control computer in the excavator.

3.1.2 Controlling valve opening

The Target Angle (command value) and Current Angle (angle obtained from the encoder) are input to each joint controller (Bucket control, Arm control, Boom control, Swing control). Then, the valve openings (valve openings) for the hydraulic valves to control the hydraulic cylinders and motors are calculated.

Each controller consists of a "Dead Time Compensator" and a "PID Controller". The "Dead Time Compensator" compensates for the delay between the time the external computer commands the valve opening degree and when the actuator operates, which is detected as a change in joint angle. We confirmed this delay is 0.35 seconds for each joint through prior measurements. Since the delay causes the controller to be less responsive, we used the Smith method to predict the current angle after the time elapsed.

The PID Controller is provided in a PID [7] package. The valve opening is obtained via the PID Controller from the target angle and the predicted value of the current angle after a time elapse.

3.1.3 Driving each joint

Valve opening information is sent to the main valve system via the solenoid valve system. The oil flow in each drive system ' s hydraulic circuit is controlled according to the valve opening of the main valve system. Then, each joint moves.

3.2 The hydraulic excavator in physical simulation

For simulating the behavior of a hydraulic excavator in a physical simulation, we have constructed the system shown in Figure 3. When a target angle is given to each joint, the torque will be output from each controller (Bucket Control, Arm Control, Boom Control, and Swing Control).

Each controller is composed of a " Dead Time Function " and an " Articulation Body " [8] The Dead Time Function outputs the target angle a certain amount of time ago to simulate the latency of a real excavator. angle a certain amount of time ago to simulate the latency of a real machine. Articulation Body is a Unity component for simulating objects with hierarchical link structures, such as robot arms. It can construct rotating and sliding joints so that angles, angular velocities, and forces can control each joint.

· Articulation Joint Type: Revolute

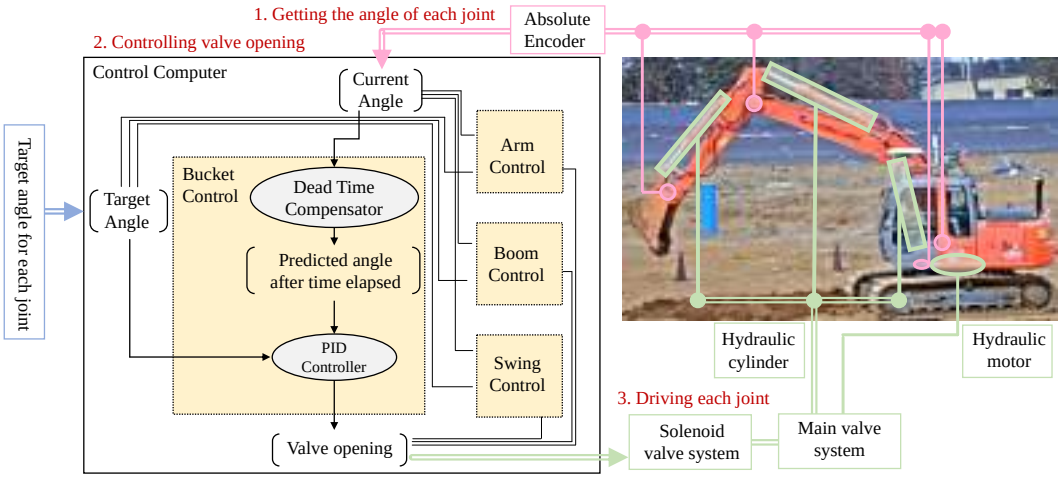


Figure 2. System configuration of the actual autonomous excavator. Each joint is position controlled by repeating the following steps: 1. Getting the angle in each joint, 2. Controlling valve opening, 3. Driving each joint.

Table 1. ZX120 spec about dynamic model

Weight [kg]	Bucket link	386
	Arm link	392
	Boom link	992
	Swing link	6514
	Crawler unit	3816
Link length [m]	Bucket link	1.23
	Arm link	2.519
Max tool force [kN]	Bucket tool force	99
	Arm tool force	65
Max torque [kNm]	Bucket joint	121.77
	Arm joint	236.187
	Boom joint	331.817
	Swing joint	34.3

· Drive Type: Force

Other parameters are described in Sections 3.2.1 ˜3.2.3.

3.2.1 Geometry and weight

The mass of the hydraulic excavator was referenced from the ZX120 model specifications on the catalog [9]. The mass of each component of the excavator is shown in Table 1. The geometry of the hydraulic excavator was scanned using a 3D LiDAR on the excavator shown in Figure 2.

Geometrical data for each link were extracted from the resulting 3D point cloud. The hydraulic excavator is scanned in 3D. The obtained point cloud will be meshed. The mesh is split for each link, and a URDF file is created and imported into the physical simulation. The moment of inertia for each link was derived under the assumption that the mass is uniformly distributed across the obtained geometry. The DH parameters were obtained through measurements of the excavator.

3.2.2 Maximum torque of each joint

We describe how to approximate the maximum output torque for each link, respectively. The maximum torque at the bucket and arm joints is calculated from a value called the ”tool force”. The tool force is the force applied to the blade edge when the bucket or the arm is moved. The method for measuring tool force is defined in ISO 6015:2006 [10]. From those definitions, the maximum torque at each joint of the bucket and arm is obtained to be 121.77 kNm and 236.187 kNm, respectively.

The maximum torque at the boom joint is derived from the force that can be generated by the boom cylinder, as shown in Figure 5. The hydraulic pressure inside the cylinder determines the force exerted by the cylinder. For the ZX120, the maximum pressure applied to the cylinder corresponds to the hydraulic excavator's main relief set pressure. This pressure is regulated by the main relief valve and cannot exceed this limit. According to the specifications of the ZX120, the value is given as $P_b = 34.3$ MPa. The force that can be generated by the cylinder differs during cylinder extension, F_b, and contraction, $\acute{F_b}$. Using the maximum pressure applied to the cylinder P_{bu}, the cylinder's internal diameter D, and the rod's outer diameter d, these forces can be calculated using the following

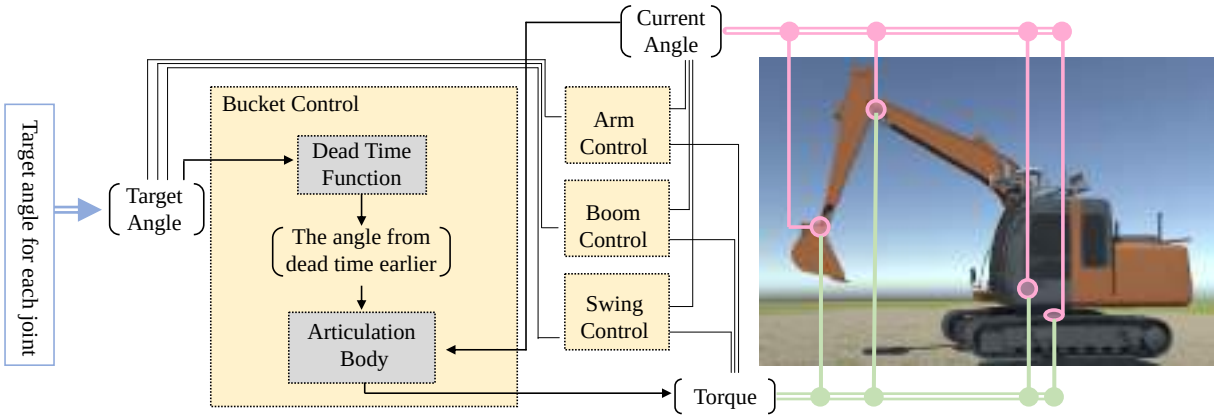


Figure 3. System configuration of the excavator in the physical simulator



Figure 4. Procedure to simulate a excavator in a physical simulator: (a) Obtaining geometry with a 3D scanner, (b) Meshing (c) Constructing and importing a URDF

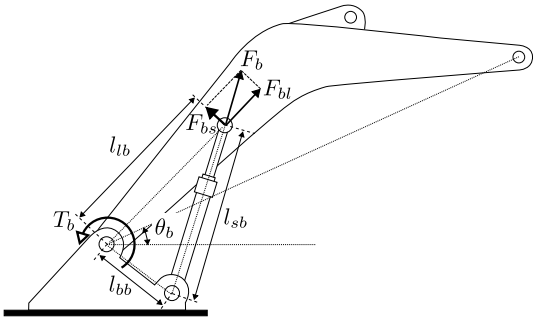


Figure 5. Calculating torque in Boom joint

equations.

$$F_b = 2 * 0.25\pi D^2 * P_{max} \quad (1)$$

$$F_b^{'} = 2 * 0.25\pi (D^2 - d^2) * P_{max} \quad (2)$$

After deriving the force that the cylinder can generate, the rotational torque around the boom link axis is calculated. The rotational torque is generated by the force F_b produced by the cylinder, specifically the component of this force F_{bs} that is perpendicular to the line formed between the boom rotation axis and the rod pin of the boom cylinder, as shown in Figure 5. The relationship between the boom angle θ_b and the rotational torque is shown in Figure 6. The boom's rotation range is from 0.77 [rad] (about 44°) to -1.22 [rad] (about -70°), and the maximum torque occurs at the point where the angle is 0 [rad]. While there are two possible values due to cylinder extension and contraction, the data for the direction in which the boom lifts off the ground (i.e., cylinder extension) is adopted here. The maximum torque is 331.816 kNm.

The maximum torque for the swing axis is taken from the datasheet [9] as 34.3 kNm. The calculation results for the maximum torque are summarized in Table 1.

3.2.3 The gain configuration for torque output

The torque at each joint is calculated as follows.

$$T = K_s(\theta - \theta^*) - K_d(\omega - \omega^*). \quad (3)$$

where, θ and θ^* are the current and target joint angles, respectively, and ω and ω^* are the current and target joint angular velocities, respectively. By adjusting the parameter K_s, which determines the angle gain, and the parameter K_d, which determines the angular velocity gain, the response of the joint to inputs can be modified. The joint torque T is constrained by the Force limit parameter of the Articulation Body. The maximum torque for each joint is applied as the Force limit to ensure that torques exceeding the set values are not output.

4 Estimation of parameters in the excavator dynamics model

We estimate the Stiffness K_s and Damping K_d parameters of the Articulation Body, which compose each joint in a simulation of the excavator. We select values of Stiffness K_s and Damping K_d, which are close to the real machine's response by tuning them. The closeness of the response

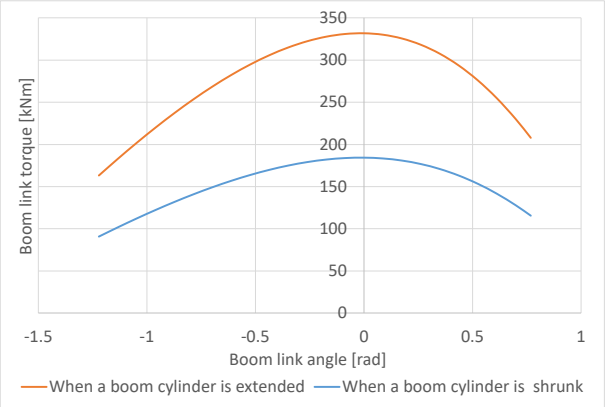


Figure 6. Changes in boom torque with respect to the boom joint angle

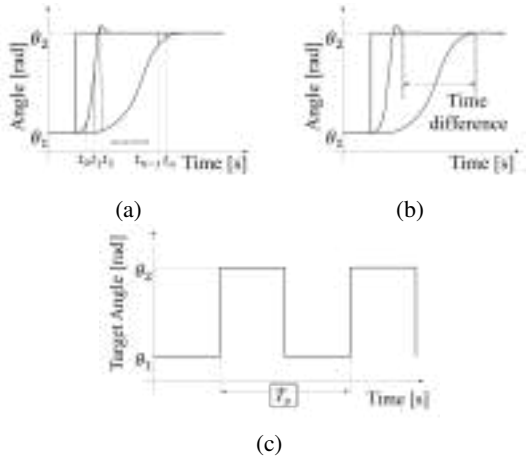


Figure 7. Parameter estimation details

to the actual machine is evaluated in the following two points.

A. **Mean square error of response time change**

Figure 7a shows the mean square of the errors in the response time change. From the Current Angle θ^r of the actual machine and the Current Angle θ^s of the simulator obtained at each sampling time $t_0, t_1, t_2 \ldots t_{n-1}, t_n$, calculate,

$$S = \frac{1}{n}\sum_{i=1}^{n}(\theta^s_{t_i} - \theta^r_{t_i})^2 \tag{4}$$

where n is the number of data samples and S is the mean square error of the time series response. We change the values of K_s and K_d, then select the one for which the mean of squares S is smaller.

B. **Convergence time difference to the target angle**

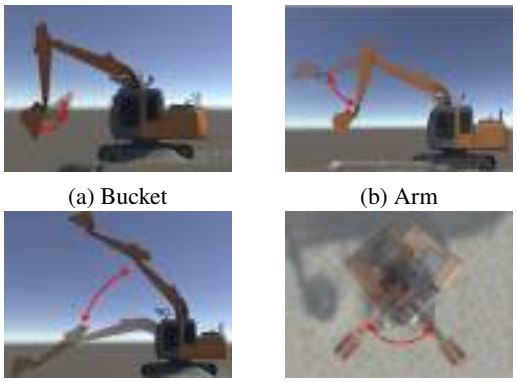


Figure 8. Movement during the parameter estimation at each joint

Table 2. PC specification for simulation

CPU	Intel(R) Core(TM) i9-14900HX 2.20 GHz
GPU	NVIDIA GeForce RTX 4060 Laptop GPU
RAM	32.0 GB
Storage	SSD (NVMe) 1TB

Figure 7b shows the convergence time difference with respect to the target angle. When the convergence time to the target angle in the real machine is T_s^r and the convergence time to the target angle in the simulation is T_s^s, the absolute value of the convergence time difference $|T_s^d|$ is

$$|T_s^d| = |T_s^r - T_s^s|. \tag{5}$$

We select the one with smaller $|T_s^d|$ by tuning the parameters.

For executing the simulation, we used the PC shown in Table 2. Prior to the simulation, we confirmed that the simulation was performed in real time.

4.1 Setting the target angle

The time variation of the target angle is shown in Figure 7c. The θ_1 and θ_2 angles are given alternately at a fixed period T_s.The values of θ_1 and θ_2 and the period T_s at the Bucket, Arm, Boom, and Swing joints are shown in Table 3. The motion of the hydraulic excavator is shown in Figure 8.

4.2 Minimizing the mean square error of response time change

The mean square error S of the time response change when the parameters are changed is shown in Figure 9. In this paper, we describe the results of changing the Arm

Damping K_d \ Stiffness K_s	200000	500000	1000000	2000000	5000000
100000	0.014409	0.045674	0.060357	0.069921	0.07994
200000	**0.00925**	0.019007	0.043298	0.060873	0.07216
500000	0.071661	**0.006424**	**0.011529**	0.033978	0.057349
1000000	0.146884	0.046586	**0.006035**	**0.010427**	0.04163
2000000	0.218495	0.118885	0.046317	**0.00562**	0.018368
	Low			High	[rad²]

(a) Arm

Damping K_d \ Stiffness K_s	20000000	50000000	1E+08	2E+08	5E+08
10000000	0.049675	0.083647	0.103477	0.101429	0.100319
20000000	**0.017519**	0.059383	0.083223	0.103053	0.101587
50000000	0.026298	**0.017482**	0.049627	0.0755	0.101053
1E+08	0.08897	**0.015279**	**0.018205**	0.047905	0.082697
2E+08	0.156272	0.066361	**0.015244**	**0.018245**	0.058392
	Low			High	[rad²]

(b) Boom

Figure 9. Mean squares error between actual and simulation responses

Dumping K_d	Time difference [s]
800000	0.563791
900000	0.390135
1000000	0.194506
1100000	0.094478
1200000	0.263104
1400000	0.620648
1600000	1.099094

(a) Arm

Dumping K_d	Time difference [s]
60000000	0.491356
70000000	0.148872
80000000	0.36404
90000000	0.731205
100000000	1.190947
120000000	1.942082

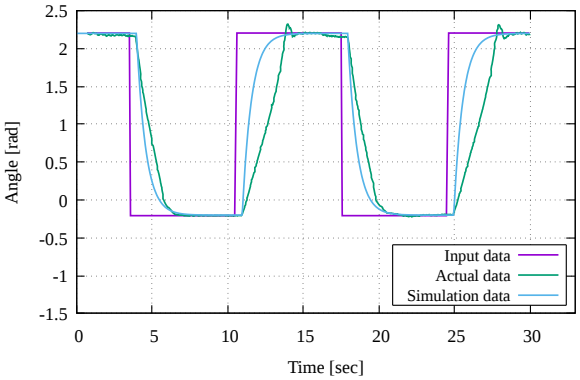
(b) Boom

Figure 10. Convergence time difference between actual and simulation

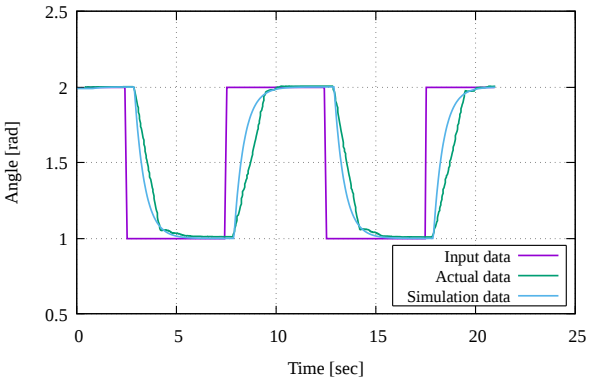
and Boom parameters. The larger value of S is shown in red, and the smaller value is shown in blue. Among the 5 x 5 = 25 samples, 6 samples with small S are shown in bold. The candidate Stiffness K_s and Damping K_d are selected from these samples. Here, a small value of Stiffness K_s is not applied. If K_s is too small, the torque to hold the weight of the link can not be provided. For Arm and Boom, the combination indicated by the red frame in Figure 9 is adopted; for Bucket and Swing, Stiffness K_s and Damping K_d are estimated similarly. The results are shown in Table 4.

4.3 Minimizing convergence time difference

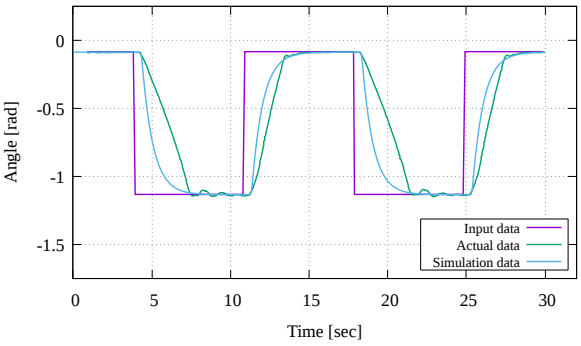
We measured the convergence time difference $|T_s^d|$ when K_d is changed while K_s selected in Section 4.2 is



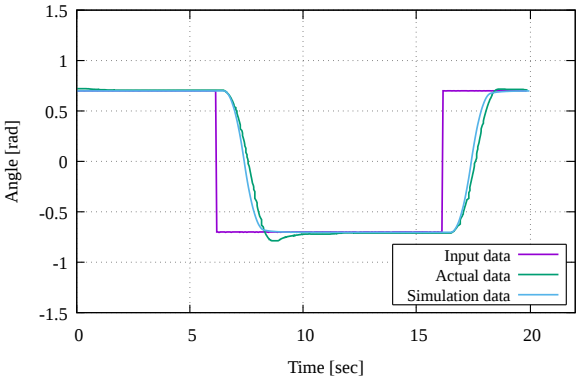
(a) Bucket



(b) Arm



(c) Boom



(d) Swing

Figure 11. The results of parameter estimation

Table 3. Detail of the target angle: θ_1 and θ_2 are given alternately at period T_s

	θ_1	θ_2	T_s
bucket	-0.2 rad	2.2 rad	7.0 s
arm	1.0 rad	2.0 rad	5.0 s
boom	-0.087 rad	-1.135	7.0 s
swing	0.785 rad	-0.785	10.0 s

Table 4. The result of minimizing the mean square error of response time change

	Bucket	Arm	Boom	Swing
Stiffness K_s	2000000	2000000	2E+8	2000000
Damping K_d	2000000	2000000	2E+8	1000000

fixed. The K_d at which $|T_s^d|$ is the smallest is selected. The results are shown in Figure 10 for Arm and Boom. The red box in the figure shows the K_d with the smallest $|T_s^d|$ among the trials. K_d for Bucket and Swing were adjusted similarly, as shown in Table 5.

Table 5. The result of minimizing convergence time difference

	Bucket	Arm	Boom	Swing
Stiffness K_s	2000000	2000000	2E+8	2000000
Damping K_d	1200000	1100000	70000000	900000

4.4 Adjustment result

The response of each joint to the target angle shown in Figure 8 is shown in Figure 11. S was obtained from Figure 4 for bucket, arm, boom, and swing, respectively: 0.179746, 0.009057, 0.034118, and 0.003911 [rad^2].

5 Evaluation of dynamics model by digging operation

We compare the behavior of the excavator during excavation by the actual machine with the behavior of the excavator during excavation within the physical simulation.

5.1 Evaluation requirements

The motion during digging is shown in Figure 12. Excavation is performed in the following order.

1. Turning the Swing to begin excavating the ground
2. Stabbing the bucket into the ground
3. Scooping and lifting



Figure 12. Digging motion

This evaluation was conducted in an experimental field owned by the Public Works Research Institute (PWRI) in Japan. The ground surface consists mainly of clayey soil (loam), which is flat.

5.2 The result

The following figure shows the time variation of the current angle in the actual machine (Actual data) and the angle in the simulation (Simulation data) when the target angle (Input data) is given to Bucket, Arm, Boom, and Swing respectively in Figure 13. data).

The mean square error S of the time response was calculated for each joint according to Formula (4). The results were 6.258×10^{-6}, 6.264×10^{-6}, 6.27×10^{-6} and 6.27×10^{-6} for bucket, arm, boom and swing respectively. The time response is shown in Figure 13.

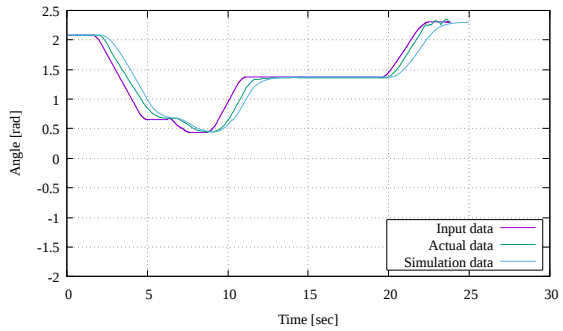
6 Conclution

This paper explained the method for constructing the dynamic model of a hydraulic excavator in the physical simulator of the open platform OPERA, developed for the automation of construction machinery and automated construction technology. Specifically, the dynamic model was configured to ensure that the hydraulic excavator in the simulation operates similarly to the actual machine, focusing on three aspects: digging force, and the response to inputs at each joint of the working mechanism.

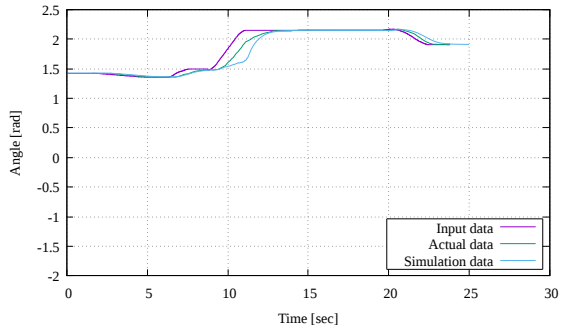
The model was adjusted so that the behaviors of the swing, boom, arm, and bucket rotational axes in OPERA matched those of the actual machine. As a result, it was confirmed that the model could replicate the behavior of the actual machine with an mean squared error of approximately 6.258×10^{-6} to 6.27×10^{-6} [rad^2] for each axis.

The following is the future works. We construct a dynamics model which more closely simulates the behavior of a real excavator.

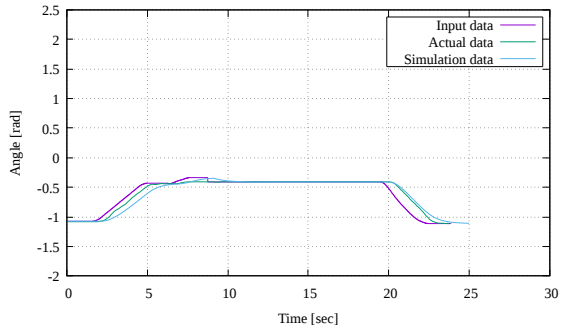
- We need to confirm the validity of the proposed method on excavators other than ZX120.



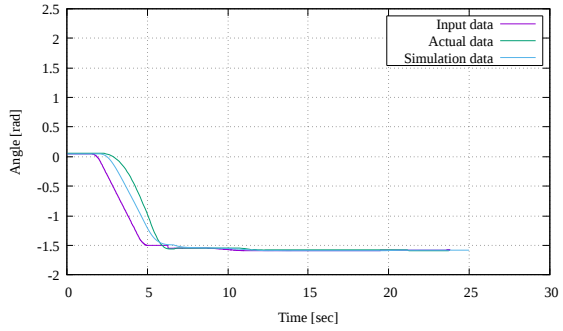
(a) Bucket



(b) Arm



(c) Boom



(d) Swing

Figure 13. Evaluation result

- We will experiment whether advanced control methods other than PID control, such as adaptive control, can be applied to excavators and whether their behavior can be simulated.
- We will devise a method of constructing a dynamics model for a machine with aging deterioration. we include in the dynamics model factors such as the decrease in power output of the excavator.

References

[1] G. Yamamuchi, d. Endo, H. Suzuki, and T. Hashimoto. Proposal of an open platform for autonomous construction machinery development. In *41th International Symposium on Automation and Robotics in Construction (ISARC 2023)*, 2023.

[2] D. Endo, G. Yamauchi, H. Suzuki, and T. Hashimoto. Research on an open source physical simulator for the autonomous construction machinery development. In *41th International Symposium on Automation and Robotics in Construction (ISARC 2024)*, 2024.

[3] CAT®Simulator Training. `https://www.cat.com/en_US/support/cat-training/simulators.html`.

[4] Advantages with the komatsu simulators. `https://komatsusimulators.com/advantages/`.

[5] OperaSim-AGX. `https://github.com/pwri-opera/OperaSim-AGX`, .

[6] OperaSim-PhysX. `https://github.com/pwri-opera/OperaSim-PhysX.git`, .

[7] pid. `https://wiki.ros.org/pid`.

[8] Unity Manual: Articulation Body component reference. `https://docs.unity3d.com/2020.3/Documentation/Manual/class-ArticulationBody.html`.

[9] Hitachi Construction Machinery Co. ZAXIS 110/ZAXIS 120 SERIES, n.d. Manufacturer's catalog.

[10] *Earth-moving machinery – Hydraulic excavator and backhoe loader boom and arm – Forces, and movements*. International Organization for Standardization, Geneva, Switzerland, 2006. ISO 6015:2006.

Personalized Emotion-Adaptive Robot Control Strategy for Human-Robot Collaboration in Construction

Francis Back[1], Leyang Wen[1], Gunwoo Yong[1], and SangHyun Lee[1]

[1]Department of Civil and Environmental Engineering, University of Michigan, USA
fbaek@umich.edu, wenleyan@umich.edu, gwyong@umich.edu, and shdpm@umich.edu

Abstract

While human-robot collaboration (HRC) is anticipated to improve construction productivity and safety, it can become crucial to manage emotional responses of collaborating human workers (co-workers) during HRC due to their critical impacts on co-workers' performance and human-robot cohesion. A potential approach involves controlling robots in ways that can adapt to potentially different emotional responses of co-workers to foster desired responses and productivity in human-robot teams. However, we still lack an effective method as existing robot control strategies tend to prioritize productivity and safety alone or consider co-workers' responses but take standardized actions across different co-workers. In this study, we propose a personalized emotion-adaptive robot control strategy using model-based reinforcement learning (RL), which aims for robots to learn each co-worker's emotional responses over interaction to plan and take actions for fostering desired responses while pursuing high productivity in human-robot teams. The effectiveness of the proposed robot control strategy was demonstrated in a virtual environment by simulating HRC with the emotional responses of 20 co-workers for a construction task. Results showed that a robot with the proposed strategy could promote desired emotional responses like happiness in all 20 co-workers for approximately 37 minutes during a 1-hour bricklaying task. While findings demonstrate the potential of the proposed robot control strategy as a way for robots to adapt to co-workers' emotional responses, further analysis will be anticipated to minimize potential trade-offs in productivity in human-robot teams that may arise from considering co-workers' responses.

Keywords –

Human-Robot Collaboration, Reinforcement Learning, Emotional Response

1 Introduction

The construction industry has suffered from stagnant productivity growth compared to other industries [1]. According to the McKinsey Global Institute, the productivity (real gross value added per labor hour) of the construction industry has grown 1% for the past two decades since 1995, while the average annual growth by the total global economy was 2.8% [2]. Although construction productivity can be considered mainly relying on labor power [3], labor power can be negatively affected by other ongoing issues in the industry, including high safety risks, the aging workforce, and the labor shortage [4-6]. The stagnant growth of productivity has remained a major challenge.

Human-robot collaboration (HRC) has recently gained attention in the construction industry, envisioned as a new paradigm shift toward co-robotic construction [7]. Recent advancements in robotic technology have introduced collaborative robots, an evolved form of conventional industrial robots, enabling collaboration with humans by combining the strengths of humans and robots [8]. For example, robots can take over physically demanding, dangerous, and repetitive tasks for human workers to focus on dexterous, problem-solving, and decision-making tasks [8]. HRC is anticipated to improve construction productivity by reducing the primary dependence on workers' labor power and increasing the level of robotic automation [9].

However, HRC is in its early stages in construction. The extent of its impact on construction productivity has not been thoroughly examined [9]. A potentially crucial factor for productive HRC is emotional response of collaborating human workers (co-workers). Emotions can be defined as mental states, associated with degrees from displeasure to pleasure and from relaxation to excitement, such as excitement, relaxation, boredom, and fear [10]. Studies have demonstrated that emotions can cause critical impacts on worker performance by affecting decision-making, problem-solving, and behaviors [11]. Additionally, studies in some fields, like social robotics, have identified that emotions can be largely influential to the human-robot cohesion by influencing acceptance and trust toward robots [12]. Within the unprecedented co-robotic work environment, the importance of emotions can become more

considerable than before to achieve the anticipated productivity growth by the human-robot teams.

Efforts have identified the potential variations in co-workers' emotional responses during HRC. Our prior work demonstrated through a lab experiment that the emotional responses of co-workers could be significantly influenced during a simulated bricklaying task with a real robot [13]. Major results showed that while every co-worker's emotional response was affected, almost every co-worker exhibited statistically different responses, which were also estimated to affect productivity in human-robot teams. These findings present that considering co-workers' emotional responses, which may vary among individuals, can become essential for fostering desired responses that can enhance both their performance as well as cohesion with robots.

A potential approach can involve controlling robots in a way that can consider each co-worker's emotional responses. However, the consideration of emotional responses has not been a major focus in robot control strategies for HRC, particularly in construction, where performance-oriented factors like productivity and safety remain the primary objectives [14,15]. While a few studies have presented the potential necessity of robot control that considers human responses [16], robots typically take standardized actions for all individuals, which may be ineffective for co-workers with different responses [13]. As such, we still lack an effective robot control strategy for robots to consider co-workers' potentially different emotional responses while pursuing higher productivity in human-robot teams.

In this study, we propose a personalized emotion-adaptive robot control strategy that aims for robots to learn each co-worker's emotional response during interactions, even with whom they encounter for the first time. This strategy allows robots to take actions that can elicit desired emotional responses (e.g., happiness and excitement), which can enhance co-workers' performance and cohesion with robots, while pursuing high productivity in human-robot teams. The proposed robot control strategy is based on reinforcement learning (RL) algorithm, an effective branch of machine learning for adaptive robot control [17]. The effectiveness of the proposed robot control strategy was evaluated through a simulated HRC for a construction task in a virtual environment, where the robot collaborated with virtually simulated 20 co-workers, whose emotional responses were identified in our prior study [13].

This study demonstrates the potential effectiveness of using reinforcement learning (RL) for an adaptive robot control strategy to promote desired emotional responses in each co-worker, including those encountered for the first time. At the same time, the findings present the need for further analysis of robot control to balance emotional responses and productivity in human-robot teams, as trade-offs between these factors may arise.

2 Model-based Reinforcement Learning for Personalized Emotion-Adaptive Robot Control Strategy

The proposed robot control strategy relies on a model-based RL, an effective branch of RL when interacting with complicated environments like human responses [18]. Model-based RL allows the robot to learn the mechanisms of co-workers' emotional responses to its actions through modeling over interactions. This enables the robot to plan and take actions for fostering desired emotional responses in co-workers while pursuing high productivity in human-robot teams. The proposed robot control strategy guides the robot to operate iteratively by 1) taking an action during HRC for a construction task (Figure 1a), 2) learning the co-worker's emotional responses (Figure 1b), and 3) planning the next action to promote desired co-worker's emotional responses and highest possible productivity in human-robot teams (Figure 1c). Notably, it is assumed that the robot can monitor co-workers' emotional responses through wearable sensors, such as brainwave analysis using an EEG headset.

2.1 Desired Emotional Response

Emotional response can be represented using a bipolar dimensional emotion model [19], which reliably and sufficiently captures emotional states (e.g., excited and bored) by estimating emotional status based on two key dimensions (Eq. 1): valence (ranging from displeasure to pleasure) and arousal (ranging from relaxation to excitement). Among various emotional

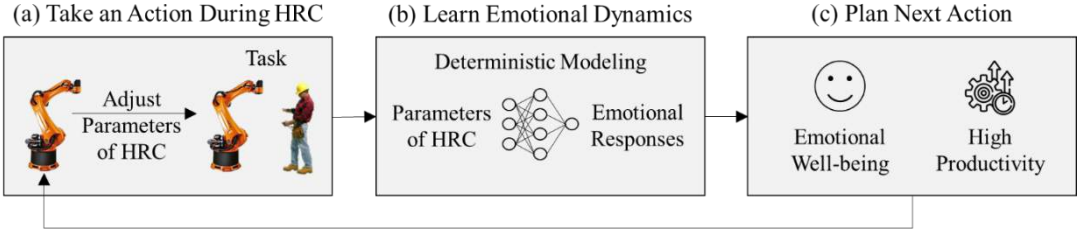


Figure 1. Overview of the proposed robot control strategy.

states, including calmness, anxiety, and boredom, this study considered positive emotions, such as happiness and excitement, as desired. The positive emotions are represented by positive valence and arousal, known to promote cohesive human-robot relationships and help co-workers with better performance while staying alert to potential safety risks in ever-changing, unstructured, and hazardous construction work environments [11,12,20].

$$e = \{valence, arousal\} \quad \text{(Eq. 1)}$$

2.2 Take an Action

In RL, the actions of a robot refer to changes in the state of the environment, which encompasses everything that the robot observes and responds to [17]. In this study, the environment is represented by the states (s_t), consisting of 1) parameters of HRC (p_t), such as robot's movement speed, which can affect co-workers' emotional responses, and 2) the corresponding emotional responses of a co-worker (e_t) at a time t (Eq. 2.1). Actions (a_t) of a robot refer to the adjustments in parameters of HRC (p_{t+1}) during a construction task, such as an updated movement speed after slowing down (Eq. 2.2). The actions directly adjust the parameters of HRC and can cause variations in emotional responses of a co-worker. Over the interaction, the robot can collect the set (D) of observations, consisting of current states (s_t), robot's action (a_t), and corresponding changes in environmental states (s_{t+1}) (Eq. 2.3). It is noteworthy that the size of D expands whenever the robot takes an action.

$$s_t = \{p_t, e_t\} \quad \text{(Eq. 2.1)}$$

$$a_t = p_{t+1} \quad \text{(Eq. 2.2)}$$

$$D = \{(s_t, a_t, s_{t+1})_{t=1\sim T}\} \quad \text{(Eq. 2.3)}$$

$$where\ s_t\ and\ a_t\ represent\ the\ state\ and\ action,\ respectively,\ at\ each\ time\ step\ t,\ ranging\ from\ the\ beginning\ (t=1)\ to\ the\ current\ time\ step\ (t=T)$$

2.3 Learn Co-worker's Emotional Responses

In this study, the proposed strategy is specifically tailored to operate within a deterministic Markov decision process framework, where the environment's transitions are deterministic given the current state and action. The set of observations, D, enables the robot to learn the deterministic state transition model $T(s_{t+1}|s_t, a_t)$ (Eq. 3), which predicts how the state (s_t) evolves based on the robot's action (a_t). As actions (a_t) correspond to the updated parameters of HRC (p_{t+1}), the learning of the transition model solely focuses on $f_{Emotion}$, which represents co-workers' emotional responses to the robot's actions. The structure of $f_{Emotion}$ can be flexibly specified when previously identified relationships exist between the robot's actions and co-workers' emotional responses. For example, in our prior study, we identified multiple linear regression relationships between robot actions, such as movement speed, and co-workers' emotional responses [13]. During interactions, the robot learns coefficients of $f_{Emotion}$ tailored to the specific co-worker using the least square optimization (Eq. 3).

$$T(s_{t+1}|s_t, a_t) = \{a_t, f_{Emotion}(a_t)\} \quad \text{(Eq. 3)}$$
$$where\ s_{t+1} = T(s_t, a_t) = \{p_{t+1}, e_{t+1}\},\ p_{t+1} = a_t, and$$
$$f_{Emotion}(a_t)\ minimizes \sum_t \|e_{t+1} - f_{Emotion}(a_t)\|^2$$

2.4 Plan the Next Action

While the robot's actions can affect co-worker's emotional responses, they also can affect the productivity of human-robot teams. For example, the robot's faster movement speed can enhance productivity in human-robot teams. We assume the robot can estimate the productivity of human-robot teams in terms of the parameters of HRC ($f_{Productivity}$) (Eq. 4). Based on the estimations, $f_{Emotion}$ and $f_{Productivity}$, the robot uses its policy to select the next action a_t^*, aiming to maximize the cumulative reward $r(s_t, a_t)$ (Eq. 4), which can lead to the highest possible productivity in human-robot teams among the candidate actions that can evoke co-workers' emotional well-being. Notably, while traditional RL framework often uses value functions to represent expected cumulative rewards, this study focuses on direct optimization of immediate rewards $r(s_t, a_t)$ using the transition model $f_{Emotion}$. Using a value function to predict cumulative rewards over longer term can be challenging in the context of human response-adaptive robot control, as human responses may change over time even in short periods (e.g., seconds). In this context, we propose the robot focus on immediate rewards, which can allow effective respond to emotional responses. However, for this planning approach to be effective, the robot must first observe co-workers' emotional responses across a wide range of scenarios to accurately learn the transition model. To achieve this, the robot incorporates an exploration strategy, randomly selecting a percentage of its actions to gather diverse data on co-workers' responses. Initially, 100% of the robot's actions are random until collecting 48 observations of co-worker's emotional responses, and this percentage decreased to 20%. This empirically determined strategy balances exploration and exploitation, enabling the robot to refine its transition model and improve action planning over the

interaction.

$$a_t^* = \underset{a}{\operatorname{argmax}} \sum_a r(s_t, a_t) \quad \text{(Eq. 4)}$$

$$where\ r(s_t, a_t) = \begin{cases} f_{Productivity}(p_{t+1} = a_t) & if\ f_{Emotion}(a_t) > 0 \\ 0 & Otherwise \end{cases}$$

3 Simulated HRC for a Construction Task in a Virtual Environment

In our prior study, we analyzed the emotional responses of 20 people, consisting of 13 men and 7 women, while they collaborated with a real robot in a lab environment for a bricklaying task, which can be considered a representative example of distributing tasks between humans and robots in construction [13]; robots take over physically demanding tasks (e.g., lifting bricks) while workers focus on the dexterous tasks like finishing. Participants were recruited from students with various academic backgrounds (mean age: 29; standard deviation of age: 3.5; minimum age: 23; and maximum age: 36), including 12 from civil and environmental engineering, some with construction field experience, 4 from other engineering departments like electrical and mechanical, and 4 from non-engineering departments like music and social science. Their varying levels of construction experience, technical proficiency, and exposure to robotic technologies were expected to exemplify a wide range of emotional responses during HRC, also anticipated among construction workers. As an early attempt to demonstrate the potential effectiveness of the proposed robot control strategy in adapting to wide spectrum of different emotional responses, we adopted the emotional responses of the student participants, instead of exploring and simulating emotional responses of construction workers, whose responses might be affected by occupational biases, such as concerns of job loss and replacement by robots [21].

In this context, we virtually simulated a robot to collaborate with each of the 20 co-workers to conduct a bricklaying task for one hour, a plausible work duration before a break. We replicated the lab environment from the prior study [13] within a virtual environment, where parameters of HRC and emotional responses were computationally simulated. The bricklaying task was conducted by a team of one co-worker and one robot (Figure 2); the robot delivered bricks for the co-worker to lay them in a line. Notably, the simulated bricklaying task conducts some portion of masonry work to demonstrate the effectiveness of the proposed robot control strategy during a representative task distribution between humans and robots, rather than replicating the entire masonry process, including scaffold installation, mortar mixing, and brick cutting.

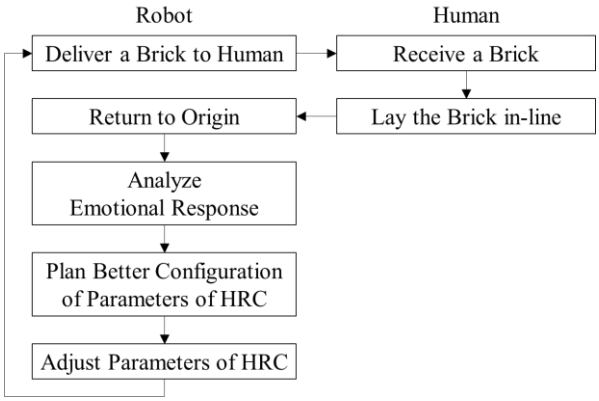


Figure 2. Workflow of the bricklaying task.

The robot was allowed to take actions over five parameters of HRC, which were identified to affect co-workers' emotional responses [13]: 1) robot's movement speed; 2) robot's arm swing speed; 3) proximity between humans and robots; 4) level of autonomy; and 5) leader of collaboration. All the actions were to be made within the specified ranges based on the robot's capability and physical dimension of the lab environment [13]: movement speed from 27.8 cm/s to 143.8 cm/s; arm swing speed from 23.8 cm/s to 109.1 cm/s; proximity as either direct (100 cm) or indirect (200 cm); level of autonomy as either manual or auto; and leader as either human or robot.

In response to the robot's actions, the emotional responses of 20 co-workers were virtually simulated by using two regression models, one for valence and another for arousal, which were obtained from the prior study (Eq. 5.1) [13]. Given robot's actions, these regression models estimate the corresponding valence and arousal of a co-worker at a given time (sec) since the start of HRC. While the robot is aware of these regression models as the $f_{Emotion}$ (Eq. 3), it does not know the coefficients specific to any co-worker but should learn the coefficients tailored to each co-worker through interactions. It is assumed that the robot observes a co-worker's valence and arousal by analyzing EEG signals collected via EEG headset.

Notably, the regression models (Eq. 5.1) were obtained by analyzing EEG signals with noise reduction techniques, including wavelet independent component analysis, which have been demonstrated to be effective in alleviating artifacts—such as eye-blinks, heart rate, and muscle movement—common sources of noise in EEG signals, even in onsite construction settings. However, despite noise reduction, EEG signals can still be affected by noise. In this regard, the estimates for each co-worker were adjusted using means, standard deviations, and noise parameters derived from EEG-based valence and arousal measurements in the prior

study (Eq. 5.2) [13].

$$y_i = c_1 x_1^2 + c_2 x_1 + c_3 x_2^2 + c_4 x_2 + c_5 x_3 + c_6 x_4 + c_7 x_5 + c_8(Time) + \beta \quad \text{(Eq. 5.1)}$$

$$where\ y_i = \begin{cases} Valence \\ Arousal' \end{cases}$$

$$x_j = \begin{cases} Move\ Speed \\ Arm\ Speed \\ Proximity \\ Level\ of\ Autonomy \\ Leader\ of\ Collaboration \end{cases}, \beta = constant,$$

$$and\ Time\ is\ the\ elapsed\ time\ of\ HRC\ (sec)\ since\ the\ start\ of\ HRC$$

$$\tilde{y}_i = \sigma_i y_i + \mu_i + \epsilon_i \quad \text{(Eq. 5.2)}$$

$$where\ \epsilon_i \sim N(0, \delta_i^2),\ and\ \sigma_i, \mu_i, and\ \epsilon_i\ denote\ standard\ deviation, mean, and\ noise\ of\ y_i$$

The robot could estimate productivity in human-robot teams according to its actions during the bricklaying task (i.e., $f_{Productivity}$ in Eq. 4) by following the method used in the prior study [13]. The estimated productivity was represented by the number of bricks laid for one hour by a human-robot team in terms of the five parameters of HRC. Specifically, the bricklaying times for one robot and one human based on five HRC parameters were calculated (Eq. 6.1 and 6.2) and then aggregated into a human-robot team's bricklaying time (Eq. 6.3), which was subsequently converted into the number of bricks for one human-robot team-hour (Eq. 6.4).

$$Time\ (sec)\ for\ a\ robot\ to\ lay\ a\ brick = Body\ Traverse\ Time + Arm\ Traverse\ Time + Delay\ by\ Manual\ Control = \frac{Body\ Travel\ Distance}{Movement\ Speed} + \frac{Arm\ Travel\ Distance}{Arm\ Swing\ Speed} + 1(Manual) * \alpha \quad \text{(Eq. 6.1)}$$

$$where\ Body\ Travel\ Distance\ is\ 6.1m,\ Arm\ Travel\ Distance\ is\ 2.4m,\ and\ \alpha\ is\ delay\ of\ 3\ seconds\ by\ the\ manual\ control$$

$$Time\ (sec)\ for\ a\ human\ to\ lay\ a\ brick = Bricklaying\ Time + Delays\ by\ Indirect\ Proximity\ and\ Human\ Leader = Bricklaying\ Time + 1(Human\ Leader) * \beta + 1(Indirect) * \gamma \quad \text{(Eq. 6.2)}$$

$$where\ Bricklaying\ Time\ is\ 5\ seconds,\ \beta\ is\ delay\ of\ 1\ sec\ by\ human\ leader,\ and\ \gamma\ is\ delay\ of\ 2\ sec\ by\ the\ indirect\ proximity$$

$$Time\ (sec)\ for\ a\ Human–Robot\ team\ to\ lay\ a\ brick = (Eq.\ 5.1) + (Eq.\ 5.2) \quad \text{(Eq. 6.3)}$$

$$Estimated\ Team\ Productivity = \left\{\frac{1\ Brick}{(Eq.\ 6.3)\ Team–sec}\right\} \times \left(\frac{3{,}600\ Team–sec}{Team–hour}\right) = \frac{3{,}600}{Eq.\ 5.3}\ Bricks\ per\ Team–hour \quad \text{(Eq. 6.4)}$$

We compared four types of robot control strategies (Table 1). The first type only prioritized productivity without considering co-workers' emotional responses, representing the currently predominant approach of robot control during HRC. For example, the robot would maximize its movement speed for high productivity. The second type employed a standardized planning approach, which takes uniform actions across different co-workers. Specifically, the robot first identified a set of actions that were effective in evoking positive emotional responses (e.g., positive valence and arousal) and high productivity for a specific co-worker. This co-worker was selected because the emotional responses were the most similar to those of as many other co-workers as possible. The identified actions were then applied uniformly to all other co-workers. The third type is the proposed personalized emotion-adaptive robot control strategy, which would allow the robot to learn the emotional responses of each co-worker, leading it to take actions for fostering positive emotional responses while pursuing high productivity.

Table 1. Four Types of Robot Control Strategies.

Four Types		Consistency	Objectives
1	Productivity-prioritized (Represent Practice)	Same for All People	• Maximum Productivity
2	Standardized Emotion-Adaptive (Represent Other Studies)	Same for All People	• Maximum Productivity • Positive Emotions
3	Personalized Emotion-Adaptive (Proposed Behavior)	Personalized	
4	Fully Emotion-Aware (Ground Truth)	Personalized	

The fourth type represents the theoretical optimum (ground truth) when a robot is fully aware of the emotional responses of co-workers (i.e., regression models of Eq. 4) even before encountering them. This type of robot control may not be practical but can provide insights into how the robot can behave if it is fully aware of co-worker's emotional responses as an ideal case.

We specifically compared the four types of robot control strategies in terms of 1) the number of virtual co-workers who exhibited the desired emotional responses during the 1 team-hour of bricklaying, 2) the average duration of the desired emotional responses across the 20 co-workers, and 3) the average of the number of bricks laid for 1 team-hour. Notably, the robot with the proposed control strategy spent 15 minutes with each co-worker prior to the bricklaying task, as this time was required to initially explore and understand the co-workers' emotional responses.

4 Results and Discussion

The proposed personalized emotion-adaptive robot control strategy could promote the desired positive emotional responses for all 20 virtual co-workers (100%) for around 37 minutes during the 1 hour of simulated bricklaying task (3rd type in Table 2). Every co-worker could experience positive emotions for at least 20 minutes. The results present the potential effectiveness of the proposed robot control strategy in promoting desired emotional responses in co-workers, which can be critical to enhance co-workers' performance and human-robot cohesion for better productivity in human-robot teams.

The comparison of these results with those of the first and second types of robot control strategies is particularly noteworthy. These strategies could only promote positive emotional responses in 5 and 7 virtual co-workers, respectively, and the average durations of positive emotions across 20 co-workers were 15 minutes and 21 minutes, respectively. Although the strategies were effective for some virtual co-workers to exhibit positive emotions during the entire 60 minutes of bricklaying, they were not effective at all for the majority of virtual co-workers. These results present that promoting desired emotional responses in co-workers can require deliberate efforts to consider each co-worker's potentially different emotional responses. Otherwise, prolonged HRC with negative emotional status (e.g., anxiousness and fear) can cause co-workers to have reduced trust in robots, decreased collaboration efficiency, and potentially lead to psychological discomfort or stress, ultimately affecting the overall productivity and cohesiveness of human-robot teams [11,22].

It is noteworthy that the proposed robot control strategy enables the robot to execute personally different actions by considering individuals' emotional responses. As shown in Figure 3, the robot's actions were quite different to different co-workers. While emotional responses can be different among individuals, the proposed robot control strategy holds potential to consider each co-worker's emotional responses for fostering desired responses during HRC.

While promising, however, the proposed robot control strategy still has a room for improvement. The fourth type of robot control strategy shows that all 20 virtual co-workers could theoretically exhibit the positive emotions for around 58 minutes during the bricklaying task, while laying a greater number of bricks. These results present that advancing the proposed robot control strategy to promote desired emotional responses in co-workers for longer duration while maintaining higher productivity can be a direction for improvement.

Meanwhile, although the first type of robot control strategy was not effective in promoting positive emotions in co-workers, it could achieve higher productivity in human-robot teams compared to the proposed and the ideal robot control strategies (types 3 and 4 in Table 2). This can be interpreted that considering and promoting co-workers' desired responses may sacrifice the productivity in human-robot teams.

Such a potential loss in productivity in human-robot

Table 2. Emotional Responses and Productivity by the Four Types of Robot Control Strategies.

Four Types	# of Co-workers w/ Positive Emotions	Avg. Duration of Positive Emotions	Est. Productivity in Human-Robot Teams
1	5 out of 20 (25%)	15 mins (Min: 0, Max: 60)	314 bricks per one team-hour
2	7 out of 20 (35%)	21 mins (Min: 0, Max: 60)	256 bricks per one team-hour
3	20 out of 20 (100%)	37 mins (Min: 20, Max: 60)	244 bricks per one team-hour
4	20 out of 20 (100%)	58 mins (Min: 20, Max: 60)	286 bricks per one team-hour

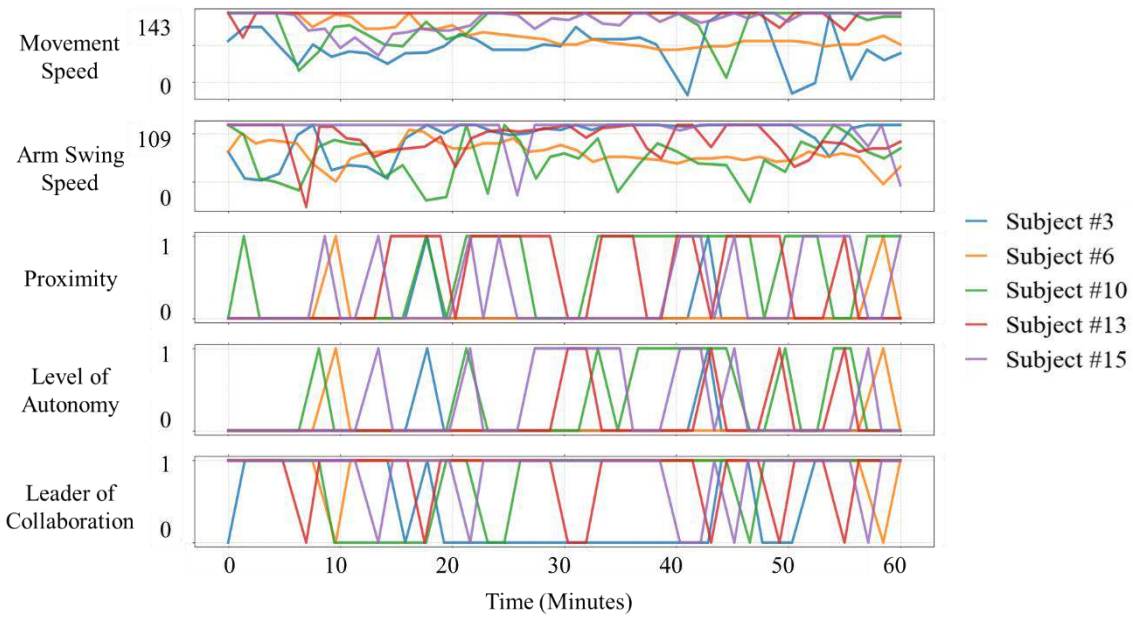


Figure 3. Robot's actions to five different subjects during the bricklaying task.

teams while promoting desired emotional responses in co-workers can raise ethical questions. For example, to what extent is a productivity loss acceptable to ensure co-workers' emotional well-being? Conversely, to what extent can a sacrifice in co-workers' emotional responses be justified to enhance productivity? This presents the clear need for further analysis regarding how we can balance emotional well-being and productivity, particularly in the construction industry, where productivity is crucial.

Grounded on the findings of this study, our future study can be two-fold: 1) demonstrate the effectiveness of the personalized adaptive robot control strategy during the interaction with actual co-workers, as this study simulated the responses of co-workers using the regression models; and 2) examine its effectiveness in other types of HRC in construction, which can involve HRC for different construction tasks, HRC with different construction robots, and HRC without direct physical interactions between humans and robots. The findings of this study are expected to provide a solid foundation for future research, aiming to achieve the emotional well-being of co-workers as well as more productive HRC in construction.

5 Conclusion

Emotions are crucial human responses, affecting human performance and human-robot cohesion, which can be potentially important for the success of HRC in construction. In this study, we propose a robot control strategy that aims to promote desired emotional responses in each co-worker while pursuing the high productivity of human-robot teams. We demonstrated the effectiveness of the proposed robot control strategy by comparing it with predominant strategies of robot control, such as those prioritizing productivity or standardized across people, through a simulated bricklaying task in a virtual environment with 20 people as co-workers. The proposed robot control strategy could lead to the desired positive emotions in all 20 co-workers for around 37 minutes during the 1-hour bricklaying task. Our results present the proposed personalized emotion-adaptive robot control strategy as an effective strategy to aim for promoting desired co-workers' emotional responses, which can contribute to enhancing co-workers' performance and human-robot cohesion. At the same time, the results raise a major question about how we can minimize the potential sacrifice in productivity in human-robot teams, which may arise as we consider co-workers' responses. While future research are expected, aiming to demonstrate the effectiveness of the proposed robot control strategy with actual co-workers for different construction tasks, the findings of this study can serve as a solid foundation to achieve more productive and cohesive human-robot teams in construction.

References

[1] L. Sveikauskas, S. Rowe, J.D. Mildenberger, Measuring productivity growth in construction, Monthly Lab. Rev. 141 (2018) 1.

[2] McKinsey Global Institute, Reinventing Construction: A Route To Higher Productivity,

Washington, D. C., 2017. http://www.mckinsey.com/industries/capital-projects-and-infrastructure/our-insights/reinventing-construction-through-a-productivity-revolution (accessed February 17, 2022).
[3] A.M. Jarkas, C.G. Bitar, Factors affecting construction labor productivity in Kuwait, Journal of Construction Engineering and Management 138 (7) (2012) 811-820.
[4] R. Sacks, A. Perlman, R. Barak, Construction safety training using immersive virtual reality, Construction Management and Economics 31 (9) (2013) 1005-1017.
[5] N.V. Schwatka, L.M. Butler, J.R. Rosecrance, An aging workforce and injury in the construction industry, Epidemiologic reviews 34 (1) (2012) 156-167.
[6] Skilled Labor Shortage: Persistent Challenge, Modern Contractor Solutions. https://mcsmag.com/skilled-labor-shortage-persistent-challenge/ (accessed February 3, 2022).
[7] C.-J. Liang, X. Wang, V.R. Kamat, C.C. Menassa, Human–robot collaboration in construction: Classification and research trends, Journal of Construction Engineering and Management 147 (10) (2021) 03121006.
[8] A. Vysocky, P. Novak, Human-Robot collaboration in industry, MM Science Journal 9 (2) (2016) 903-906.
[9] M. Wu, J.-R. Lin, X.-H. Zhang, How human-robot collaboration impacts construction productivity: An agent-based multi-fidelity modeling approach, Advanced Engineering Informatics 52 (2022) 101589.
[10] E.A. Kensinger, Remembering emotional experiences: The contribution of valence and arousal, Reviews in the Neurosciences 15 (4) (2004) 241-252.
[11] G.A. Fordjour, A.P. Chan, P. Amoah, A.A. Fordjour, Exploring the effects of occupational psychological disorders on construction employees and the construction industry, Occupational diseases and environmental medicine 8 (01) (2020) 1.
[12] S.H.-W. Chuah, J. Yu, The future of service: The power of emotion in human-robot interaction, Journal of Retailing and Consumer Services 61 (2021) 102551.
[13] F. Baek, D. Kim, G. Lee, B. Choi, S. Lee, Emotional Response Modeling for Human-Robot Collaboration in Construction., The 22nd International Conference on Construction Applications of Virtual Reality, Seoul, Korea, 2022.
[14] R. Li, Z. Zou, Enhancing construction robot learning for collaborative and long-horizon tasks using generative adversarial imitation learning, Advanced Engineering Informatics 58 (2023) 102140.
[15] Y. Sun, I. Jeelani, M. Gheisari, Safe human-robot collaboration in construction: A conceptual perspective, Journal of Safety Research (2023).
[16] Y. Liu, M. Habibnezhad, H. Jebelli, Brainwave-driven human-robot collaboration in construction, Automation in Construction 124 (2021) 103556.
[17] L.P. Kaelbling, M.L. Littman, A.W. Moore, Reinforcement learning: A survey, Journal of artificial intelligence research 4 (1996) 237-285.
[18] A.S. Polydoros, L. Nalpantidis, Survey of model-based reinforcement learning: Applications on robotics, Journal of Intelligent & Robotic Systems 86 (2) (2017) 153-173.
[19] J.A. Russell, A. Weiss, G.A. Mendelsohn, Affect grid: a single-item scale of pleasure and arousal, Journal of personality and social psychology 57 (3) (1989) 493.
[20] J. Wang, P.-C. Liao, Re-thinking the mediating role of emotional valence and arousal between personal factors and occupational safety attention levels, International Journal of Environmental Research and Public Health 18 (11) (2021) 5511.
[21] M. Morikawa, Who are afraid of losing their jobs to artificial intelligence and robots? Evidence from a survey, GLO Discussion Paper, 2017.
[22] A. Weidemann, N. Rußwinkel, The Role of Frustration in Human–Robot Interaction–What Is Needed for a Successful Collaboration?, Frontiers in psychology (2021) 707.

GPU-Accelerated Collision-Free Path Planning for Multi-Axis Robots in Construction Automation

Ilija Vukorep[1] Rolf Starke[1] Arastoo Khajehee[2] Nicolas Rogeau[2] Yasushi Ikeda[2]

[1]Faculty of Architecture, Civil Engineering and Urban Planning, BTU Cottbus-Senftenberg, Germany
[2]Department of Architecture, Graduate School of Engineering, The University of Tokyo, Japan
ilija.vukorep@b-tu.de, rolf.starke@b-tu.de, arastoo@arch1.t.u-tokyo.ac.jp, nicolas@arch1.t.u-tokyo.ac.jp, yasushi@arch1.t.u-tokyo.ac.jp

Abstract –

The Architecture, Engineering, and Construction (AEC) sector faces increasing pressure for higher production rates amidst a growing shortage of skilled labor, driving the demand for advanced robotic applications to enhance precision, efficiency, and adaptability in complex environments.

This paper introduces a software setup designed to ensure collision-free movements for multi-axis robots in AEC scenarios. Our approach leverages the NVIDIA cuRobo framework's robust capabilities, seamlessly integrated with Grasshopper for Rhino 3D software (GH), a tool widely recognized for its versatility in parametric design.

The integration of these technologies allows for the efficient online generation of optimal path movements, avoiding collisions even in highly intricate settings and changing environments. This is achieved in a remarkably short timeframe, enhancing productivity and reducing downtime. NVIDIAs framework's GPU-driven architecture paired with our GH parametric and controlling setup is a significant advancement, validated through a case study involving a complex, tree-like structure constructed from timber sticks. Using a six-axis robotic arm, the study demonstrates the system's capability to navigate and manipulate within congested spaces efficiently. With this enhanced automation workflow, new possibilities emerge for robotic applications, from industrial automation to sophisticated construction projects. Our GH software also allows visualization and exchange with URDF-models and better planning of collision logic, which was previously only possible with ROS and Nvidia Isaac technology.

Keywords –
collision-free path planning, multi-axis robots, AI robotic automation, parametric design

1 Introduction

Collision-free path planning is one of the most complex challenges in robotics, particularly in the AEC industry, where multi-axis robots must navigate dynamic, congested, and constrained environments [1]. We define collision-free in this study as the ability of the system to generate paths that avoid both self-collisions and external obstacles while optimizing for smooth motion and execution speed. Traditional path planning methods, such as manual programming or basic algorithms, are time-consuming and prone to errors, making them unsuitable for the intricate geometries and real-time adaptability required in construction scenarios. While advancements in sampling-based and grid-based algorithms have improved pathfinding capabilities, their integration into widely used design tools remains limited. Most existing solutions in platforms like Grasshopper for Rhino3D (GH) do not support real-time collision avoidance, leaving a critical gap in robotic automation workflows.

This paper addresses this gap between computational robotic planning and architectural workflows, making robotic automation more accessible to AEC professionals by introducing a novel framework that integrates NVIDIA cuRobo [2], a GPU-accelerated motion planning library, with GH. By leveraging cuRobo's advanced capabilities, the system enables real-time, collision-free path planning, even in dynamic environments. This streamlined approach eliminates many of the barriers associated with traditional methods, offering a robust solution for automating complex robotic tasks. The framework's effectiveness is validated through a case study of a robotic timber structure assembly, demonstrating its ability to navigate full cluttered environments with precision and reliability. This advancement represents a significant step toward solving one of the core challenges of robotic automation in the AEC industry.

2 Background & Related Work

The integration of multi-axis robots into the AEC industry has significantly advanced automated fabrication and assembly processes. A critical component of this integration is the development of efficient, collision-free path planning algorithms that can be seamlessly incorporated into architectural design software environments.

Path planning for multi-axis robots involves computing trajectories that avoid obstacles, adhere to the robot's kinematic constraints, and optimize performance metrics like speed and precision. Traditional methods often relied on manual programming, which is time-consuming and susceptible to errors. Recent advancements have focused on automating this process to enhance efficiency and adaptability in complex architectural tasks. One of the most common algorithms comes from the group of **Sampling-Based Algorithms**, such as Rapidly-exploring Random Trees (RRT) and Probabilistic Roadmaps (PRM), that have been widely adopted for robotic path planning in software solutions [3]. These algorithms explore the robot's configuration space by randomly sampling points and connecting them to construct feasible paths. For instance, RRT incrementally builds a tree by randomly sampling the configuration space, effectively handling high-dimensional spaces typical of multi-axis robots. However, their direct application within design environments has been limited, necessitating the development of more accessible tools for architects and designers. In contrary to this, the **Grid-Based Algorithms**, such as the A* algorithm, operate on a discretized representation of the environment [4,5]. They evaluate possible paths based on cost functions to determine the most efficient route. In construction scenarios, an improved A* algorithm with a dynamic weight heuristic has been developed for global path planning, enhancing the robot's ability to navigate complex sites. Our paper describes the implementation of NVIDIA-cuRobo Algorithm that falls into the category of **Optimization-Based Algorithms.** In recent years, optimization-based algorithms have gained prominence in robotic motion planning, particularly for complex tasks requiring smooth and efficient trajectories. These algorithms frame motion planning as an optimization problem, aiming to find the best path according to specific criteria, such as minimizing time, energy, or/and avoiding obstacles.

To bridge the gap between advanced path planning algorithms and architectural design, several plugins have been developed to enable robot programming within design software. For instance, the Software RoboDK [6], which incorporates the PRM motion planner, integrates path planning software interfaces into computational design environments to automatically generate collision-free paths within the robot's workspace. Another example is MoveIt [7], which integrates seamlessly with various computational platforms to facilitate motion planning, collision checking, and trajectory generation. Unlike RoboDK, which focuses on predefined paths within industrial environments, MoveIt is often used for more dynamic and adaptive scenarios, making it suitable for robotics applications that require real-time decision-making or complex manipulations.

Those two software solutions are commonly used in industries such as automotive manufacturing and aerospace, where fabrication setups often rely on highly predefined and controlled processes. Althought RoboDK has a GH plugin and MoveIt can be integrated within Compas-Fab [8| framework the collision-free path planning is not trivial and robust. ROS and its simulation platforms, Gazebo and PyBullet, are excellent for developing robotic systems similar to those presented in this paper. Their integration with NVIDIA Omniverse also makes them potential candidates for automatic collision-free path planning. However, these systems require specialized robotics and programming expertise, posing a barrier to adoption in the AEC industry. While no definitive study exists on the most widely used software in AEC robotic fabrication, the prevalence of Rhino and Grasshopper (GH) plugins across major robotic and software companies suggests its dominant role. GH's visual programming capabilities lower the technical barrier, making it a crucial bridge between design and fabrication, particularly in parametric workflows with evolving design constraints.

For GH, several plugins are available for robotic control and path planning, including HAL, KUKA|prc, Robot Components, Robots (by Visose), FU-Robot, and Machina [9] that makes it a serious tool for robotic control. However, despite their utility, none of these plugins currently support collision-free path planning, primarily due to the complex nature of the task and the substantial computational power required. Real-time, collision-free path planning remains a significant challenge within these design environments. The intricacy of architectural geometries, parametric setups, the presence of dynamic elements such as moving objects and actors, and the high level of customization required, demand more advanced solutions. These solutions must be capable of managing complex design constraints while providing immediate feedback to designers.

3 cuRobo

A notable advancement in the area of automatic collision-free path planning for robots is NVIDIA's cuRobo, a GPU-accelerated library [2]. By leveraging the parallel processing capabilities of GPUs, cuRobo formulates motion planning as a global optimization problem, enabling the simultaneous evaluation of

multiple potential solutions. This parallelism allows cuRobo to solve complex motion generation tasks in approximately 50 milliseconds on average, achieving speeds up to 60 times faster than traditional trajectory optimization methods. It's efficiency stems from combining straightforward optimization techniques with numerous parallel starting points, or "seeds." It utilizes the L-BFGS method—a popular optimization algorithm for estimating step directions—alongside a novel parallel noisy line search strategy and a particle-based optimization solver. Additionally, the algorithm incorporates a parallel geometric planner capable of generating plans within 20 milliseconds and a collision-free inverse kinematics (IK) solver that can process over 7,000 queries per second. These capabilities make cuRobo a powerful tool for real-time robotic applications, including those in the AEC industry, where rapid and reliable motion planning is essential for tasks like automated fabrication and assembly.

The software requires installation on a robust system equipped with a CUDA-capable NVIDIA GPU of Volta or newer architecture, operating on Ubuntu 20.04 or 22.04, although it also works on Debian 11. A Python environment with version 3.10 is recommended, and the software depends on PyTorch version 1.15 or newer. These requirements, including the reliance on a Linux operating system and high-performance hardware, present significant barriers within the architectural and construction industries, where such setups are not commonly available. Additionally, the simulation environment, based on NVIDIA Isaac, a platform designed for developing and deploying robotics applications, further complicates testing different configurations due to its specialized requirements, such as loading Universal Robot Description Format (URDF) files, and the need for high-level technical expertise to adapt it to real-world scenarios. To address these challenges and use the paralleled speed and precision of this framework, we propose a novel setup that integrates high-performance computing with GH for AEC professionals.

4 Grasshopper – cuRobo Setup

The system setup integrates three interconnected software frameworks: cuRobo, Grasshopper and MQTT (Figure 1). **cuRobo's** AI-powered system is equipped with a Python API. It operates by loading a configuration file tailored to a specific robotic setup, such as the Universal Robots UR10e and others, to enable seamless integration and control. It operates headless by just responding with the calculated results (joint values, environment, error messages etc.). **Grasshopper in Rhino3D** (GH) serves for the visual and parametric interaction of the environment (world and robot) with the user as well for the control of the real robotic system. **MQTT**-communication framework stands between the two systems that assures low-level, easy to implement exchange of data. While the MQTT-broker Mosquitto [10] is installed on Debian Workstation, MQTT-clients were implemented inside cuRobo python code and on GH side in form custom build MQTT components, both based on Paho-clients.



Figure 1. Scheme of the system setup: Left – Debian GPU workstation with cuRobo, connected to a GH user computer via an MQTT connection and to the robot via cable.

The backbone of the whole concept is cuRobo API that comprises interconnected modules designed for efficient and precise robotic motion planning. Key components include optimization solvers (curobo.opt) for refining trajectories, kinematics handling (curobo.cuda_robot_model) for forward and inverse computations, and collision detection (curobo.geom) for safe motion planning. Modules like curobo.graph enable geometric pathfinding, while curobo.rollout evaluates motion strategies by mapping actions to costs. High-level wrappers (curobo.wrap) simplify task programming, including collision-free reaching, and curobo.types standardizes data structures. The curobo.wrap.reacher module provides a high-level interface for programming tasks involving collision-free reaching. It simplifies the process of generating safe and efficient motion trajectories, enabling seamless implementation of reaching tasks within various robotics workflows.
Robots are represented in URDF files, along with their collision spheres in an additional YAML file. Another configuration YAML file defines the entire additional setup, as well as the dimensions of the optional end-effector. The world in cuRobo is also represented in this manner but can be programmatically reloaded after the system is initialized. This is important because the elements can change positions in a dynamic environment. Once initialized, the running program waits for a path creation command and returns either a successful path or a message explaining the reason for failure. Other commands update the world or read the collision spheres for improved visualization (Figure 2).



Figure 2. Rhino 3D / GH visualization of 1) Collision spheres of the robot; 2) Collision spheres of the attached stick; 3) World collision object of the cable connection of the robot.

GH handles the visualization and control of the setup. Inside GH, we read the shared robot URDF file and display the output of the resulting joint positions calculated by cuRobo. Additionally, GH is used for world creation, which is reflected back to cuRobo. Visualization and path rehearsal (Figure 4 and 5) are crucial for ensuring safe execution, especially in complex design environments. We enable designers to preview paths, identify potential issues, and refine movements before execution. This prevents costly errors and enhances user confidence. Additionally, we can plan attaching elements to the end-effector and send this configuration back to cuRobo, which incorporates the attached element into collision-free path planning. After detaching, the new world is updated with the element's new position. For this, a unique URDF reader for GH has been developed and incorporated (Figure 3). This makes it much easier to build novel robotic setups as well as collision sphere setups.

The third part is the communication system, whose purpose is to simply incorporate subscription and publisher capabilities on both sides. MQTT can be optimal for secure and rapid communication between subscribers and publishers in the form of topics. It requires an MQTT broker, which most platforms provide, and clients are widely available. For the case study, we incorporated four request topics and five response topics. Depending on the network quality and data size, it will deliver an almost immediate response (see Latency chapter).

The control inside GH consists in these three elements:

Exchange procedures GH with cuRobo. Here we can send the world setup, this means all objects in the real or simulated world, from GH to cuRobo in a form of a list of elements with their dimensions and poses. Vice versa, we can pull the world from cuRobo to cross-check how it sees the world. Furthermore, we can also pull information about the collision spheres for cross-checking the YAML file. This helps a lot also to identify the position of the robot from the cuRobo side.

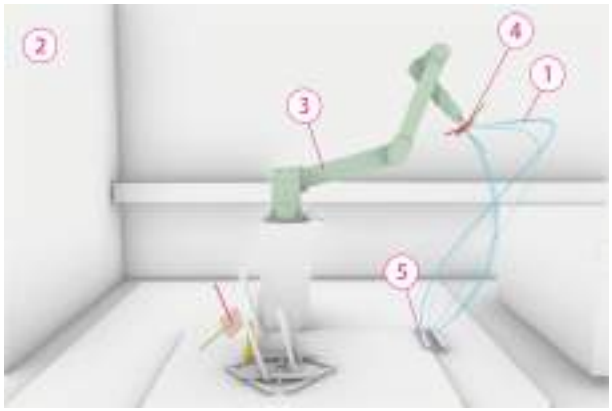


Figure 4. Representation of the real world translated in Curobo and visualized in Rhino3D / GH. 1) automated path options considering the world and self-collision; 2) world perception of Curobo; 3) Robot representation from URDF-file and cuRobo joint positions; 4) start position; 5) goal position



Figure 5. the original scenery

Path queries from GH to cuRobo in form of sending the initial start position (joints) and the desired goal position (plane, position, and pose). This fires back an array of joint positions that can be previewed. Every query will generate a unique new collision-free path. It considers the environment and prevents self-collision. In addition to the path query, we can attach or detach objects to or from the end-effector, in the case of gripping an object or similar. In this case of attaching, the object is removed from the world, and additional collision spheres will be created in the volume of the object as a part of the robot. All subsequent movements of the robot take these new collision spheres into account. In the case of not finding a valid solution for the target pose, start or goal collision states, or other failings, we can provide GH with feedback from cuRobo.

Robot control commands send the calculated new joints to the robot. Here, we can also add other commands, such as gripper control, depending on the robot being used.

5 Latency

The great advantage of cuRobo is its speed. In contrast to other motion planners described in the previous chapter, which often take several seconds to optimize a trajectory, cuRobo requires only 1/5–1/10 of a second. Black Coffee Robotics [11] compared RRTConnect-based MoveIt2 (MI) and cuRobo. In an obstacle-free environment, the results are similar (cuRobo/MI – 100%/96% success rate – 0.19 sec./0.17 sec.). In an environment with obstacles, cuRobo outperformed MI in terms of reliability and trajectory quality (cuRobo/MI – 90.32%/62.50% success rate – 0.69 sec./0.33 sec.).

The response in a system with multiple elements, as described in this paper, is dependent on several factors, such as the computers used and the connection between them. In our setup, cuRobo runs on a Linux workstation with an Nvidia RTX A4000 GPU with 16GB RAM. The computers are connected over a VPN network with a minimum 200 Mb/s download speed using the MQTT framework. GH is running on a MacBook Pro (2021). The average path calculation measured in GH, from activating the toggle until receiving the joint positions, is approximately 200–250 ms (Figure 4). With such speed, an almost continuous flow preview of possible paths for moving elements is possible.

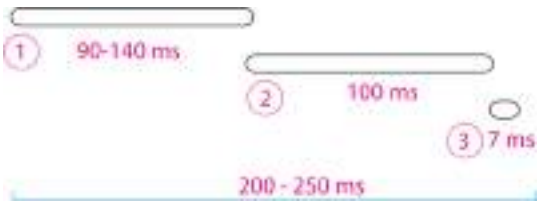


Figure 6. Approximated processing time 1) Path calculation (Python API time measure); 2) MQTT transfer (Mqtt-explorer timestamp); 3) GH path creation (GH profiler).

6 Case Study

To test the automatic collision-free path planning with conventional path planning, we required a sufficiently complex structure that went beyond the typical simple stacking examples. For this purpose, we utilized an ongoing collaborative project with the University of Tokyo. The task involved constructing a wooden structure composed of identical modular elements measuring 18x18x300 mm. In this "classical path planning scenario," two robotic setups were established in Japan and Germany to test remote collaboration scenarios for robotically assembled structures. In each facility, a collaborative robotic arm (cobot - respectively UR10E and UR10) was installed in front of a wooden base on which to perform the assembly. Both setups' gripper end effectors were the OnRobot RG6 finger gripper. The different tool mounting technology caused slight variations in the tool center points. Despite these small differences in the robots' end-effector positions relative to their frames of reference, both setups had very similar reachable workspaces.

In this modular construction system, the sticks are glued together along one of their long faces to form tree-like structures. The simplicity of the assembly rules, paired with the complex architectural results they create, makes it an excellent platform for experimenting with robotic assemblies [11]. The outcome of the original tests was a twin structure in both locations together with the design and fabrication data that was exchanged through a synchronized JSON file stored on a shared Dropbox folder. The JSON file included geometrical information

that was used to build another twin structure [12].

We replicated the experiment of the structure assembly of the first 12 sticks, and instead of manually or parametrically constructing the path, we relied solely on the cuRobo path-planning capabilities. The 3D model was built in GH, and all sticks were positioned at the pick position. The assembly consisted of several parts where all movements and attaching/detaching were made by cuRobo:

1. Movement from home to one of three offset gripping positions of the stick to be picked.
2. Opening the gripper and a movement towards the picking position.
3. Closing the gripper and a movement to the offset position.
4. Virtual attaching the stick to the robot.
5. Movement towards the offset placing position.
6. Movement to the placing position.
7. Opening the gripper and virtual detaching the stick from the robot.
8. Movement to the offset position.
9. Movement back to the home position.

All movements had their preview state in the form of a curve to ensure that the movements were possible and within the desired area. As soon as a preview was validated as valid, it was sent to the robot for execution. The gripper openings and closings were controlled manually through the same GH controller.

The movements were extremely accurate and immediately available. As we had the previews of the movements, we could execute the robot without fear of colliding with other objects, even with the stick in the gripper (Figure 5). Sometimes, it moved very near to objects, which made us nervous.

After the fifth stick, a collision still happened. We forgot to account for the cable connection to the robot that was facing towards the sticks. The joint position of the shoulder squeezed the connector and disconnected it from the power. We had to release the joints manually and rotate the shoulder away. After we integrated the connector into the world (Figure 3), this unusual position never repeated.

As we tested the cuRobo URDF file of a UR10e robot on a real UR10 robot, discrepancies in the calculated positions limited us to only 12 sticks. These discrepancies arose because the base of the UR10e is 15 mm higher than that of the UR10, along with some minor differences in arm length and will be improved in the next tests.

This specific case study demonstrated the technical capabilities of cuRobo in generating fast and precise collision-free paths, seamlessly integrating with Grasshopper for real-time visualization and robotic control. However, limitations emerged, such as discrepancies between simulated and real-world robot models, the need for manual adjustments to account for unmodeled elements like cables, and the lack of real-time dynamic obstacle handling.

7 Conclusion and Discussion

This study successfully integrates advanced AI-based software for collision-free path planning into an architectural workflow, offering significant potential for robotic automation in the AEC industry. By combining the NVIDIA cuRobo framework with Grasshopper in Rhino3D, the system eliminates the need for complex environments like NVIDIA SIM, streamlining the process and lowering adoption barriers. The case study demonstrates the framework's precision and efficiency in generating accurate collision-free paths, enabling complex assembly tasks with minimal intervention. Real-time visualization further empowers users to plan and execute robotic movements confidently.

The system excels in speed, precision, and accessibility. The GPU-accelerated cuRobo framework outperforms traditional motion planners, generating results in near real-time. Its integration into Grasshopper ensures AEC professionals can work within familiar environments, enhancing usability. Additionally, the system's ability to adapt to dynamic changes, such as object attachment and detachment, underscores its flexibility for construction applications.

Despite these strengths, challenges remain. The dual-hardware setup and the need for a CUDA-capable Linux system pose barriers for widespread adoption. Compatibility issues, as seen in the discrepancies between UR10e and UR10 hardware, highlight the need for precise configuration, and oversights in collision modelling, such as the robot's cable connection, demonstrate the importance of comprehensive planning.

Future research will focus on incorporating real-time scanning and dynamic feedback to handle variable environments, as well as developing intuitive control methods like voice or semantic commands. Expanding compatibility to include more robot models and releasing an open-source library is planned.

Acknowledgments

We gratefully acknowledge the funding of the project by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 519047184.

References

[1] Latombe, J.-C., Robot Motion Planning. Springer US, Boston, MA., 1991

[2] Sundaralingam, B., Hari, S.K.S., Fishman, A.,

Garrett, C., Wyk, K.V., Blukis, V., Millane, A., Oleynikova, H., Handa, A., Ramos, F., Ratliff, N., Fox, D., cuRobo: Parallelized Collision-Free Minimum-Jerk Robot Motion Generation, 2023

[3] Cui, C., Wang, Z., Sui, J., Zhang, Y., Guo, C., An improved RRT behavioral planning method for robots based on PTM algorithm. Sci Rep 14, 21776, 2024

[4] Ye, X., Guo, H., Guo, Z., Luo, Z., A construction robot path planning method based on safe space and worker trajectory prediction. ISARC Proceedings 2024 Proceedings of the 41st ISARC, Lille, France, pages 227–235, 2024

[5] Liu, L., Wang, X., Yang, X., Liu, H., Li, J., Wang, P., Path planning techniques for mobile robots: Review and prospect. Expert Systems with Applications 227, 120254, 2023

[6] Collision-Free Motion Planner - RoboDK Documentation [WWW Document], n.d. URL https://robodk.com/doc/en/Collision-Avoidance-Collision-Free-Motion-Planner.html (accessed 10.24.24).

[7] Motion Planning — MoveIt Documentation: Rolling documentation [WWW Document], n.d. URL https://moveit.picknik.ai/main/doc/concepts/motion_planning.html (accessed 12.25.24).

[8] Planning scene and collision objects — COMPAS FAB [WWW Document], n.d. URL https://compas.dev/compas_fab/latest/examples/03_backends_ros/05_collision_objects.html (accessed 12.25.24).

[9] Foo4Rhino [WWW Document], n.d. . Food4Rhino. URL https://www.food4rhino.com/en/browse (accessed 12.18.24).

[10] Eclipse Mosquitto [WWW Document], 2018. . Eclipse Mosquitto. URL https://mosquitto.org/ (accessed 12.25.24).

[11] Kaneko, T., Khajehee, A. & Ikeda, Y. Heuristic Fabrication: An Interactive Robotic Building System for Enhancing Human Participation in Timber Structures. (2024).

[12] Khajehee, A., Rogeau, N., Abdelaziz, M., Kotov, A., Starke, R. Vukorep, I. & Ikeda, Y. Collaborative Exploration of Complex Design Spaces: Integrating constraints from two remote robotic setups. Proceedings of CAADRIA, Tokyo, Japan (2025) (In Press)

Investigating the Evolving Robot-Building Relationship via a Past-Present-Prospect Model

Junjie Chen[a], and Weisheng Lu[a]

[a] Department of Real Estate and Construction, The University of Hong Kong, Hong Kong, China
E-mail: chenjj10@hku.hk, wilsonlu@hku.hk

Abstract –

Architects and construction engineers have long been intrigued by a future where robots become pervasive in the built environment. Recent technological advancement has made this once utopian vision appear more feasible than ever. However, in the discourse of construction robotics, there seems to be insufficient discussion on the relationship between robots and buildings, the two key entities involved. This study delves into the robot-building dyad by taking a historical perspective to examine how their interactive dynamics evolves over time: (a) from an objectified dualism in the past, (b) to an increasingly blurry boundary in the present, and (c) to a prospective symbiosis with mutual benefits. We argue that this shifting dynamic is underpinned by large-scale digitalization and the empowerment of artificial intelligence. Blue-sky prospective scenarios are discussed, where the robots become an inherent part of the building (like the MEP system), or the building itself becomes a robot (like Howl's moving castle). It is our hope to stimulate further discussion on the critical robot-building relationship, gaining a clearer view of which could help overcome the theoretical dilemma hindering wide-spread robotization in the built environment.

Keywords –

Construction robotics; Robot-inclusive built environment; Robot-building dynamics; Robot-oriented design.

1 Introduction

The architecture, engineering, construction and operation (AECO) sector is facing enormous challenges in recent years. Deeply rooted in the complexity and fragmented nature of the industry, its productivity has been stagnant over the past 20 years [1]. The problem is further exacerbated by the lack of skilled labor force, and the increasingly stringent environmental requirements imposed by climate change. To make the situation even worse, many developed economies in the world are experiencing simultaneous deterioration of their building mass and demographic profiles, which is usually referred to as a "double aging" issue [2]. The aging building stock is awaiting proper maintenance or redevelopment on the one hand; and on the other, the shrinking population fails to provide solid workforce for facility upkeep.

Robotization offers a promising solution for the demographic challenge, as well as performance improvement in the AECO sector. For their superiority over humans in terms of consistency, precision, efficiency, and durability, the use of robots is expected to enhance productivity, ensure safety, and reduce wastes [3]. The benefits of robotization have been recognized from early on. In 1980s, pioneering attempts were first made to deploy robots in construction, with Japen as the leading player [4]. These explorations were profit-driven in essence, and thus were largely spearheaded by industrial practitioners. It was not until ten years later that the academia started to follow up. Signified by the establishment of The International Association for Automation and Robotics in Construction (IAARC) in 1990, systematic research efforts have been made to examine construction robotics from a theoretical perspective. The research input has been stagnant afterwards due to technological limitations, but the advancements in AI and automation in recent years re-attract significant attention to this area.

Despite its ups and downs, the discourse on construction robotics seems to be predominantly technology-oriented. Seldom has previous research taken a broader view to examine the relationship between robots and buildings – the two core entities explicitly expressed in the terms "Construction Robotics". Clarification on this somewhat overlooked dyad is essential because of a mutually reliant but also exclusive nature. They are mutually reliant, in a sense that the robots require the physical environment provided by the building to carry out their activities, whereas the building depends on the robots to materialize, inspect, and maintain. At the same time, the distinctive differences between their inherent characteristics also lead to a series

of compatibility problems, whether it is regarding the autonomy level, functional purposes, or just the sheer size of them.

The few attempts to research the topic mainly happened during late 1980s and early 1990s. Back then, the field of construction robotics was still in its infancy. Inspired by the success of robotics in manufacturing, research at that time tried to draw useful experience by separating the building as individual entity from the dyad and comparing it to the production line [5]. Similar to "lean" production in the manufacturing industry, Koskela [5] proposed a process improvement in construction to embrace robotization, but the fragmented and project-based nature of construction makes it more difficult to achieve standardization. Warszawski and Sangrey [6] conducted an in-depth analysis on the respective features of industrial robots and building activities, and proposed possible adaptations of the construction process and building components for efficient applications of robots. Skibniewski and Nof [7] evaluated the readiness of existing construction environment for robotization, and recommended required changes in work organization and construction system. Despite the foundation laid by these pioneering studies, the robot-building relationship has never been formally defined and investigated. Nor has it been examined via a systemic framework in terms of how the relationship evolves over time.

To fill in the gap, this paper aims at examining the dyadic relationship between robots and buildings and its evolvement through a historical perspective. A "Past-Present-Prospect" model is conceptually proposed, which views robots and buildings as dynamically evolving entities that change with the advancement of technologies such as digitalization, industrialization, and AI. The research contributes to the field of construction robotics by theorizing current dilemma faced by construction robotics as a robot-building compatibility problem, and paving the way to its resolution with a clearer view on the robot-building relationship.

2 A conceptual model of the evolving robot-building relationship

A "Past-Present-Prospect" model is proposed to help interpret the evolving robot-building relationship, as shown in Figure 1.



Figure 1. Conceptual model of evolving robot-building relationship.

The study of a subject matter would not have been possible without a clear definition of it. This is also the same when it comes to investigating the robot-building relationship with a historical perspective. The connotations of term "robot" and "building" shift gradually over time, and the scope of objects they refer to does not remain static. For the sake of discussion, we tend to adopt the broadest possible definitions of both. According to the Cambridge Dictionary, a building is a structure with walls and a roof; it can also refer to the process or business of making structures. As for robots, they are generally accepted as machines capable of carrying out a complex series of actions automatically. Despite the varying level of autonomy (e.g., Teleoperation, Supervisory control, Autonomous) [8], an object has to be a machine to become a robot.

As shown in Figure 1, with the risk of over-simplification, both robots and buildings are going through a process of transformation from objects — passive entities that cannot react to external stimulus, to agents — proactive entities with the ability to perceive, plan, and react.

In the past, a robot is nothing more than a deaf-and-dumb machine. Despite being through several revolutionary innovations in terms of their power source (from horse power, to steam and to internal combustion) [9], they were used merely as objective tools to materialize the built environment. The building, on the other hand, is solely made of bricks and mortar [10]. As objects, they can neither sense nor think, let along reacting to a change of the environment. With this objectified nature, the relationship between the machines and buildings is purely utilitarian. Machines are used to construct the building, and the building, as gradually being materialized, offers the physical environment where the robots operate.

At present, with the arrival of the digital era, this relationship gradually changes. A widely accepted but less radical solution is to retrofit existing machines and buildings with an array of digital technologies, in particular the ability to sense intrinsic or environmental changes [11]. As they are being digitalized, the boundary between robots and buildings are becoming vague. The interaction between digital machines and digital buildings not only happens in the physical space, but also in the digital world. This is because the robot can simultaneously sense the built environment and update it to the corresponding digital model as it is being materialized. The other way around, the digital building model, e.g., a building information model (BIM), not only serves as an environment model in the virtual space useful for robot simulation, but also provides a reference map for the robots to perceive their environments.

In the long run, as robots and buildings become gradually intelligentised, they would eventually become

agents. On the one hand, the machines will eventually become robots in the most rigorous sense, meaning that they are able to sense environmental changes, to process and interpret them, and to response independently. On the other hand, the once bricks-and-mortar structures will be turned into smart buildings, which perceive and proactively response to internal (e.g., occupancy) or external changes (e.g., weather, sunlight). This dyad would then eventually become a mutualistic symbiosis, where they are nurtured by each other, and mutually enhanced. However, as a millennium-old profession, construction is deeply rooted in its history, local culture, geographical conditions, etc., and notoriously reluctant to changes. The road to autonomy (and the resulting symbiosis) would be by no means easy. Therefore, researchers [11, 12] have suggested distinct paradigms, which either propose an incremental shift based on existing mature tools and techniques, or strive for a radical change by adopting emerging technologies.

3 Past-to-Present: Blurring boundary between robots and building

The shifting robot-building dynamics from the past to present is underpinned by digitalization. Augmented by an array of digital technologies (e.g., smart sensing, automation, IoT, CAD, simulation), the once deaf-and-dumb machines and buildings are empowered by a sensing ability to understand a change of their environments and their digital representations for simulations. They thus become digital machines and buildings.

The various construction machines used in the building sector evolve as the broader industrial context changes. From an energy source perspective, they have gone through a transformation from being steam-driven to diesel-driven. The latest equipment of computer systems and various sensors elevate them to another level as digital machines, which are able to perceive their own operating status and external environment. When it comes to a digital building, interpretation concerning its definition, and connotation varies. Watson [13] confined the scope to only include a digital model that contains structured information of a physical building, whereas others consider a building augmented by an array of IT infrastructure (e.g., a building control system) as a digital building. This study accepts both interpretations, and defines a digital building as an IT-augmented man-made structure, and its digital representation used in various phases of project cycle.

The digitalization of machines and buildings means they can escape the constraints imposed by their specific physical embodiment. Therefore, interaction between the digital machines and digital buildings no longer remains at a materialism level, but enters the realm of cyber space. This allows the digital models of the machines and buildings talk to each other, and boundary lying in-between becomes blurry. On the one hand, the digital machines not only materialize the buildings, but can also map their surroundings simultaneously along the way; on the other hand, the digital building not only serves as the physical environment indispensable for the machines to operate, but also provide a digital model of the environment to empower the machines (e.g., to understand the environment).

3.1 Mapping buildings with digital machines

A major stream of recent works has been focused on sensifying traditional construction machinery. They are aligned with the incremental approach proposed in [11, 12], which renovates legacy machinery with suite of sensors and computing units. The purpose is to turn the machinery from purely a production tool to instruments with both production and measurement utility. Examples can be drawn from a field called intelligent compaction. In these studies, an array of sensors is installed to a conventional roller used for earthworks/pavement compaction. A typical setup includes an accelerometer mounted on the drum to infer material compactness from vibration, a GNSS unit to track position of the measurement, and some other sensors to collect values deemed necessary, e.g., temperature for asphalt pavement [14]. Thanks to the empowerment of sensors, the rollers can measure whether the materials have reached a desired level of compactness as it carries out the compaction work. Therefore, the moment the compaction finishes, a quality map of the entire working area can be expected. Such sensory augmentation goes beyond compacting machines. Any mechanical-based construction machinery can be given senses to with the help of modern computing instruments. Niskanen et al. [15] developed a 2D profile-augmented excavator, which can dynamically map earthwork surface as it operates. Sun et al. [16] sensified a dozer system with RTK GPS and inertial measurement unit (IMU) to collect machine pose information for construction guidance. The digital machines not only materialize the built environment in a physical sense, but also enrich the digital building/infrastructure model by constantly updating its condition information from the physical world.

The mapping function of digital machines does not stop at the construction phase. They have a critical role to play in facility operation and maintenance as well. These include, but not limit to, flying drones to inspect high-rise façade [17], and quadruped robots or unmanned ground vehicle (UGV) for indoor inspection. In many cases, such robots still rely on humans to remotely control their movements. With high-resolution cameras or LiDAR scanners onboard, they are able to collect first-hand data of the facility condition. These data are then

automatically processed by AI, which can detect defects of various types (cracks, dampness, spalling, etc.). 3D reconstruction and registration techniques are then applied to incorporate defect information (semantics and geometries) into the digital building model such as BIM. Figure 2 summaries a typical workflow.



Figure 2. Workflow of mapping building condition with drones.

3.2 Empowering machines with digital buildings

The other around, a digital building can empower digital machines in various ways. Most notably, a digital building model like BIM can contribute to robot perception of its surrounding by providing a readily available reference map. Two typical problems for robot perception are a) positioning and pose estimation, and b) object recognition/scene understanding.

In terms of robot positioning, several previous studies attempted to explore BIM as an internal world model to estimate the robot's spatial location and orientation. However, there is a cross-domain gap to overcome, which is caused by distinct texture and appearance differences between BIM and its corresponding real-world scene. To address this domain gap, end-to-end training-based approaches have been applied to directly regress robot pose information from BIM renderings [18]. A different approach is to use BIM renderings as a pool of space-aware features that can be compared with real-life photographs for pose estimation. Ha et al. [19] used high-level visual features extracted by CNN as registration targets to avoid the domain gap. Chen et al. [20], on the other hand, proposed a neural rendering approach to turn textureless BIM renderings into photorealistic ones with vivid texture. The synthetic images are not only more plausible to humans, but also more machine-readable (see Figure 3). Rich feature correspondences can be found between robot-collected photos and the synthetic renderings, which then allows pose estimation by solving a typical PnP problem.

By positioning current robot view in BIM, the rich semantic information from the digital building model becomes easily accessible for the robot's perception of the world. This typically involves: (a) Extracting current view from BIM; (b) Overlapping it onto camera view; (c) Retrieving semantics of arbitrary pixel in BIM; (d) Facilitating robot perception with the semantics from BIM. Figure 4 shows the perception of a drone in a building façade scene, where no training has been conducted for object recognition. Instead, it purely relies on semantic information from BIM.



Figure 3. Addressing domain gap between BIM and real scene via neural rendering (with SIFT as an example) [20].



Figure 4. Aligning aerial photograph with BIM rendering to facilitate robot perception.

4 Present-to-Prospect: Towards mutualistic symbiosis as intelligent agents

Looking ahead, both robots and building would eventually become agents with the ability of independent decision-making and autonomy. As agents, they will form a mutualistic symbiosis, where robots and buildings would rely on each other to grow and develop, and benefit from the association. To embrace this future, we need to (a) make robot smart, (b) make building smart, and (c) facilitate robot-building communication. Thanks to the advancing AI, many of these transformations are underway [21].

4.1 Making robot smart

Despite rapid development in the field of robotics, delegating robots fully with independency and autonomy remains risky and challenging. This is even the truth in the built environment with high occupancy dynamics, spatial complexity and affordance diversity.

It is thus imperative to make robots smart, especially regarding spatial intelligence [22], an ability for the robots to process spatial data, make predictions and act upon those predictions. As robots always function in a specific space context (e.g., the built environment), this spatial intelligence is particularly relevant. Connectivism

and symbolism have long been two mainstream schools of thoughts in AI. While the super performance of neural networks in recent years has put connectivism under the limelight, it is believed a convergence of top-down symbolism and bottom-up connectivism is critical when it comes to empowering robots with spatial intelligence.

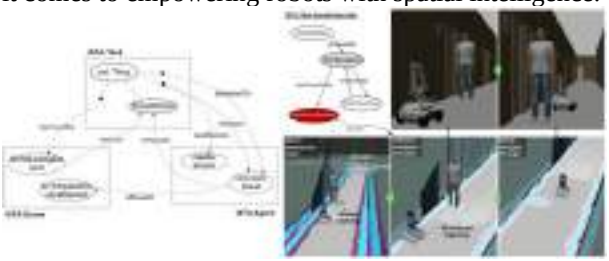


Figure 5. Knowledge graph for smart robot-based facility inspection.

Figure 5 shows a top-down knowledge graph for robot-based facility management (FM). It outlines how the robots should behave when facing different scenarios in the built environment. Combined with bottom-up connectivism neural networks (e.g., for human detection), it can empower robots to carry out advanced actions such as collision avoidance with humans in the corridor. FM robots are just an example. Thanks to the rapid advancement of AI, the entire AECO sector is adopting robots with increasing level of autonomy (LoA). Melenbrink et al. [11] conducted a comprehensive review of existing onsite autonomous construction robots, and rated their LoA according to a categorization system adapted from autonomous vehicle.

4.2 Making building smart

To make building smart, a plethora of studies have been conducted, either from a theoretical perspective or from an implementation standpoint. Ranging from smart home [23], cognitive building [24], to cognitive facility management [25], numerous terms have been proposed to explain and foresee how the once brick-and-mortar buildings would become intelligent with the empowerment of AI.

Different from the current digital buildings that are only able to sense and collect operational data, a smart building evolves to another level where it can react and response to a change of its external environment and internal occupancy situation. Such smart buildings can be built from the bottom up by the empowerment of intelligence to discrete building components (e.g., precast beams, columns, and walls) [26], and even to fine-grain construction materials [27]. As building components are intelligentized, they would eventually become agents that can better co-work with the robots as a material-robot system [28].

Once commissioning, the smart building can comprehensively analyse the data collected by the various sensors scattered in the building, and then independently make decisions given the current conditions. The decision would next be issued, activating actuators around the buildings to implement the decided plans. These smart building reactions can be as trivial as light control responding to room occupancy, or can be much more complicated in scenarios such as window curtain control in response to changing natural lighting. Figure 6 shows a smart living system adopted in the WCH student hostel project at The University of Hong Kong (HKU). It has integrated a smart light control system, an electronic lock system that can be controlled by students on their mobile phones, and a smart camera system to ensure safety of the dormitory area.



Figure 6. Smart building system adopted by HKU WCH project.

4.3 Facilitating robot-building communication

Smooth communication is a premise for taking advantages from each other between robots and buildings, further leading to the formation of the symbiosis. To this end, the robots and buildings need to be able to talk to each other, which entails a common language between the two parties. This, unfortunately, is not the case in current practice, as robotics and AECO industry have developed their own data schemas, i.e., the Unified Robot Description Format (URDF) for robots, and the Industry Foundation Classes (IFC) for buildings.

To mediate the gap, a potential solution is to create a common data environment between the robots and buildings based on IFC. The basic idea is to represent robots and relevant information with IFC, so that they can be compatible with mainstream building software for various applications, such as design, remote control, digital twin and so on. Figure 7 shows a model view definition for robot representation in IFC. Following this data model, a format translator has been developed to convert many of the existing URDF-based robot representations to IFC.

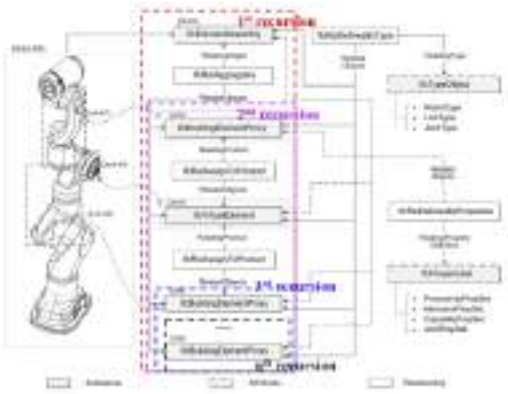


Figure 7. Model view definition for IFC-based robot representation.

Such an IFC-based common language can facilitate robot-building communication, enabling various applications such as robot-inclusive design. Many studies have underscored the need of rethinking building design and process improvement for robot adoption [29-31]. However, software tools are in absence to develop designs catered to the specificity of each project for robotic construction. IFC-based robot-building communication offers a means to resolve the fundamental interoperability problem in developing those design tools. Figure 8 shows an example of robot-inclusive design of a building MEP system. With the common language of IFC, the model of an MEP inspection robot can be directly imported into Revit. Since the robot and building models are in the same design software, one can easily identify clashes between them, and then timely adjust the design.

For construction/operation phase applications, as the IFC robot and building models are integrated in the same environment, the exchangeable information can be leveraged to enable application such as remote control. For example, as shown in Figure 9, by clicking a point in the building model, the robot can automatically navigate to that position. The robot arms and grippers can also be teleoperated by a remote human for more sophisticated tasks such as construction waste sorting.

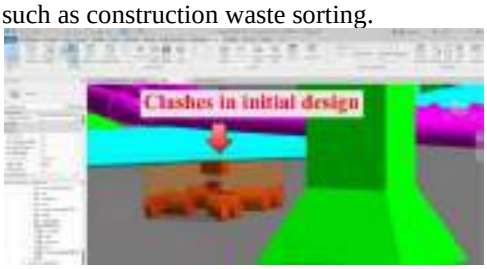


Figure 8. Inclusive design for MEP inspection robots.

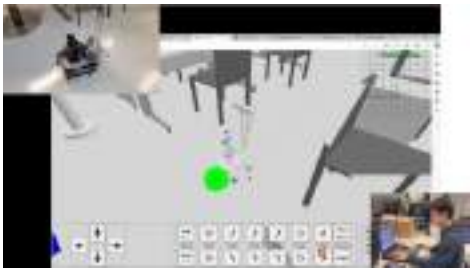


Figure 9. Robot teleoperation enabled by freely exchanged building information.

5 Discussion

Influenced by technological development, societal expectation, and cultural perception, the dyadic relationship between robots and buildings evolves over time. From a purely utilitarian dualism as objects, to increasingly blurry boundaries, and to a prospective mutualistic symbiosis, this shifting interactive dynamics calls for a fundamental rethink of what define a robot and a building, how they should be designed, and where their boundaries lie. Table 1 provides two different models to frame future robot-building relationship, and technological development required for their realization.

5.1 Robot as an inherent part of building

A way to formulate the problem is to consider robots as an inherent part of buildings. This requires a systematic thinking to take robots into account from the outset of building design. A comparable analogy can be drawn from the MEP in a building system, which needs to be considered at the design stage via multi-disciplinary collaboration for clash detection. This is gradually gaining attentions in the field of robotics. For example, in 2022, an office building called 1784 was constructed in Seongnam City, Korea, which claims to be the world's first robot-friendly building. The building was designed to allow a large population of autonomous mobile robots operating within it, and therefore is equipped with robot-friendly infrastructure, which includes world's first elevators exclusively for robots. During the Beijing Winter games in 2022, robots were widely deployed. Among them, there are robotic systems hanging from the ceiling in cafeteria, serving select cuisines individually. This is another example scenario where robots become part of the building system.

Table 1. Models and research required for future robot-building relationship

Model	R&D required

Robots as building parts	Integrated design and engineering, AI-powered generative design, IoT, BMS, Human-robot co-existence
Buildings as robots	Electromechanical building materials, Ambient intelligence, Urban science and design

Further research and development (R&D) are needed to widely deploy robots as parts of the building system. Fundamentally, an integrated mindset is required to design and engineer the building system with robots integrated. The robot design considerations, along with the uncertainty when it interfacing with buildings, is expected to significantly complicate the system design. The complexity can be mitigated by emerging AI-powered design tools. Once the robot-integrated buildings are put in place, the building management systems (BMS) will need to be substantially upgraded to ensure the robot swarms are under control, where the IoT is essential for robot-robot and robot-building communication. In the long run, with more and more robots entering the living environment, the problem of human-robot co-existence would become prominent to research for well-being.

5.2 Building as a robotic system

Taking a more radical step, would the building itself ends up becoming robots? We have been seeing buildings increasingly empowered and computerized by the use of electronics and mechanics. In certain smart home settings, it is already a reality that the furniture and layout of a room can automatically transform to meet occupants' need. In 1970s, when the idea of modular building was still popular in Japen, landmark projects such as the Nakagin Capsule Tower have been designed and developed following the broader context of urban metabolism. The Nakagin building was designed with intentional impermanence, following the principles of growth and transformation over time [32]. With all the mentioned precedents, a robotics building that can move and transform in response to external environment might not be as futuristic as it sounds after all.

R&D is required in the following directions. First, the science of building materials needs a fundamental rethink. For a building to sense and response like robots, its constituent components should be of electromechanical characteristics, so as to be compatible with the current silicon-based artificial intelligence. Second, ambient intelligence is critical to make computing pervasive in the built environment. In doing so, spontaneous response becomes possible. Last but not least, as robotic agents, the dogmatic view of building being places would be challenged. This will then pose new questions about what forms a city, which requires systemic research from a larger scale of the urban system. With continuous research input from the above directions, the Howl's Moving Castle will perhaps become a reality in the foreseeable future.

6 Conclusions

Despite being the most critical dyad in the robotization of the built environment, the relationship between robots and building has seldom been researched systematically. This paper fills in the gap by taking a historical perspective to examine the evolving interactive dynamics between robots and buildings over time: (a) from an objectified dualism in the past, (b) to an increasingly blurred boundary in the present, and (c) to a prospective symbiosis with mutual benefits. The shifting dynamics is argued to be underpinned by large-scale digitalization and the empowerment of artificial intelligence. Blue-sky prospective scenarios have been discussed, where the robots become an inherent part of the building, or the building itself becomes a robot. The paper hopes to stimulate further discussion on the critical relationship between robots and buildings. Its clarification is deemed indispensable to overcome the theoretical dilemma hindering wide-spread robotization in the built environment.

References

[1] PSGO. (2018). *Construction 2.0*. Retrieved March 20 from https://www.psgo.gov.hk/assets/pdf/Construction-2-0-en.pdf

[2] Lee K. and Ling K. K. Tackling Double Ageing with Double Smart. *HKIP Planning and Development*, *33*, 2019.

[3] IFR. (2023). *Presentation of World Robotics Report*. Retrieved May 27 from https://ifr.org/img/worldrobotics/2023_WR_extended_version.pdf

[4] Taylor M., Wamuziri S. and Smith I. Automated construction in Japan. pages 34-41, 2003.

[5] Koskela L. Process improvement and automation in construction: Opposing or complementing approaches? , 1992.

[6] Warszawski A. and Sangrey D. A. Robotics in building construction. *Journal of construction engineering and management*, *111*(3), 260-280, 1985.

[7] Skibniewski M. J. and Nof S. Y. A framework for programmable and flexible construction systems. *Robotics and Autonomous Systems*, *5*(2), 135-150, 1989.

[8] Goodrich M. A. and Schultz A. C. Human–robot interaction: a survey. *Foundations and*

Trends® in Human–Computer Interaction, *1*(3), 203-275, 2008.

[9] Xu Q. and Chang G. K. Adaptive quality control and acceptance of pavement material density for intelligent road construction. *Automation in Construction*, *62*, 78-88, 2016.

[10] [Record #2762 is using a reference type undefined in this output style.]

[11] Melenbrink N., Werfel J. and Menges A. On-site autonomous construction robots: Towards unsupervised building. *Automation in construction*, *119*, 103312, 2020.

[12] Bock T. The future of construction automation: Technological disruption and the upcoming ubiquity of robotics. *Automation in Construction*, *59*, 113-121, 2015.

[13] Watson A. Digital buildings–Challenges and opportunities. *Advanced engineering informatics*, *25*(4), 573-581, 2011.

[14] Liu D., Chen J. and Li S. Collaborative operation and real-time control of roller fleet for asphalt pavement compaction. *Automation in Construction*, *98*, 16-29, 2019.

[15] Niskanen I., Immonen M., Makkonen T., Keränen P., Tyni P., Hallman L., Hiltunen M., Kolli T., Louhisalmi Y. and Kostamovaara J. 4D modeling of soil surface during excavation using a solid-state 2D profilometer mounted on the arm of an excavator. *Automation in Construction*, *112*, 103112, 2020.

[16] Sun D.-I., Kim S. H., Lee Y. S., Lee S. K. and Han C. S. Pose and position estimation of dozer blade in 3-dimensional by integration of IMU with two RTK GPSs. 2017.

[17] Chen J., Lu W. and Lou J. Automatic concrete defect detection and reconstruction by aligning aerial images onto semantic-rich building information model. *Computer-Aided Civil and Infrastructure Engineering*, *38*(8), 1079-1098, 2023.

[18] Acharya D., Khoshelham K. and Winter S. BIM-PoseNet: Indoor camera localisation using a 3D indoor model and deep learning from synthetic images. *ISPRS Journal of Photogrammetry and Remote Sensing*, *150*, 245-258, 2019.

[19] Ha I., Kim H., Park S. and Kim H. Image retrieval using BIM and features from pretrained VGG network for indoor localization. *Building and Environment*, *140*, 23-31, 2018.

[20] Chen J., Li S., Liu D. and Lu W. Indoor camera pose estimation via style-transfer 3D models. *Computer-Aided Civil and Infrastructure Engineering*, *37*(3), 335-353, 2022.

[21] Gharbia M., Chang-Richards A., Lu Y., Zhong R. Y. and Li H. Robotic technologies for on-site building construction: A systematic review. *Journal of Building Engineering*, *32*, 101584, 2020.

[22] Yang J., Yang S., Gupta A. W., Han R., Fei-Fei L. and Xie S. Thinking in Space: How Multimodal Large Language Models See, Remember, and Recall Spaces. *arXiv preprint arXiv:2412.14171*, 2024.

[23] Ghayvat H., Mukhopadhyay S., Gui X. and Suryadevara N. WSN-and IOT-based smart homes and their extension to smart buildings. *sensors*, *15*(5), 10350-10379, 2015.

[24] Pasini D., Ventura S. M., Rinaldi S., Bellagente P., Flammini A. and Ciribini A. L. C. Exploiting Internet of Things and building information modeling framework for management of cognitive buildings. pages 1-6, 2016.

[25] Xu J., Lu W., Xue F. and Chen K. 'Cognitive facility management': Definition, system architecture, and example scenario. *Automation in Construction*, *107*, 102922, 2019.

[26] Niu Y., Lu W., Chen K., Huang G. G. and Anumba C. Smart Construction Objects. *Journal of Computing in Civil Engineering*, *30*(4), 04015070, 2016.

[27] Wood D., Yablonina M., Aflalo M., Chen J., Tahanzadeh B. and Menges A. Cyber physical macro material as a UAV [re] configurable architectural system. pages 320-335, 2019.

[28] Werfel J., Petersen K. and Nagpal R. Designing Collective Behavior in a Termite-Inspired Robot Construction Team. *Science*, *343*(6172), 754-758, 2014.

[29] Bock T. (1988). *Robot-Oriented Design* 5th International Symposium on Automation and Robotics in Construction, Tokyo.

[30] Sun Z., Mei H., Pan W., Zhang Z. and Shan J. A robotic arm based design method for modular building in cold region. *Sustainability*, *14*(3), 1452, 2022.

[31] Scott Howe A. Designing for automated construction. *Automation in Construction*, *9*(3), 259-276, 2000.

[32] Ishida A. Metabolic Impermanence: The Nakagin Capsule Tower. *Inflection-Journal of Melbourne School of Design*, *4*, 2017.

Autonomous Construction Robot with Telescopic Arm for Drywall Panel Installation

Ayshin Bagherzadeh[1], **Vahid Dolatkhah**[2], **Amin Mirfakhar**[3], **Omid Valinezhad**[2], **and Mohammad Javad Haghi**[1]

[1] Department of Mechanical Engineering, AmirKabir University of Technology, Tehran, Iran
[2] Department of Mechanical Engineering, Islamic Azad University Science and Research Branch, Tehran, Iran
[3] Department of Mechanical Engineering, Sharif University of Technology, Tehran, Iran

Abstract–
Conventional techniques for installing panels face several problems, such as human errors, time inefficiency, and safety risks. The panel installation robot, outfitted with sophisticated mechanical, electronic, and software technologies, proficiently addresses these challenges. The robot employs omnidirectional motion mechanisms, a telescopic robotic arm, and precision installation tools, including drills and nail guns, to facilitate the accurate and safe execution of installation operations. The robot's positioning system, with advanced cameras and image processing technology, can precisely identify the runners' location and modify the arm's movements accordingly. The semi-automatic control via a web user interface enables operators to remotely monitor and control the installation process. Furthermore, the robot's safety mechanism, utilizing distance sensors and cameras, averts collisions with obstacles, so guaranteeing safer operation. This unique design decreases installation time by 10% to 30%, depending on the length of the panel, while simultaneously enhancing productivity, precision, and addressing environmental concerns.

Keywords –
Construction Robotics, Automation in Construction, Drywall Panel Installation, and Telescopic Robotic Arm

1 Introduction

The incorporation of Building Information Modeling (BIM) as a holistic tool for design, planning, and project management within the construction sector, alongside significant advancements in robotics, has generated novel prospects for utilizing robots in unstructured settings like construction sites [1]. These developments allow robots to adapt effectively to the intricate and dynamic conditions characteristic of construction environments, ensuring dependable and efficient performance. In the last thirty years, construction robotics has advanced significantly, being utilized in tasks such as painting, welding, exterior cladding installation, major component assembly, precast concrete panel manufacture, and site cleaning [2,3]. Drywall installation, which demands high levels of accuracy and technical expertise, is typically performed by experienced workers. However, installation complexity, and human error make this process challenging. Incorrect measurements, cutting, and drywall misalignment cause uneven surfaces and gaps, increasing repair and rework expenses. Furthermore, the time-consuming nature of drywall installation usually causes delays in large construction projects. To address these challenges, robotic technologies have emerged as an innovative solution, particularly in the installation of interior panels like drywall. The robot studied in this research is specifically designed for the installation of various types of building panels. The robot features a fully automated system for installing drywall panels, which, unlike semi-automated or operator-dependent systems in previous studies, requires no human intervention and offers higher safety and installation speed [4,5]. By automating the installation process, robots not only reduce the dependency on skilled labor but also minimize errors and ensure consistent quality in construction projects. This efficiency is not only about speed but also about the ability to work continuously without the need for breaks, thereby streamlining the construction process. The robot presented in this study (Wall-bot) is designed to operate on various surfaces and install different types of panels, such as drywall. Unlike many previous systems that were limited to specific materials or a particular type of installation, this robot is equipped with two distinct installation tools—an electric drill for metal and a nail gun for concrete [6-8]. These features enhance installation efficiency and align with the evolving requirements of the construction industry. The innovative design of Wall-bot met the requirements of construction tasks. It aims to provide fundamental design

elements while addressing panel installation limitations. Ultimately, this system is market-ready and can reduce installation time, which is a significant advantage compared to systems that are still in the experimental stage or operate slower than manual methods [9]. By decreasing project time and the need for specialized labor, overall costs are significantly reduced. These features enhance installation efficiency and align with the evolving requirements of the construction industry.

2 Design and Implementation

This section details the conceptual design and implementation of various components of the robot, which has been specifically developed for the installation of panels in construction environments. The robot's design incorporates mechanical, electronic, and software systems that operate seamlessly and with high precision to efficiently perform the panel installation process.

2.1 Conceptual design

A robotic system has been designed and developed to overcome the obstacles associated with the installation of panels. The complete process is designed to resolve critical issues and enhance the installation process. Drywall panels are an ideal choice for automation technologies due to their lightweight structure, ease of transportation and installation, and lower execution costs. However, the installation process of these panels faces challenges, such as the need for high precision, ensuring safety in construction environments, and a significant demand for time and labor. In this study, a robotic system that is equipped with advanced technology has been developed to facilitate the installation of panels with superior precision, speed, and efficiency. This section details the conceptual design and implementation of various components of the robot, which has been specifically developed for the installation of panels in construction environments.

One of the most notable features of this robot is its advanced telescopic arm. This arm is specifically designed to provide precise adjustment of the angle and position of panels based on the ceiling-to-floor distance. With five degrees of freedom, the telescopic arm is fully automated for the transportation, positioning, and installation of non-load-bearing building panels. The robotic arm has two linear actuators arranged in parallel along the vertical axis. These actuators are designed to adjust the height of the panel. Additionally, the arm has a linear actuator that enables the forward and backward movement of the entire robotic arm. This actuator, in conjunction with another actuator connected via a metal linkage—one end attached to the robot's body and the other end to the telescopic arm—enables angular adjustments in the robotic arm. Another significant design challenge was the transportation of panels during the installation process. To address this, the robotic arm is equipped with a panel gripper, enabling it to perform all stages of transportation, positioning, and installation fully automatically, without requiring manual labor. For mounting panels onto ceilings, walls, and floors, the telescopic arm is equipped with two types of installation tools designed for attaching panels to both metallic and concrete surfaces. This system operates as an electric drill and a nail gun, guaranteeing accurate and secure panel installation.

The robot is equipped with an advanced positioning and control system specifically designed for panel installation operations in complex construction environments. This system utilizes advanced technologies, such as smart cameras, image processing, and a web user interface (Web UI), to ensure precise movement of the robot and accurate panel installation. Smart cameras are used to ascertain the precise positions of runners (structures onto which panels are affixed). These cameras capture both chromatic pictures and three-dimensional information about the surrounding environment. The collected data is processed to analyze the geometric features of the surfaces, boundaries, and lines. This analysis enables accurate determination of the panel installation positions. The Edge Detection algorithm is used to identify the boundaries and lines of the runners, thereby determining the exact locations for panel installation. The main problem in the installation of building panels is the need for high accuracy and flexibility in complex environmental conditions. This procedure is traditionally performed manually, often resulting in increased installation time, less precision, and possible safety hazards for workers. To address these challenges, an advanced control system has been designed for Wall-bot, automating several tasks, such as guiding the robot to predetermined locations while also allowing manual control by an operator when necessary. The technology employs a web user interface (Web UI) that enables the operator to precisely control the robot in various work conditions and determine the exact locations of the runners. The Web UI, functioning via a local network, synchronizes the robot's wheels and arm, enabling users to control the robot using any device connected to a wireless network. The Web UI includes three primary components:

- **Camera Live View:** It provides live video streams from the cameras mounted at the center of the robot's arm. The strategic positioning of this camera provides the user with a precise view of the working area, enabling more accurate management of operations.
- **Robot Wheel Control:** This section is designed for managing the movement and positioning of the

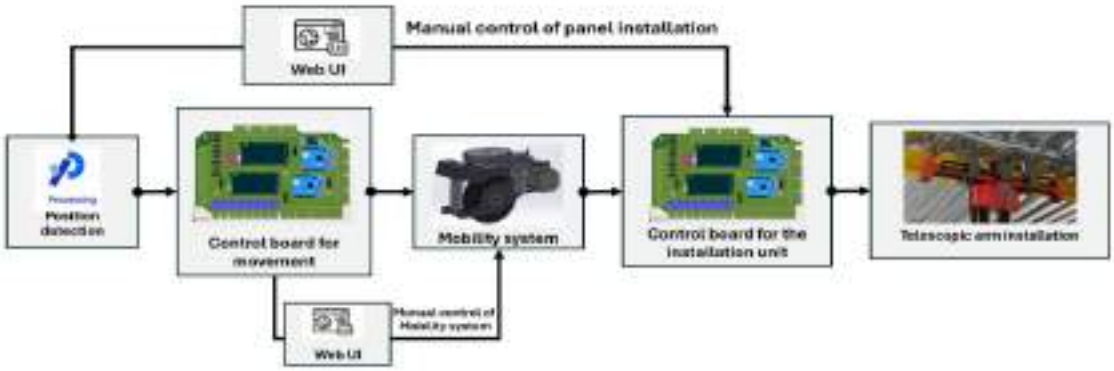


Figure 1. Block diagram illustrating the process of the panel installation. The process begins with position detection using an image processing system and cameras. The mobility system, controlled by an advanced system, enables accurate movement to the target location. Sensors enhance safety and prevent collisions. Finally, the telescopic arm securely installs the panel. Manual control via a Web UI for necessary adjustments.

robot's wheels. Default Modes presents five images, representing the default angles of the wheels. Users can select their preferred image and click the 'Run' button to transmit the angle data to the robot, thereby initiating the movement of the wheels to the desired location. This feature is designed for quick and simple adjustments. The sixth image enables the manual modification of each wheel's angles and robot's linear speed for long-distance straight movements, providing flexibility for users to set the desired speed based on their specific requirements. Furthermore, an emergency 'Stop' button is included to immediately halt the robot in critical situations.

- **Robot Telescopic Arm Control:** This section focuses on the control of the robotic arm and its attached equipment. A set of buttons is designed to restore the motors to their initial positions, which is especially beneficial during startup or while reconfiguring the arm. Additionally, the user can precisely designate the position of each arm motor by inputting numerical values. Once the settings are applied and the execute button is pressed, the arm moves to the specified position. This functionality is ideal for accurate operations. Three toggle buttons are provided for the operation of the auxiliary equipment attached to the robotic arm. These three toggle buttons serve specific functions: they allow precise positioning of the laser pointer, activate the magnet for holding metallic parts, and enable the lighting system to improve visibility in low-light environments. Additionally, the 'Install Runner' button is used to activate the telescopic arm mechanism. At this stage, the arm's motors elevate the runner towards the ceiling, guaranteeing its adherence to the surface and proper positioning. Upon installation, the tools shoot fasteners to affix the runner to the ceiling. After completing the operation, the motors automatically stop for safety and efficiency.

The robot's mobility system is a crucial element of its design. The robot's mobility system, featuring four driven wheels, enables autonomous navigation on construction sites. Each wheel moves linearly and rotates about its vertical axis, providing omnidirectional motion for enhanced maneuverability in complex interior spaces.

A belt and pulley system has been integrated to transfer power from the motors to the wheels. This mechanism guarantees accuracy and increased torque transfer, allowing the robot to function efficiently in various conditions.

A spring-damper system is crucial for enabling smooth and steady robot movement in construction sites with sloped or uneven surfaces, as it absorbs undesirable vibrations during transportation and panel installation. Thus, the robot is equipped with four hydraulic shock absorbers, located next to the wheels due to space limitations for installation.

In addition to the mechanisms that are specifically designed for the working environment, the robot must be equipped with sensors that are capable of detecting obstacles. These sensors measure the distance to obstacles to avoid accidents, ensuring safe and efficient operation in dynamic construction sites. As shown in Figure 1, the panel installation robot follows a designed operation procedure to guarantee the precise and efficient positioning of panels. The procedure begins with the identification of the panel installation position through the use of an image processing system and cameras.

Following the determination of the position, the robot employs its omnidirectional mobility system, which is managed by advanced control systems to guarantee smooth and accurate movement, to navigate to the desired destination. Additionally, sensors are integrated into the system to guarantee safety and prevent collisions during operation. Finally, the telescopic arm, equipped with advanced installation tools, precisely attaches the panels in place. This integrated design optimizes the quality of the project, reduces construction time and costs.

2.2 Implementation of the Robot

Wall Bot is equipped with advanced mechanisms, including mobility systems, a robotic arm, and installation tools, enabling high-speed and precise installation of building panels. Figure 2 illustrates the robot's entire structure and its mechanisms. These mechanisms, enable transportation through difficult pathways, adjust angles and heights, and install panels with accuracy and reliability. The integration of advanced control systems with precise positioning capabilities and an efficient user interface enhances the accuracy and efficiency of the installation process.

A detailed explanation of its main mechanisms is provided below:

2.2.1 Robotic Arm and Motion Mechanism

The Cartesian robotic arm is an essential component of the robot, capable of autonomously positioning, and installing panels at diverse heights and angles. This arm is designed to accurately adjust its position, guaranteeing efficient and precise operation throughout the panel installation procedure. The following information describes the primary mechanics of this arm:

- **Telescopic Mechanism:** The robotic arm is equipped with an advanced telescopic mechanism that enables variable adjustment of height and depth





Figure 3. Detailed View of the Telescopic Mechanism in the Robotic Arm

of motion. To achieve precise and linear movement, a stepper motor with a torque of 85 kg·cm is employed. This motor, in conjunction with a ball screw featuring a 16 mm diameter and a 5 mm pitch, provides highly accurate and controlled linear motion (see figure 3).

- **Motion Mechanisms:** The robotic arm utilizes advanced mechanisms, including stepper motors and ball screws, to achieve accurate vertical and horizontal motions. This mechanism employs a stepper motor with a torque of 45 kg·cm, paired with a ball screw that has a diameter of 16 mm and a pitch of 5 mm. This combination facilitates the accurate and delicate motions, enabling the arm to properly install panels in various positions.
- **Rails and Connections:** Various components of the robotic arm affixed to linear rails and carriages, facilitating the arm's vertical and horizontal motions. These mechanisms enable the arm to transport panels along different paths with accuracy.
- **Panel Gripper:** The robotic arm is equipped with an advanced panel gripper responsible for holding and stabilizing panels during installation operations. This gripper can position panels in various orientations and, with the help of linear actuators and stepper motors, precisely adjust their angles before securely installing them in the desired location. The gripper's design prevents damage to the panels during handling and ensures a safe and highly accurate installation process.
- **Angle Adjustment Mechanisms:** The robotic arm is equipped with an advanced mechanism comprising a linear actuator and a metallic linkage. The linear actuator is responsible for moving the arm forward and backward, while the metallic linkage is connected on one side to the robot's body and on the other side to the telescopic arm structure. These two mechanisms can operate either synchronously or independently. When operating in synchronized mode, the robotic arm moves directly forward and backward without any angular deviation. However, if the two mechanisms operate at different speeds or directions, a positive or negative angle relative to the

vertical axis is generated, allowing the arm to be adjusted at various angles. This mechanism, by combining horizontal and vertical movements, enables the arm to achieve a wide range of angular adjustments, ensuring precise positioning of the panels. The angular adjustment mechanism of the robot is particularly effective for installing panels in locations where direct installation is not feasible due to obstacles on the ceiling or floor of the building.

2.2.2 Installation Tools

The accurate and efficient operation of installation tools is one of the key features of Wall bot. These tools must operate precisely and impeccably in many construction areas, including uneven surfaces, varying heights. The robot's installation tools are specially designed for various conditions.

- **Electric Drills and Nail Guns:** The robotic arm is equipped with two types of installation tools, enabling the installation of panels on ceilings, walls, and floors.

1. Installation on Metal Surfaces: For metal ceilings, the system employs an electric drill using self-tapping screws. These screws automatically install the runner (the panel support structure) to the metal ceiling without the need for additional pre-drilled holes.
2. Installation on Concrete Surfaces: The robot employs a nail gun system for concrete ceilings. This tool accurately embeds nails into the concrete substrate to firmly fasten the panels. The nail gun technology functions similarly to industrial nailers, facilitating rapid and secure installation of panels on robust and resilient ceilings. Furthermore, each tool is affixed to a linear system that enables positional adjustment according to the unique requirements of the ceiling. This capability allows the robot to do installations with enhanced precision at any location on the ceiling, considering its type and condition.

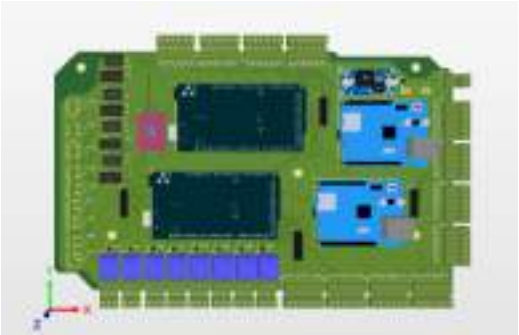
2.2.3 Cameras for Runner Positioning and Sensors

The runner's location on the ceiling and the robot's movement across obstacles are achieved by two complimentary techniques:

- **Smart Cameras**: Cameras mounted to the robot capture the surrounding environment and transmit this data for image processing. This processed data directs the robot to the precise area for panel installation by determining the runner's position on the ceiling. The robot is equipped with RGB cameras
- and 3D depth cameras, crucial for analyzing the geometric characteristics of the surroundings. The robot's cameras map the area to identify obstacles and determine safe, efficient pathways.
- **Sensors:** The robot's operation in construction environments necessitates the top priority of preventing collisions with obstacles and ensuring safety. The robot's positioning system is designed to autonomously detect obstacles and modify its trajectory using cameras and supplementary sensors. LiDAR (Light Detection and Ranging) sensors provide intricate 3D mapping of the surroundings, allowing the robot to identify obstacles, measure distances, and move securely. These sensors augment the robot's spatial awareness and enable accurate mobility in congested areas. Moreover, the device sends audible and visual alerts to the operator upon the detection of a possible threat. This capability is vital in construction settings, where many impediments and risks may exist. Through the integration of cameras and sensors, the robot has an advanced safety system, guaranteeing secure functionality in dynamic and difficult conditions.

3 Electronic Circuit Board of Wall-Bot

Wall-bot is an advanced machine for drywall installation. The robot must move along the surface and install each panel at the desired positions. This robot has a total of 12 stepper motors, 4 DC motors, 2 drills, and several additional sensors, including two distinct control systems. A Control Board have been designed to make the connections between the microcontroller boards, Ethernet connections, sensors and actuators (see figure 4). The board comprises two distinct Atmega2560 processing boards integrated to facilitate connections. Two boards connect to each other over the Ethernet Cables though a router and then connect to a PC with another Ethernet Cable. The board is required a 12V Main Power supply to run with Maximum current of 10 Amps. A buck switching converter transforms the power into the requisite 5V for the majority of digital components. Given the presence of two control boards within the unit, each board can be reviewed separately.



3.1 Main control for the movement

Four DC motors are coupled to BTS7960B external drivers, each connected with the board via 2x4 IDC connectors. These DC motors are mounted on four stepper motors, which rotate the DC motors to enable the robot to move in any direction across the surface. The stepper motors are controlled with TB66xx series drivers, with each driver connected through a 6-pin, 0.2-inch pitch terminal socket. To accurately measure motor speed, AS5600 Magnetic Angle Encoder are used to precisely measure the motor speed of all tires. The Encoder link to the Arduino processor via 2x3 IDC connectors, with the connection facilitated by the TCA9845 multiplexer. This design guarantees accurate speed measurement and effective motor control. The robot's movements are controlled by a singular ATmega2560 Arduino processor, which implements the control algorithms and manages communication. Commands are sent from the primary PC by a W5100 Ethernet module, facilitating uninterrupted and instantaneous control of the robot's motions. The robot is equipped with advanced communication components, stepper motors, and DC motors, which are incorporated to enhance its adaptability, precision, and mobility for intricate positioning tasks.

3.2 Main control for the installation unit

This unit consists of five actuators driven by eight stepper motors, nine proximity sensors for positioning, and many relays for the control of auxiliary equipment such as drills and fans. The system is controlled by an ATmega2560 Arduino board, which regulates motor function and sensor feedback for accurate operation. The main PC and a router are connected through an Ethernet module in this device, which allows for real-time monitoring and control. Buffer gates and ULN2xxx driver circuits are used to guarantee reliable connection and avert interaction problems among components. These devices provide electrical stability, hence augmenting the system's resilience and reducing possible interference among components. The seamless integration of actuators, sensors, and communication modules guarantees accurate performance, making the unit very efficient for intricate operations. The incorporation of gate ICs in electronic board design offers substantial protection for the primary processor by reducing potential damage. Nevertheless, to attain full electrical isolation and protection against electrostatic discharges (ESDs) and electrical shocks, optocouplers are utilized. These components guarantee strong isolation and improved system reliability, especially in electrically noisy settings. One of the key advantages of this design is the implementation of a dual-processor system. By employing two processors simultaneously, conflicts between movement control and the primary tasks of the robot are effectively prevented, resulting in smoother operation and task execution. A further advantage is the reduction of interference among I²C sensors, since the design has a specialized I²C module. Furthermore, all proximity sensors are completely isolated, guaranteeing precise sensor functionality and reducing the risk of signal disruption. The board is equipped with robust protection mechanisms on the power lines, including safeguards against overvoltage, overcurrent, and polarity mismatches. These protections improve its durability and resilience under challenging operating conditions.

Table 1 Denavit-Hartenberg (D-H) parameters of the robotic arm.

parameter	$\boldsymbol{\alpha_{i-1}}$	$\boldsymbol{a_{i-1}}$	$\boldsymbol{\theta_i}$	$\boldsymbol{d_i}$
i=1	0	0	0	d_1
i=2	0	0	0	$d_2 + l_1$
i=3	0	0	θ_1	l_2
i=4	$-\frac{\pi}{2}$	0	0	$d_3 + l_3$

4 Kinematic Modelling for the Telescopic Arm

For both forward and inverse kinematic analysis, this section models the telescopic arm as a 4-DOF robotic arm with three prismatic joints and one revolute joint, neglecting the fifth degree of freedom, which enables the angle adjustment of the telescopic arm, to simplify the analysis.

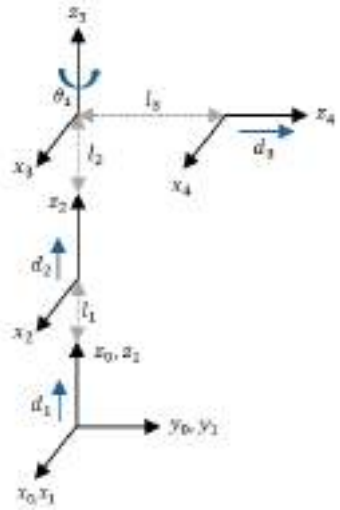


Figure 5 Assigned coordinate frames for each link of the robotic arm

4.1 Forward kinematics model

Forward kinematics calculates the position and orientation of the end-effector in 3D space by using the joint variables, such as joint angles or linear displacements. This calculation is performed by applying homogeneous transformation matrices that describe how the robot's joints and links are connected [10]. The Denavit-Hartenberg (DH) convention is a widely used approach for modeling robotic arms. It defines four parameters to represent the kinematic chains of robotic arms: twist angle, link length, link offset, and joint angle. Fig 5 depicts a coordinate frame attached to each link of the robotic arm. Table 1 lists DH parameters for the telescopic arm. The homogenous transformation matrix (see Equation (2)) that relates the global coordinate system is derived from Equation (1):

$$T_{06} = T_{01}\, T_{12}\, T_{23}\, T_{34}\, T_{45}\, T_{56} \tag{1}$$

$$T_{04} = \begin{bmatrix} \cos(\theta_1) & 0 & -\sin(\theta_1) & -\sin(\theta_1) * (d_3 + l_3) \\ \sin(\theta_1) & 0 & \cos(\theta_1) & \cos(\theta_1) * (d_3 + l_3) \\ 0 & -1 & 0 & l_1 + l_2 + d_1 + d_2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2}$$

Based on the transformation matrix, five points within the range of motion for each joint, including the initial and final points, were analysed. The position values for the selected points were calculated using MATLAB code and are presented in Table 2.

4.2 Inverse kinematics model

Inverse kinematics is the method of calculating the joint parameters required for a robotic arm to reach a desired position and orientation, functioning as the reverse of forward kinematics. There are two methods for solving inverse kinematics: the geometrical method and the algebraic method. The algebraic method is more complex than the geometrical method, as it involves solving a set of mathematical equations derived from the robot's kinematic chain. The joint values for the telescopic arm are determined from following equations using the geometrical method shown in Figure 1. θ_1 is derived from Equation (3), and d_3 is derived from Equation (4). To determine d_1 and d_2 from the equation, one of the values must be predefined by the operator. It is recommended to initially set d_2 at its maximum displacement (80 cm), and then derive d_1 from Equation (5).

$$\theta_1 = \tan^{-1}(-\frac{x}{y}) \tag{3}$$

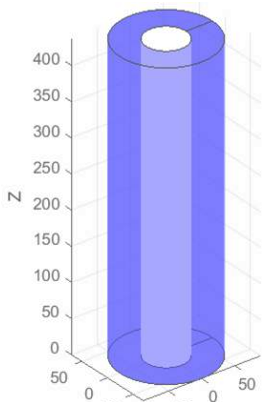
$$d_3 = \sqrt{x^2 + y^2} - l_3 \tag{4}$$

$$d_1 + d_2 = z - l_1 - l_2 \tag{5}$$

The telescopic arm has vertical movement along the Z-axis, while the horizontal link rotates along the Z-axis. Additionally, the radius of the installation tools is adjustable. The combination of these movements results in the installation range forming a hollow cylinder, as shown in figure 6.

5 Conclusion

This study demonstrates that the developed robot has the potential to revolutionize the construction industry, particularly in panel installation. The robot enhances efficiency, minimizes human errors and physical strain, and guarantees superior quality and precision in operations by automating the installation procedure. The advanced image processing technology integrated into the robot not only improves accuracy in panel placement but also highlights its high flexibility in installing panels of diverse sizes and weights. The robot's standout features, such as precise environmental analysis, the design of flexible arms, and coordinated movements enable it to effectively perform complex tasks in dynamic and challenging environments. Moreover, its ability to interact and coordinate with other robots facilitates the execution of larger projects with greater efficiency and speed. Future advancements in machine learning algorithms could enhance the robot's decision-making capabilities and adaptability to diverse conditions. Energy sustainability is expected to be incorporated into these systems in the future. Optimizing the robot's energy consumption through advanced energy management technologies, such as intelligent consumption control algorithms, can maximize its operational efficiency. Additionally, integrating the robot with renewable energy sources would enable more



$l_1 = 40\ cm$
$l_2 = 18\ cm$
$l_3 = 30\ cm$
$d_1 < 300\ cm$
$d_2 < 80\ cm$
$d_3 < 40\ cm$

Figure **6** The installation range of the robotic arm, formed as a hollow cylinder

Table 2 Calculated position values for five selected points within the range of motion for each joint

parameter	x (cm)	y (cm)	z (cm)
$d_1 = 0,\quad d_2 = 0$ $d_3 = 0,\quad \theta = 0$	0	30	58
$d_1 = 75,\quad d_2 = 20$ $d_3 = 10,\quad \theta = 7.5^{\circ}$	-5.2	39.66	153
$d_1 = 150,\quad d_2 = 40$ $d_3 = 20,\quad \theta = 15^{\circ}$	-12.94	48.3	248
$d_1 = 225,\quad d_2 = 60$ $d_3 = 30,\quad \theta = 22.5^{\circ}$	-22.96	55.43	343
$d_1 = 300,\quad d_2 = 80$ $d_3 = 40,\quad \theta = 30^{\circ}$	-35	60.62	438

sustainable and environmentally friendly project implementation. Furthermore, it is expected that the integration of robotics with 3D printing technology will be a key focus of future advancements. This integration could provide opportunities for designing and constructing complex building components. With 3D printing, it would become possible to produce parts that are precise, fast, and minimally wasteful, while robots could be equipped to install and assemble these components with high accuracy. This future development is anticipated to reduce costs, expedite project schedules, and align with sustainable construction standards. In terms of social aspects, this technology reduces laborious and repetitive tasks, allowing human workers to focus more on the creative and strategic aspects of projects. This shift not only enhances job satisfaction but also emphasizes the necessity of implementing safety standards and professional ethics to increase human-robot interaction. The Wall-bot, with its distinctive features, provides an innovative shift in civil engineering techniques. Utilizing this innovation may optimize project schedules, improve building quality, and foster robust and sustainable infrastructure, so offering greater prospects for the construction sector. Due to continuous developments, these robots are anticipated to function with enhanced speed, precision, and efficiency, significantly influencing the future of construction.

6 References

[1] Charehzehi, A., Chai, C. S., Yusof, A. M., Chong, modeling in construction conflict management. International Journal of Engineering Business Management, 9:1–12, 2017.

[2] Asadi, E., Li, B., and Chen, I.-M. Pictobot: A cooperative painting robot for interior finishing of industrial developments. *IEEE Robotics & Automation Magazine*, 25(2):82–94, June 2018

[3] B.R.K. Mantha, B. García de Soto, C.C. Menassa, and V.R. Kamat, “Robots in indoor and outdoor environments,” Chapter 16. In A. Sawhney, M. Riley & J. Irizarry (Eds.). Construction 4.0: An Innovation Platform for the Built Environment (pp. 307-325). 1st Edition. London: Routledge, 2020.

[4] Choi, J. H., Lee, K. Y., & Han, C. S. AUTOMATION SYSTEM FOR CURTAIN WALL INSTALLATION Ⅱ (ASCI Ⅱ). International Symposium on Automation and Robotics in Construction (ISARC), Seoul, Rep. Korea, 2006.

[5] Lee, K. Y., Lee, S. Y., Choi, J. H., Lee, S. H., & Han, C. S. The application of the human-robot cooperative system for construction robot manipulating and installing heavy materials. *SICE-ICASE International Joint Conference* ,Busan, Korea, 2006.

[6] van Gassel, F., Schrijver, P., & Lichtenberg, J. Assembling Wall Panels with Robotic Technologies. *Proceedings of the 23rd ISARC*, Tokyo, Japan, 2006.

[7] Han, C. S., Lee, S. Y., Lee, K. Y., & Park, B. S. A Multidegree-of-Freedom Manipulator for Curtain-Wall Installation. *Journal of Field Robotics*, 23(5), 347–360, 2006.

[8] Lee, S., & Moon, J. I. Case Studies on Glazing Robot Technology on Construction Sites. *Proceedings of the 32nd ISARC*, Finland, 2015.

[9] Lee, S., Lee, J., Han, C., and Lee, S. "Human Robot Cooperative Control and Task Planning for a Glass Ceiling Installation Robot." *Proceedings of the 25th International Symposium on Automation and Robotics in Construction (ISARC-2008)*, Vilnius, Lithuania, June 26-29, 2008.

[10] Kruthika, K., Kumar, K. B. M., and Lakshminarayanan, S. "Design and Development of a Robotic Arm." *Proceedings of M.S. Ramaiah University of Applied Sciences*, Bangalore, India.

Path planning for LiDAR equipped UGV considering coverage and efficiency in building inspection

Yun Cao[1], Ying Lo[1] and Cheng Zhang[1]

[1]Department of Civil Engineering, Design School, Xi'an Jiaotong-Liverpool University, Suzhou, China
yun.cao23@student.xjtlu.edu.cn y.lo18@student.xjtlu.edu.cn cheng.zhang@xjtlu.edu.cn

Abstract –

Construction quality inspection plays a crucial role in ensuring safety and quality at all stages of a project. However, traditional quality inspection has certain limitations, such as time consuming, expensive equipment, and cumbersome processes. Therefore, improving inspection technologies and processes by utilizing Unmanned Ground Vehicle (UGV) equipped with relatively low-cost LiDAR to enhance data collection coverage and efficiency during inspections has become a pressing issue in current construction quality management.

The main objective of this paper is to explore methods for enhancing the data acquisition capabilities of UGV equipped with low-cost LiDAR in construction environments and to experimentally verify the impact of path planning methods on data accuracy. The research focuses on investigating a new algorithm that simulates LiDAR scanning by UGV in point cloud space through 3D modeling and iteratively selects paths using Genetic Algorithm (GA) to preserve the path with the highest point cloud data collection efficiency. The experimental results extract the coordinates of the selected paths based on different scenarios, providing data support for the application of UGV on construction sites and promoting the adoption of intelligent building inspection methods.

Keywords –

Building Inspection; Unmanned Ground Vehicles; Path Planning

1 Introduction

Building inspections are critical to ensuring construction safety and engineering quality within the built environment. This process not only directly affects the structural stability of buildings but also effectively reduces safety hazards during construction, thereby safeguarding lives and property [1]. However, traditional building inspection methods, such as supervisory inspections, third-party testing, self-inspection, and non-destructive testing techniques, still face numerous limitations in practical applications. These limitations include lengthy inspection cycles, cumbersome and complex equipment, and subjective evaluation results that lack consistency and comprehensiveness [2,3]. Furthermore, these methods are constrained by technological means and inspection scope, making it difficult to adequately address potential issues during construction, thereby limiting their applicability in complex construction environments. These shortcomings may adversely impact construction progress to some extent. Therefore, optimizing inspection technologies and processes to improve the efficiency and accuracy of building quality inspections has become an urgent issue in the field of construction quality management [4].

Today, the development of building quality inspection technologies demonstrates a significant trend toward informatization, intelligence, and automation [5]. The use of robots for detecting and monitoring construction environments has gradually become a research hotspot [1]. Their application plays a critical role in reducing human error and improving the efficiency of inspection tasks. Common unmanned platforms include UGV, which exhibit significant payload and functional advantages, enabling them to carry larger-scale inspection equipment [6]. This includes cameras for visual data collection, LiDAR and laser scanners for 3D data acquisition, ground-penetrating radar (GPR) for behind-wall and underground detection, as well as components like inertial measurement units (IMUs) with ultra-wideband (UWB) sensors and global positioning systems (GPS) [7], which play an essential role in positioning and navigation. Among these, LiDAR, with its exceptional ranging accuracy and reliability, has become an essential tool in the architecture, engineering, and construction (AEC) industry, showing promising applications in environmental modeling and mapping [8]. Specifically, comparing the dimensions, positions, and orientations of building components detected on-site with the design parameters in Building Information Modeling (BIM) can effectively identify discrepancies with the BIM model [9].

However, the noisy construction environment significantly interferes with LiDAR scanning, making it a key challenge to achieve high-quality point cloud data acquisition in complex and cluttered environments [10]. Recent research has mainly focused on optimizing trajectory planning for unmanned platforms to enable efficient and high-precision data acquisition. These methods take into account factors such as energy efficiency, planning time, data redundancy reduction, and adaptability. They aim to address challenges such as minimizing energy consumption, reducing travel time, and maximizing data collection efficiency [11]-[13]. However, research focusing on solving path planning problems through path optimization algorithms to enhance data collection coverage remains relatively limited. This limitation constrains the depth and scope of theoretical advancements in this field.

2 Literature review

In order to achieve efficient environmental perception and navigation, path planning algorithms play a crucial role in this process. Traditional path planning algorithms are categorized into cell decomposition methods, roadmap-based methods, graph search-based path planning methods, and sampling-based path planning algorithms [14]. In both cell decomposition and roadmap methods, the shortest path between the starting and target positions is often determined using graph search algorithms. Graph-based path planning methods typically rely on environmental maps as prior input information, making them suitable for static planning scenarios. Among these, Dijkstra's algorithm is one of the most well-known graph traversal search algorithms, designed to find the shortest path between two nodes in a graph [15]. However, when the search space is large, the computation time of Dijkstra algorithm increases significantly. To address this, Peter E. Hart proposed the A* algorithm based on Dijkstra algorithm [16]. As a heuristic search method, A* is widely used in path planning for autonomous driving and robotics. By introducing a heuristic approach to guide the search towards the target node, A* effectively reduces the number of graph nodes explored, thus accelerating the search process. Through comparison with Figure 1, it is evident that the A* algorithm reduces the computational load compared to the Dijkstra algorithm.

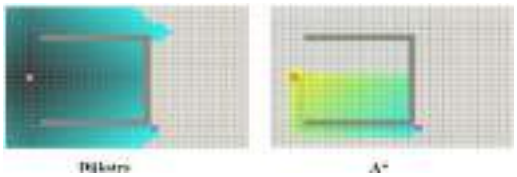


Figure 1. Comparison of A* and Dijkstra [16]

In addition, to adapt path planning algorithms to different application requirements (such as dynamic planning, variable target nodes, and path angles) [17], improving computational efficiency is crucial. The A* algorithm performs exceptionally well in this regard, making it particularly advantageous in static environments [16]. In recent years, researchers have worked on improving the A* algorithm to overcome its limitations.Wang et al. proposed an enhanced version of the A* algorithm that takes into account vehicle contour, speed, and dynamics constraints, thereby providing safer and smoother path planning for self-driving cars [18]. Bao et al. introduced a dynamic weight-adjusted 16-way search method that reduces redundant search nodes, thus improving path planning efficiency. These improvements aim to optimize path lengths, reduce turn times, and improve overall performance [19]. Although the A* algorithm still faces many challenges in practical applications, it is applicable to more and more application scenarios as researchers continue to improve it, and still plays an indispensable role in path planning.

However, most of the research on path planning algorithms is used to perform path navigation, obstacle avoidance, etc. Nguyenet al. proposed a new path planning algorithm, NSPMR (Near-Shortest Path for Mobile Robot), based on the improvement of the traditional local path planning algorithm [20]. It is specifically designed to solve the path optimization and obstacle avoidance problem for mobile robots (MRs) when navigating in known or unknown environments. The algorithm generates an approximate shortest path by integrating sensor data in real time, which effectively avoids the limitations of traditional algorithms in the problems of "local minima" and "infinite loop".Dong et al. proposed a combination of the optimized A* algorithm and the dynamic window algorithm to optimize the path of a mobile robot in a known or unknown environment [21]. A* algorithm and Dynamic Window Algorithm (DWA) to solve the global optimization and dynamic obstacle avoidance problems in path planning for automated guided vehicles (AGVs) [21]. The optimized A* algorithm improves the traditional algorithm's deficiencies in path smoothness and computational efficiency by means of adaptive heuristic functions and redundant point deletion, and is capable of achieving globally optimal path planning in static environments. Baras et al. proposed a Covered Path Planning (CPP) optimization method for UGV to conserve energy with a focus on Reducing the number of UGV turns through area allocation, and significantly improving the energy utilization through path optimization [22]. In the paper, the environment is partitioned and the boundaries are finely adjusted to reduce the region mixing phenomenon, while the node exchange mechanism is introduced to optimize the

generated coverage paths, thus reducing the number of path turns and energy consumption. Experimental results show that the method outperforms traditional algorithms, such as A*-DARP+STC, in different obstacle ratios and complex terrains, and exhibits significant advantages in terms of the number of turns and energy efficiency. The study provides a more efficient solution for path planning of multi-UGV tasks and lays a theoretical foundation for energy-aware path planning in dynamic environments.

These studies typically emphasize minimizing path length and ensuring driving safety, often overlooking the specific requirements for scan coverage and data collection accuracy in construction sites. As a result, the adaptability of existing path optimization algorithms in construction environments remains insufficient. The main difference between this study and previous BIM-based robot path planning methods is that our approach is specifically designed to address the issue of data loss often encountered by unmanned vehicles during point cloud collection due to the noisy construction site environment. The goal is to simulate the collaboration between UGV and LiDAR and optimize the path through algorithmic iterations until the path that allows the UGV to collect the most complete point cloud model is obtained, along with its corresponding path coordinates.

3 Methodology

The method of this study aims to improve the coverage and efficiency of 3D point cloud data collection using unmanned platforms equipped with LiDAR sensors. As shown in Figure 2, the process begins with LiDAR accuracy assessment, where the impact of path planning methods on point cloud accuracy is validated. The next stage is path planning via algorithm, which consists of two parts: algorithm execution and model preprocessing. Preprocessing includes converting the BIM model into a point cloud model and adjusting the coordinate system to ensure data alignment and analysis.

After preprocessing, the feasible path for the UGV is first generated using the A* algorithm. Then, the path points undergo crossover mutation, and the GA is executed to generate the path with the highest number of collected point cloud points, the data processing stage begins, where the path length and point cloud quantity are standardized, unifying the units of path length and point cloud quantity to facilitate scoring and selecting the optimal path required for the experiment. Finally, the path is evaluated using a scoring mechanism to select the final optimized path.

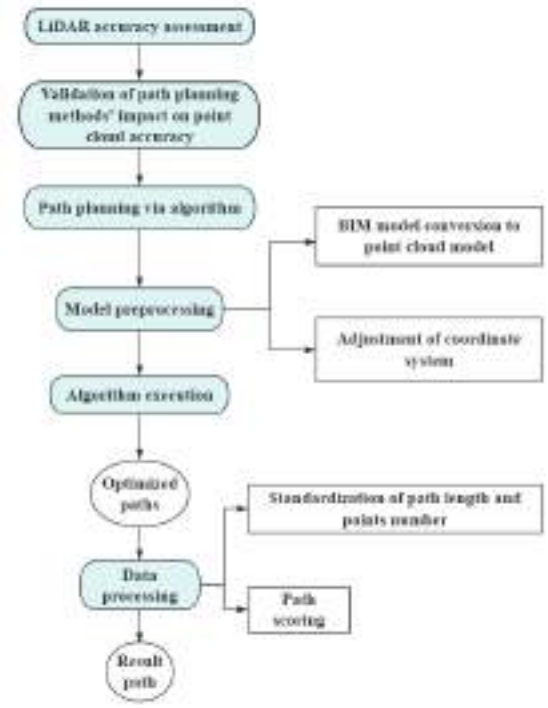


Figure 2. Proposed methodology

To verify the impact of path planning methods on the accuracy of point cloud data, experiments were conducted in both structured and unstructured scenarios (as shown in Figure 3). The experiment in the structured scene was conducted to obtain the accuracy of the UGV when scanning with the mounted LiDAR. Point cloud data was collected along different path types (2 meters, 4 meters, 6 meters, and 8 meters) using both an UGV and a Terrestrial Laser Scanner (TLS), with the high-precision point cloud data collected by the TLS serving as the baseline. After denoising, the dimensions of the target point clouds collected by the two devices were calculated using the C2C (cloud-to-cloud distance) method to ultimately obtain the scanning accuracy results.

Based on the experimental results from the structured scenario, the optimal path for UGV data collection was planned in the unstructured scenario, which simulates the construction site environment where the UGV operates in practical applications, and point cloud data was collected along both the planned path and random paths. Data processing in the unstructured scenario also included noise removal and denoising, but further comparisons of the scanning accuracy for different path types were made to verify the effectiveness of path planning methods in improving UGV scanning accuracy in complex environments.

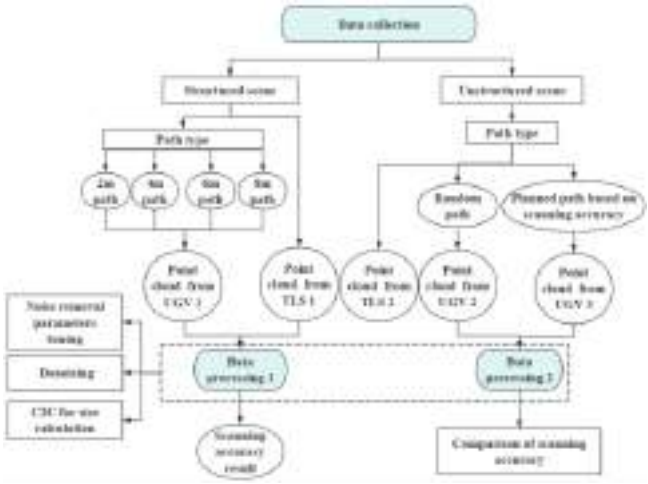


Figure 3. Experimental Workflow for verifying path planning impact on scanning accuracy

4 Case study

4.1 Theoretical verification phase

4.1.1 Mid360 precision exploration

According to existing studies, Lo et al. analyzed the data collection capability of a UGV equipped with the Mid360 in a structured scenario, with the data evaluated using the Level of Accuracy (LoA) rating. The results showed that at a distance of 2 meters, the measurement accuracy was the highest, with errors close to the sub-centimeter level, reaching an LoA30 rating, and the noise was almost negligible. In contrast, when the measurement distance increased to 8 meters, the measurement accuracy significantly decreased, with higher noise intensity and greater measurement errors, approaching 2 centimeters, and the rating dropped to LoA20. This indicates that as the measurement distance increases, the accuracy of LiDAR data is significantly affected. Additionally, at measurement distances of 2 meters and 4 meters, some point cloud data for the cardboard box were lost (as shown in Figure 4) [23].



Figure 4. Point cloud modeling for different distance conditions [23]

4.1.2 Data collection in unstructured scene

In the Civil Engineering Laboratory at university, the researchers selected a structural column as the target to evaluate the effectiveness of using a path planning method to enhance the point cloud data collection capability of an UGV equipped with the Livox Mid360 in an unstructured scenario. As shown in Figure 5, a closed-loop route was laid out with yellow tape in the underground laboratory (at a distance of 4-6 meters from the target column), the path of the UGV shown in

Figure 6. To control variables (data collection time), the UGV in this experiment maintained the same speed (0.6 m/s) and performed multiple data collections along both the planned path and random paths. A set of data files with similar file sizes was selected for accuracy comparison.



Figure 5. Column target in unstructured scene

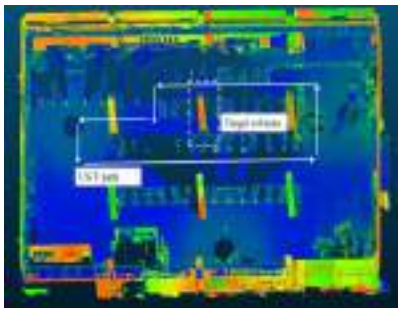


Figure 6. UGV path in unstructured scene

4.1.3 Data analysis in unstructured scenes

Table 1. Point cloud accuracy results for different paths

		TLS	Standard Deviation	UGV(2)	Standard Deviation	Error	UGV(3)	Standard Deviation	Error
Concrete Column	Width	0.608	0.0037	0.651	0.009	0.037	0.574	0.0151	0.032
	Length	0.613		0.627		0.046	0.578		0.035

The denoising process first requires separating the point cloud data of the target column collected by the TLS and UGV (with TLS data serving as the baseline) and measuring its width and length. The measurement results are shown in Table 1. It is worth noting that although the standard deviations are similar, the error range of the Livox Mid360 is slightly larger than the results of the 4m to 6m control distance experiment. This indicates that noisy environments can affect the accuracy of LiDAR data.

The point cloud data collected by the UGV shows slightly higher accuracy in path planning, but both datasets fall within the LoA20 class (Height measurements were not taken into account due to obstructions caused by the mechanical, electrical, and plumbing (MEP) systems on the ceiling). Furthermore, in the point cloud model collected along the random path, the length and width of the near-square target column were recorded as 0.651m and 0.627m, respectively, with an error of 0.24m. In summary, the data error collected by the UGV along the prescribed path was smaller than that collected along the random path.

4.2 Path planning phase

Building on this previous work, this section explores how to optimize path planning through algorithms. The focus of the research is on the global path planning problem for UGV in static environments, aiming to design a path that collects the most point clouds. During the path planning process, it is crucial to consider obstacles and danger zones in the environment, ensuring the path avoids impassable obstacles. At the same time, the kinematic and dynamic constraints of the Bunkerpro unmanned ground vehicle must be satisfied. Unlike traditional path planning methods that simplify the UGV to a particle model, this study takes more key factors into account, such as the platform's size and LiDAR placement height, in order to overcome the shortcomings of traditional methods that overlook motion constraints and environmental factors.

4.2.1 Model preparation

In practical applications, data collection usually relies on existing Building Information Models (BIM) to obtain relevant information about construction sites. Given this, the study selected a classroom and corridor model from the fifth floor of the Engineering Building at university campus for path planning. As shown in

Figure 7, the classroom model contains many obstacles (such as desks, chairs, and podiums), which is similar to the noisy environment of a construction site. The corridor model has a large coordinate range and a high point cloud density, which is similar to the point cloud data scale encountered in real construction sites. Both models are ideal choices for validating path planning and simulating LiDAR scanning effects.

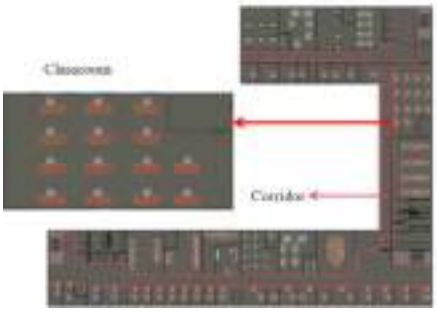


Figure 7. BIM models of classroom and corridor

In this study, the total number of points generated for the corridor and classroom models was set to 100,000, with the point cloud density set to 10 to ensure the preservation of model details. The generated point cloud models required further adjustments to ensure their dimensions and coordinates aligned with the actual building. To achieve this, the researchers imported the models into 3ds Max to adjust their coordinates and dimensions. The final result is shown in Figure 8.

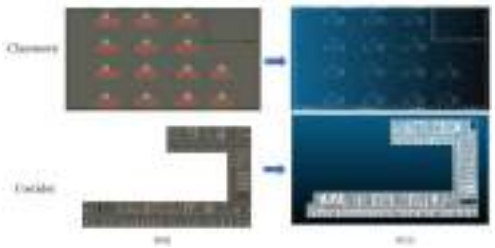


Figure 8. BIM models of classroom and corridor

4.2.2 Algorithm results

This algorithm aims to simulate data collection by an UGV equipped with LiDAR along a planned path. Due to the relatively small passable areas in the two models, the researchers set the grid coordinates to the size of the UGV (0.9m) to ensure smooth passage. The cloud-to-cloud algorithm was used to calculate the ground point cloud thickness of the model as 0.2m. In the algorithm, the parts at an elevation of 0.2m were designated as the ground for the UGV, and a feasible path map was generated (as shown in Figure 9). In addition, to ensure that the path does not collide with obstacles and that the start and end points are within valid areas, the constraint equations are used as follows:

$$g_1(x) = valid_path(x) = Ture \quad (1)$$

$$g_2(x) = start_point_valid(x) = Ture \quad (2)$$

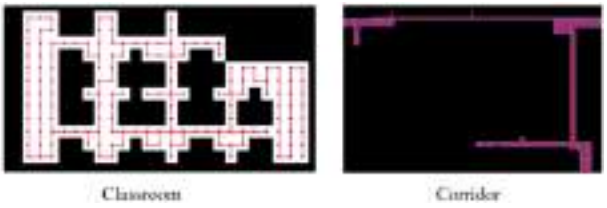


Figure 9. Feasible path map of classroom and corridor

It is important to note that during this process, the LiDAR simulates emitting rays within a specific angular range and checks where each ray intersects with the point cloud. In this study, the parameters of the Livox Mid-360 were used to set the LiDAR's field of view (fov=360°), scanning angle range (up_angle=52°, down_angle=7°), mounting height (h=0.7m), and measurement range (range_min=0.1m, range_max=70m). The rays are traced in each direction until they hit the nearest point in the point cloud. This setup ensures that the system does not detect point clouds behind obstacles, thus reducing noise interference.

In this algorithm, the objective function can be the number of point clouds covered by the path, that is, the number of valid point clouds scanned in the area traversed by the path. This reflects the optimization effect of the path and is defined as:

$$f_{objective}(\mathrm{x}) = \mathrm{coverage}(\mathrm{x}) \quad (3)$$

The researchers recorded the number of point clouds corresponding to the optimal path in each generation (as shown in Figure 10). During the simulation scanning process in the tenth generation of GA, the maximum number of point clouds was obtained, with values of 6115 and 30606, respectively. As shown in Line charts, the point cloud collection gradually stabilized in the later generations for both models. This indicates that the GA converged after several generations of optimization, and the model's performance became consistent. The number of point clouds collected reached the optimal state during the optimization process, confirming the success of the framework in this study. Furthermore, it is important to note that in GA's iterative process, it is possible for two generations to produce identical results. This happens because the fitness function of the offspring, after crossover and mutation, is lower than the optimal result of the parent generation.

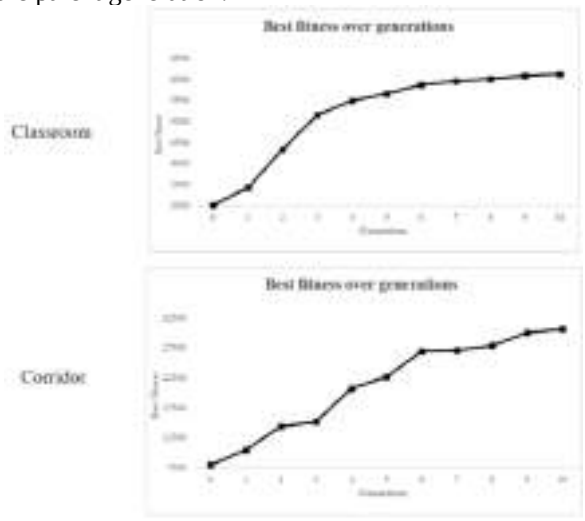


Figure 10. Convergence trends of GA

Figure 11 shows the coordinate information of the paths for the two models and the preview of the point cloud. The coordinate information of the path reflects the specific distribution of the optimal solution after optimization by GA in the coordinate space, while the point cloud preview displays the distribution of the point cloud data collected along the optimized path, providing a clear and intuitive reference.

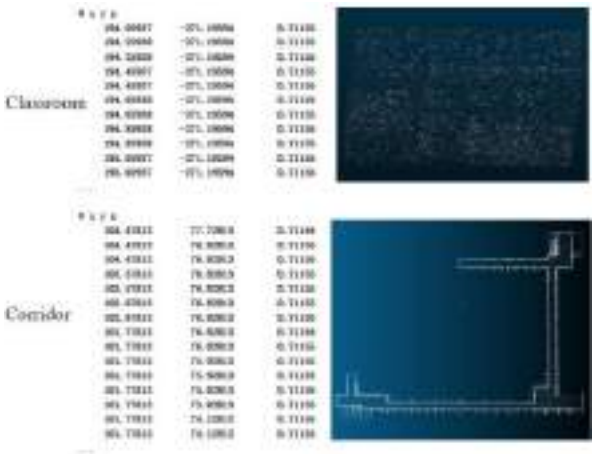


Figure 11. Path coordinate set and simulation models

Through approximate calculation, the longest path is 1417.09m and the shortest is 1203.32m. From the Figure 12, it can be seen that the path length does not show an obvious pattern of change and does not exhibit a significant convergence trend. Among them, the third-generation path, due to random variations, shows a significantly longer result compared to other generations. It can be attributed to the exploratory nature of the algorithm in the early stages of evolution, which is a common occurrence in the initial phase of GA optimization.



Figure 12. Changes in path length per generation

4.2.3 Data normalization

For point cloud path planning and scanning coverage optimization, the goal of data normalization is to bring different evaluation metrics to a unified dimension, allowing researchers to derive an optimal path based on the scoring rules they formulated.

Since the classroom model is relatively small, path length is not considered as a scoring factor for its optimal path. For the corridor model, researchers selected point cloud quantity and path length as scoring metrics, assigning weights of 80% and 20%, respectively. Path length, being a common optimization objective, means that shorter paths result in lower task costs (such as collection time, battery consumption of the autonomous vehicle, etc.).

The normalized calculation results and final scores are shown in Table 2, with the tenth generation path receiving the highest score. This indicates that the path in the tenth generation achieved the best balance during the optimization process, representing the optimal trade-off between path length and point cloud coverage. Therefore, the path from the tenth generation was selected as the optimal result for this experiment and output as the final recommended path.

Table 2. Scoring results for all paths

Generation	Best Fitness	Path length	Score
1	0.111	0.914	0.272
2	0.284	0.820	0.391
3	0.319	0.000	0.255
4	0.563	0.956	0.642
5	0.648	0.951	0.709
6	0.837	0.831	0.836
7	0.844	0.888	0.853
8	0.877	0.947	0.891
9	0.973	0.914	0.961
10	1.000	1.000	1.000

5 Discussion

This study first converted the BIM model of the fifth-floor corridor and classroom of the Engineering Building into a point cloud model. During this process, software limitations caused coordinate system confusion, requiring researchers to adjust coordinates and obstacle dimensions comprehensively using professional 3D modeling software.

Next, a path optimization algorithm (combining GA and A* algorithm) was developed in Python to maximize data collection by UGVs in three-dimensional space. The algorithm iteratively searched for the optimal path over multiple generations, eventually converging to the best path planning scheme, ensuring that the LiDAR covered all target areas as comprehensively as possible. Notably, the algorithm allows researchers to select the most suitable path based on specific needs. In the tenth generation of GA, path optimization reached its optimal state according to the weighted scoring system.

The experimental results indicate that during the path optimization process, the number of scanned point clouds increases continuously and eventually stabilizes. Researchers can flexibly choose the path from the stable results based on different experimental requirements.

Moreover, It is worth noting that the algorithm has certain CPU performance requirements and relatively low precision requirements for imported point cloud data, retaining only the essential key details. In practical applications, to improve efficiency and save time, it is recommended to limit the size of point cloud data to within 100 MB to meet practical needs while ensuring the high efficiency of the algorithm.

6 Conclusion

This study proposes an innovative path planning method that ensures UGVs can capture point cloud data more comprehensively in noisy construction environments, thereby improving model completeness. Additionally, by overcoming the limitations of existing tools, we elevate radar scanning simulation from the traditional 2D (such as the MATLAB Driving Scenario Designer) to 3D, and accurately track and record the spatial areas that can be captured within the scan range.

Experimental results show that this path planning method exhibits high adaptability, allowing adjustments based on different laser radar performance characteristics (such as scanning range and installation position). Moreover, the algorithm iteratively optimizes the path to minimize scanning blind spots, enhancing the coverage of the laser radar and consequently obtaining denser point cloud data, providing a more comprehensive data foundation for building inspections.

References

[1] Halder S, Afsari K. Robots in Inspection and Monitoring of Buildings and Infrastructure: A Systematic Review. Applied Sciences. 2023;13(4):2304.

[2] Hallermann N, Morgenthal G, Rodehorst V. Unmanned Aerial Systems (UAS) – Case Studies of Vision Based Monitoring of Ageing Structures2015.

[3] Pinto L, Bianchini F, Nova V, Passoni D. Low-Cost UAS Photogrammetry for Road Infrastructures' Inspection ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2020;XLIII-B2-2020:1145-50.

[4] Jiashen S. Exploring the influence and importance of engineering testing on quality control of construction projects. Engineering Management and Technical Discussion. 2022;4(16).

[5] Cong C. Brief Analysis of the Characteristics and Methods of Construction Engineering Inspection. Engineering Construction Technology. 2023;1(5):23-5.

[6] Lattanzi D, Miller G. Review of Robotic Infrastructure Inspection Systems. Journal of Infrastructure Systems. 2017;23:04017004.

[7] Asadi K, Jain R, Qin Z, Sun M, Noghabaei M, Cole J, et al. Vision-Based Obstacle Removal System for Autonomous Ground Vehicles Using a Robotic Arm2019. 328-35 p.

[8] Wang X, Pan H, Guo K, Yang X, Luo S. The evolution of LiDAR and its application in high precision measurement. IOP Conference Series: Earth and Environmental Science. 2020;502:012008.

[9] Ilyas M, Ying K, Nithish M, Xin J, Xinge Z, Cheah C. Robot-Assisted Object Detection for Construction Automation: Data and Information-Driven Approach. IEEE/ASME Transactions on Mechatronics. 2021;PP:1-.

[10] Kau L-J, Chiou L-J, Lo Y-H, Chen S-H. A Multi-Lidar-based Point Cloud Acquisition Platform and Data Fusion for Autonomous Vehicle in Complex Urban Environment2023. 1-5 p.

[11] Nolan P, Paley D, Kroeger K. Multi-UAS path planning for non-uniform data collection in precision agriculture2017. 1-12 p.

[12] Popovic M, Vidal-Calleja T, Hitz G, Chung J, Sa I, Siegwart R, et al. An informative path planning framework for UAV-based terrain monitoring. Autonomous Robots. 2020;44.

[13] Pan Y, Yang Y, Li W. A Deep Learning Trained by Genetic Algorithm to Improve the Efficiency of Path Planning for Data Collection With Multi-UAV. IEEE Access. 2021;PP:1-.

[14] Hua X. Research on path planning and motion control of unmanned platforms under unstructured roads [D]. Jilin: Jilin University; 2023.

[15] Wang S, Li A. Multi-adjacent-vertexes and Multi-shortest-paths Problem of Dijkstra Algorithm. Computer Science. 2014;41(6):217-24.

[16] Wang M. Mobile robot path planning based on A* algorithm [D]. Shenyang: Shenyang University of Technology; 2017.

[17] Baker ZK, Gokhale M, editors. On the Acceleration of Shortest Path Calculations in Transportation Networks. 15th Annual IEEE Symposium on Field-Programmable Custom Computing Machines (FCCM 2007); 2007 23-25 April 2007.

[18] Wang P, Liu Y, Yao W, Yu Y. Improved A-star algorithm based on multivariate fusion heuristic function for autonomous driving path planning. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering. 2022;237:095440702211006.

[19] Bao W, Li J, Pan Z, Yu R. Improved A-star Algorithm for Mobile Robot Path Planning Based on Sixteen-direction Search2022. 1332-6 p.

[20] Nguyen HT, Le HX. Path planning and Obstacle avoidance approaches for Mobile robot. ArXiv. 2016;abs/1609.01935.

[21] Dong Y, Cao J, Li G. Automatic Guided Vehicle Path Planning Based on A-Star and DWA Dual Optimization Algorithms. Science Technology and Engineering. 2023;23(30):12994-3001.

[22] Baras N, Dasygenis M. UGV Coverage Path Planning: An Energy-Efficient Approach through Turn Reduction. Electronics. 2023;12(13):2959.

[23] Lo Y, Cao Y, Zhou X, Zhang C, Chen M. Performance Analysis of Low-Cost Solid-State LiDAR for Building Elements Point-Cloud Mapping. ICiMR 2024; 22-23 August 2024, Taicang, China.

Mixed Reality Based Teleoperation System For Prefabricated Building Assembly

Haoran Wang[1], Shunjie Gu[1], Zuoqing Yang[1], Zeren Tao[1] and Qilei Sun[1,*]

[1]School of Advanced Technology, Xi'an Jiaotong-Liverpool University, China

Haoran.Wang21@student.xjtlu.edu.cn, Shunjie.Gu21@student.xjtlu.edu.cn, Zuoqing.Yang20@alumni.xjtlu.edu.cn
Zeren.Tao21@student.xjtlu.edu.cn, Qilei.Sun@xjtlu.edu.cn

Abstract -

In response to the challenges in the assembly process of prefabricated buildings and the pursuit of sustainable construction, this paper proposes an industrial robot arm teleoperation method based on mixed reality (MR) technology. Traditional assembly processes that rely on manual installation face challenges such as operator safety issues, and high energy consumption. Our proposed method utilizes MR technology to achieve immersive and remote control and uses Optimal Resolved Acceleration (ORA) and Resolved Acceleration Control (RAC) to complement the shortcomings of Resolved-Rate Motion Controller (RRMC) by introducing compliance control to enhance environmental adaptability and telemanipulation accuracy. The aluminum profile assembly and precision testing experiments have shown the feasibility and accuracy of this method. The MR teleoperation framework allows for compliance control, and despite some noise, the robotic arm exhibits good accuracy. Future work will focus on expanding task complexity, improving system accuracy, integrating intelligent functions, and exploring scalability for large-scale construction projects.

Keywords -

Mixed Reality; Teleoperation; Robotic arm; Prefabricated Building; Sustainable Construction

1 Introduction

In recent years, with the increasing global industry attention to sustainability, integrating low-carbon and environmental practices into the construction industry has become a focus. For example, the European Union continuously updates its Building Energy Efficiency Directive, requiring building projects to standardize carbon emission management from design to construction to operation and maintenance, and transform them into environmentally friendly projects. Industry and market demand are also driving the construction industry to transition towards low-carbon operations.

Prefabricated construction refers to transferring a large amount of on-site work from traditional construction methods to factories, where building components and accessories (such as floor slabs, wall panels, stairs, balconies, etc.) are processed and manufactured, transported to the construction site, and assembled and installed on site through reliable connection methods. Compared to traditional buildings, buildings using prefabricated components can reduce greenhouse gas emissions by 8.06%. In addition, using alternative green concrete materials in the prefabrication process of prefabricated buildings can reduce the total greenhouse gas emissions by 2.83-12.05% [1]. Xiang [2] also pointed out that the construction industry has caused a significant amount of carbon emissions worldwide. The adoption of prefabricated buildings has reduced the industry's environmental pressure. The comparison between prefabrication methods (such as assembly) and traditional in-situ casting methods shows that in all categories of influence, prefabrication methods have greater environmental advantages than traditional methods [3].

Although prefabricated structures have many advantages compared to traditional construction techniques, there are still many disadvantages in the assembly process. Due to differences in operational proficiency and on-site conditions, prefabricated buildings often and energy. In order to solve the above problems, this article proposes a teleoperation method for industrial robot arms based on mixed reality technology, which is used for assembly work in prefabricated buildings. This method enables us to detect hand displacement and angle deflection through MR devices while wearing them, and accurately control the robot arm to make responsive actions to complete assembly work. Compared to using cranes and manual assembly methods. This method can complete assembly work more accurately and quickly while maintaining low carbon emissions. In order to solve the challenge of realizing slosh-free assembly in the assembly process of prefabricated buildings and improve the efficiency and accuracy of the assembly work and the computational efficiency of the slosh-free assembly trajectory under the condition of real-time feedback, ORA and RAC with compliance control algorithm is proposed. This algorithm introduced the compliance matrix into the existing motion controller to make the robot arm perform the assembly task more

smoothly and naturally[4]. Ensure that the robot arm can not only meet the accuracy of assembly without shaking in the condition of human-computer interaction, but also reduce the vibration and impact caused by rigid movement, further improving the response ability to external forces. This method supports remote assembly by experienced workers, avoiding the cost increase and additional carbon emissions associated with personnel mobility. In addition, it can replace operators in some extremely dangerous situations and still accurately and effectively complete assembly work, ensuring the safety of construction personnel.

This article mainly consists of six parts: we begin by discussing introduction, followed by related work of this study. The methodology for implementing the article will be presented in the third section. The experimental verification of the technology and the analysis of the results are presented in the fourth and fifth parts, respectively. The final section includes a summary of this study and plans for future research work.

2 Related work

The integration of mixed reality (MR) technology enriches and enhances collaborative experiences by seamlessly merging digital and physical elements. For instance, MR systems empower operators to engage with both digital and physical environments, thereby reducing redundant hardware and maximizing workspace efficiency [5]. MR technology amalgamates the strengths of AR and VR, augmenting user immersion and facilitating interaction with holographic physical devices within the real environment. This enables humans to remotely operate robotic arms, significantly mitigating the environmental footprint associated with substantial human resource deployment. Moreover, immersive systems diminish the necessity for construction industry operators to travel, consequently reducing the carbon footprint linked to transportation. Case in point, the Australian company GRAINTECH has implemented Microsoft's HoloLens mixed reality technology to enhance remote assistance and collaboration, reportedly slashing travel expenses and carbon emissions by 68% [6]. This showcases the potential of MR technology in reducing the need for physical travel and lowering the carbon footprint. The MR remote control method can also prevent humans from working in extremely dangerous situations, improving worker safety and work efficiency [7].

MR remote-controlled robots have achieved considerable success in assembly tasks. Takahashi [8]proposed a virtual reality-based robot teaching interface, offering a user-friendly environment where novice operators can effortlessly instruct robots. Robots in the workplace replicate the actions of operators within a virtual workspace to perform assembly tasks. Jarecki and Lee [9] introduced a mixed reality (MR) user interface (UI)-based system device that supports the remote operation of mobile robotic arms. Their findings suggest that MR-based UIs may be perceived as more intuitive by operators compared to conventional control interfaces. Zhang et al. [10] integrated collaborative robots with mixed reality technology to achieve human-machine collaborative assembly, enhancing assembly efficiency and presenting a novel and viable assembly model for the intelligent collaborative assembly of large-scale and complex spatial products.

In the operation scenario of an industrial assembly line, the robot arm must be able to accurately control its motion trajectory and speed. Although the current research on gesture-based building robot control has been relatively abundant. However, in this dynamic environment, it is difficult to ensure that the assembly slosh-free when moving rapidly due to inertia. Humans need special skills to perform tasks efficiently without shaking. Therefore, how to balance the agility of the robot arm and the maneuverability in slosh-free, and improve the assembly accuracy and efficiency is an important challenge. We can improve the agility of the robot system by designing advanced control methods, so that the robot arm can achieve superhuman slosh-free motion. In addition, in the case of real-time feedback, the calculation of slosh-free trajectory requires a lot of computational resources, so the design of an efficient optimized solver is also crucial. To solve these problems, this paper proposes a solution based on conformance control algorithm ORA (Optimization trajectory algorithm) and RAC (real-time adjustment algorithm). The algorithm is improved on the basis of ORA and RAC, and the compliance matrix is introduced to ensure the high precision and efficiency of the robot arm during assembly. The tolerance of the robot arm to the rigid external force is enhanced, so that the worker's operation is safer and smoother. This approach greatly facilitates remote assembly by experienced workers, significantly reducing safety risks and the additional carbon emissions associated with too many workers.

3 Methodology

The methodology integrates Mixed Reality (MR) technology with advanced motion control algorithms to enhance robotic arm manipulation. It employs Transmission Control Protocol/Internet Protocol (TCP/IP) [11] for Unity [12]and xArm 6 communication [13], ensuring synchronized movements. Data transmission aligns position and rotation data, minimizing operational gaps. The system was first introduced by the Resolved-Rate Motion Controller (RRMC), improved with the Optimal Resolved Acceleration (ORA) algorithm for precision control in the context of assembly, and finally combined with compliance control for adaptability. The framework (as shown

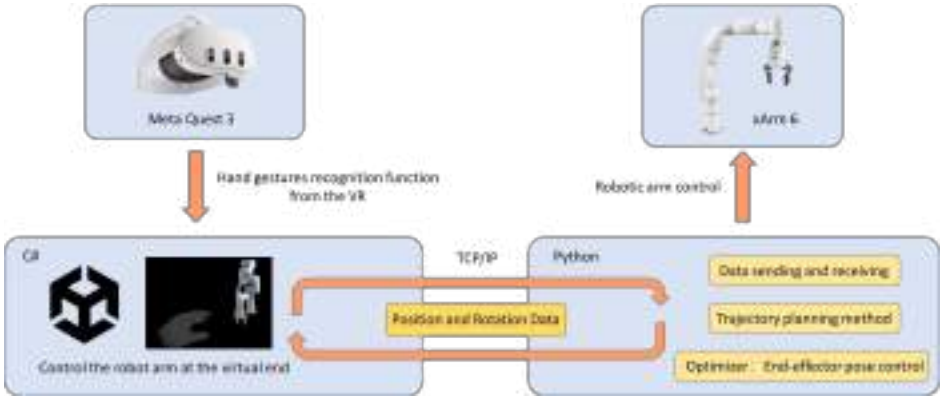


Figure 1. The workflow of the communication between Unity and xArm-6

in Figure 1) aims to improve assembly accuracy and efficiency while maintaining low carbon emissions, in line with sustainable building goals.

3.1 TCP/IP Connection

As depicted in Figure 1, the TCP/IP connection method was used in this study primarily to fulfil the communication function between Unity based on C # and xArm 6. Python implemented both the sending and receiving of control information, as well as the optimization functions of xArm 6. The objective was to transmit the data of hand recognition coordinates and attitude angles (position x, y, z, and rotation x, y, z) data provided by Meta Quest 3's SDK function in the VR headset and Unity scene to the xArm 6, with which a control connection had already been established, through the specified IP address and port, thus achieving the control of the real end (xArm 6) by the virtual end (Unity and VR). Meanwhile, the target angles of each link of the robotic arm could also be reversely transmitted to Unity through the PyCharm end to control the movement of the virtual xArm 6 model, thus realizing the control of the virtual end by the real end.

3.2 Data Transmission

The hand gesture recognition technology (i.e. pinch interaction) in VR equipment captures the spatial coordinates of the five fingers and the palm's center point. The palm's center point data is used to map the end-effector of the robotic arm, enabling precise control of its position and rotation. The pinching distance (spatial separation between fingertips) is directly mapped to the gripper aperture, where a smaller distance corresponds to a narrower gripper closure, providing intuitive control.

In the data transfer module, a calibration coefficient of 1000 is applied to position data in Python to align Unity's meter (m) units with Python's millimeter (mm) units. This ensures that the actual displacement values remain aligned between the two systems. For large industrial robotic arms, this coefficient can be adjusted to handle tasks like automobile assembly, where precise positioning of large components is required. The robotic arm's motion is controlled using coordinate and attitude rotation differences between consecutive frames, enabling accurate trajectory tracking based on real-time hand position and orientation changes.

3.3 Manipulator motion control and trajectory optimization

3.3.1 Resolved-Rate Motion Controller (RRMC)

For attitude control at the end of a robotic arm, we first introduce Resolved-Rate Motion Controller (RRMC) [14]. This is a closed-loop attitude controller that is applied directly to the robotic arm as a first-order differential equation to generate the motion trajectory of the end-effector with velocity-level commands.

$$\mathbf{T} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{1a}$$

$$\rho(\mathbf{T}) = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \tag{1b}$$

$$\tau(\mathbf{T}) = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} \tag{1c}$$

where $\rho(\mathbf{T}) \in \mathbb{R}^3$ constitutes the rotation component, $\tau(\mathbf{T}) \in \mathbb{R}^3$ constitutes the translation component.

The translation and rotation of the end effector with respect to the difference between the current attitude and the desired attitude is represented by the error vector e:

$$e = \frac{\tau\left({}^0T_{e^U}\right) - \tau\left({}^0T_{e^R}\right)}{\alpha\left(\rho\left({}^0T_{e^U}\right)\rho\left({}^0T_{e^R}\right)^T\right)} \in \mathbb{R}^6 \tag{2}$$



Figure 2. Hand gesture recognition points for the teleoperation

where ${}^0T_{e^R}$ is the end effect pose in the base frame, ${}^0T_{e^U}$ is the desired end effect pose in the base frame. $\alpha(\cdot)$: **SO**(Special Orthogonal Group) $\in \mathbb{R}^3$ turns rotation matrix into an equivalent Euler vector.The Euler vector specifies the rotation axis and rotation angle.

The v in the RRMC equation is determined by the error e, when $v = ke$ and k is the proportional gain term. We limit v to some value v_m by modifying the magnitude of the proportional gain term k until the error e drops below some value e_m:

$$v = \begin{cases} k_e \frac{v_m}{||k_e||} & \text{if } ||k_e|| > v_m \\ k_e & \text{if } v_m \geq ||k_e|| > ||e||_m \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

The RRMC ensures that the error decreases at the same rate until the speed gradually decreases to a smooth stopping of the arm.

3.3.2 Optimal Resolved Acceleration (ORA) with Resolved Acceleration Control (RAC)

RRMC relies more on the computation of Jacobi matrices and pseudo-inverses to deal with external forces and constraints indirectly through pliability matrices and avoids dealing directly with constraints arising in the system, leading to controller degradation. To solve this problem, we introduce quadratic programming techniques to achieve the desired results. Quadratic programming (QP) can transform the RRMC into an optimization problem and quickly generate the optimal control strategy in a short time. As a general optimization algorithm, it can provide additional functions directly in the formulation of RRMC and show good results when dealing with equality, inequality and cost constraints. Therefore, QP applied to the control of a robotic arm is a quadratic planning process that includes the physical constraints of the whole system, and does not encounter the problem of generating very high control speeds due to the presence of non-square and rank-loss Jacobi matrices.

In addition, RRMC, as a first-order differential equation, generates motion trajectories only at the velocity-level command. Precise positioning and trajectory tracking of the robotic arm is achieved by accurately controlling the acceleration in response to rapidly changing task demands and environmental conditions. For the required accuracy in the context of assembled buildings, we can introduce Optimal Resolved Acceleration (ORA) for optimization. ORA is primarily concerned with optimizing the acceleration of the robot, and is usually formulated through a quadratic programming problem with formulas that contain the cost terms for acceleration and velocity, as well as velocity and acceleration constraints in the task space. At the acceleration-level command, optimal joint accelerations and velocities are found by minimizing the cost functions of acceleration and velocity.

We use the resolved acceleration control (RAC) method, which maps the acceleration command $u_a(t)$ in the task space to the joint space [15]:

$$u_j(t) = [q(t),\ \dot{q(t)}, \ddot{q(t)}] \tag{4}$$

By direct application of second-order differential equations at the acceleration-level command, an optimal control problem under constraints is solved in each iteration to ensure the feasibility of the joint configurations. For an intial joint space configuration $u_j(t) = [q(t),\ \dot{q(t)}, \ddot{q(t)}]$, in each iteration of the QP solution, acceleration command $\ddot{x_d}$ and kinematics contraints of robotic arm can be showed:

$$\min_{q,\dot{q},\ddot{q},\delta} [q, \dot{q}, \ddot{q}, \delta] P [q, \dot{q}, \ddot{q}, \delta]^T \tag{5a}$$

Contraints:

$$q = q_0 + \dot{q}_0 \Delta t \tag{1}$$

$$\dot{q} = \dot{q}_0 + \ddot{q} \Delta t \tag{2}$$

$$\ddot{x_d} + \delta = J_0 \ddot{q} + \dot{q_0} \bigotimes H_0 \dot{q_0} \tag{5b}$$

where $P \in \mathbb{R}^{27\times27}$, Δt is time-step, $J_0 = J(q_0)$, $H_0 = H(q_0)$.

The introduction of slack variables $\delta \in \mathbb{R}^6$ to the task-space controller in the minimization of the acceleration command $\ddot{x_d}$, when the variables include $q, \dot{q}, \ddot{q}, \delta$, ensures that the optimal control problem is reachable to an optimal solution. The slack variable δ is in the cost function and is activated only if the acceleration command u_a does not satisfy the kinematic constraints. For ease of understanding and computation, we rewrite the objective function:

$$\min V(\ddot{q}) = \ddot{q}^T P \ddot{q} + \dot{q}^T Q \dot{q} \tag{6a}$$

Constraints:

$$\ddot{x_d} = J(q)\ddot{q} + \dot{J(q)}\dot{q} \quad (6b)$$

$$\dot{x_d} = J(q)\dot{q} \quad (6c)$$

where $\dot{x_d}, \ddot{x_d}$ are desired mission space velocity and acceleration.

Both P and Q are positive definite matrices used to weight the cost of joint acceleration and velocity.

For the first part of the optimization objective of the objective function $\ddot{q}^T P\ddot{q}$, minimizing joint acceleration $\ddot{q}$ and controlling acceleration amplitude.

For the second part of the optimization objective of the objective function $\dot{q}^T Q\dot{q}$, minimizing joint velocity $\dot{q}$ and controlling velocity smooth.

3.3.3 Compliance Control and Environmental Adaptability

Since we need to meet accuracy not only in the context of assembly-type building installation, but also in applications where humans collaborate with robotic arms, or operate in uncertain environments, we introduced suppleness control to robotic arm operation to avoid collisions and injuries.

In order to enhance the compliance, a compliance matrix S is introduced into the objective function $V(\ddot{q})$. This compliance matrix S, weighs the relationship between joint acceleration $\ddot{q}$ and end-actuator acceleration $\dot{x_d}$ while enhancing the collaborative capability of the robotic arm. The objective function $V(\ddot{q})$ can be modified as:

$$\min V(\ddot{q}) = \ddot{q}^T P\ddot{q} + \dot{q}^T Q\dot{q} + (\ddot{x} - \ddot{x_d})^T S(\ddot{x} - \ddot{x_d}) \quad (7a)$$

Constraints:

$$\ddot{x_d} = J(q)\ddot{q} + \dot{J(q)}\dot{q} \quad (7b)$$

$$\dot{x_d} = J(q)\dot{q} \quad (7c)$$

where $\dot{x_d}, \ddot{x_d}$ are desired mission space velocity and acceleration. For the compliance matrix S, which reflects the weight of the system on the end-effector acceleration error. S as a symmetric positive definite matrix for tuning the sensitivity of the system to different directional pliability:

$$S = \mathrm{diag}(\lambda_x, \lambda_y, \lambda_z) \quad (8)$$

where $\lambda_x, \lambda_y, \lambda_z$ denote the weights on the acceleration errors in Cartesian space for the three directions, respectively. When $\lambda_x, \lambda_y, \lambda_z$ is small, the system is less sensitive to errors and exhibits a higher degree of suppleness; conversely, it exhibits rigidity.

In human-robot collaboration, the compliance matrix S can be dynamically adjusted to enhance the compliance and adaptability of the robotic arm according to environmental demands. The compliance matrix S can also be combined with an impedance control strategy to further improve the responsiveness of the robotic arm to external forces.

3.4 Apparatus

3.4.1 Manipulator

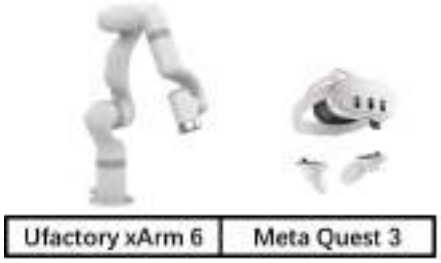


Figure 3. Hardware required for the teleoperation

In order to verify the feasibility of our approach to this control, we chose a robotic arm of the xArm 6 model from Ufactory as shown in Figure 3 . It has high load and accuracy, with a maximum load of 5kg and repeatability of ±0.1mm. xArm-Python-SDK provides a complete Python interface, which makes it easy for researchers to quickly get started and perform new algorithms. Compared with other industrial robots, xArm 6 is less expensive and more cost-effective. Therefore, this approach was used to validate the feasibility of the control method proposed in the paper.

3.4.2 Teleoperator

Meta Quest 3 (Figure 3) supports 6 Degree of Freedom (DoF) tracking function with excellent motion tracking performance. It is capable of accurately capturing the user's movements with an average latency of only 12 milliseconds, which enables natural and smooth interactions, allowing us to manipulate the virtual objects and carry out experimental simulations in a more precise way, improving the accuracy and efficiency of scientific research.

4 Experiment

The experimental section provides a detailed introduction to two key experiments: aluminum profile assembly based on MR teleoperation framework and robot arm trajectory accuracy testing. The aluminum profile assembly experiment demonstrated the effectiveness of the remote control framework in controlling the construction of assembly tasks for the robotic arm. The trajectory accuracy test of the robot arm evaluates its accuracy by comparing its trajectory with the trajectory of the hand. The results indicate that although there are some errors in the accuracy of the robotic arm trajectory, the overall control accuracy

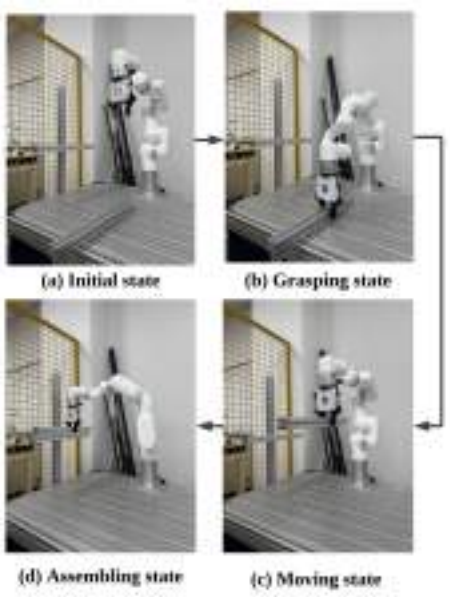


Figure 4. The teleoperation of assembling aluminum profiles. Subfigure (a) represents the initial position of the robotic arm, at which point there is no movement of the robotic arm. (b) illustrates that robotic arm received instructions to clamp the aluminum profile. Subfigure (c) shows arm moves according to the trajectory of the hand. And (d) presents the state that robotic arm completes the assembly process.

is considerably high. Both experiments emphasize the potential of remote operating systems in sustainable building practices.

4.1 Aluminum Profile Assembly

To explore whether a robotic arm can replace human labor to perform various assembly tasks, it is necessary to test its flexibility. When the robotic arm is operated using the MR telecontrol frame, it is tested whether the robotic arm can assemble the components as flexibly as the traditional assembly method. For this purpose, a small experiment was designed to use the MR device to manipulate the robotic arm to pick up aluminum profiles on the workbench and place them on a shelf built from aluminum profiles. The experimental process is shown in the Figure 4 .

The key to this experiment is to observe whether the robot arm can complete the task quickly and flexibly. If the robotic arm is able to complete the task successfully, then the framework is feasible and the assembly task will not be affected by remote control or MR interaction. This manipulation method can be further tested to manipulate larger industrial robotic arms to accomplish some assembly tasks.



(a) (b) (c)

Figure 5. Movement of the manipulator and hand during the assembly process: (a) The robotic arm is controlled via hand interaction using MR-based gesture recognition to clamp aluminum profiles. (b) The robotic arm follows the trajectory of the hand interaction to move to the designated location. (c) Completion of the assembly task by the robotic arm.

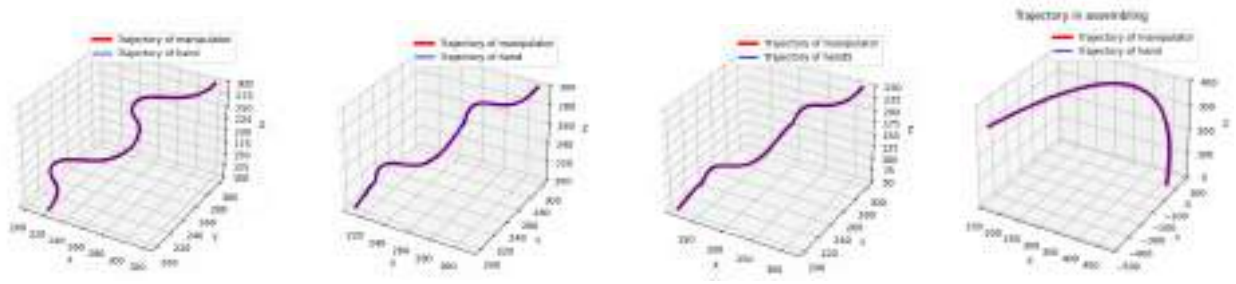
4.2 Accuracy test

Another measure of teleoperation of robotic arms to replace humans in completing assembly work is accuracy, as robots need to fully follow the operator's hand movements to complete task responses. The higher the precision, the higher the success rate of the robot arm in completing assembly work. In this experiment, MR equipment will be used to detect three different human-demonstrated paths (experiment $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ shown in Figure 5). In experiments, the operator's hand will operate the robotic arm within a reasonable range to draw different trajectories. Then observe whether the trajectory generated by the robot arm is consistent with the trajectory of the hand. The degree of overlap between two trajectories is the precision of control.

5 Result and Analysis

5.1 Result of Aluminum Profile Assembly

As shown in the Figure 5 , the operator finally controlled the robotic arm to successfully complete the assembly work of the aluminum profile. This indicates that the MR teleoperation framework can flexibly control robot arms to complete simple assembly tasks. In addition, setting aside various physical controllers, the control method of directly detecting the position and state of the hand is also more convenient for operators to understand its control mode. We extracted 200 points from the hand trajectory and robotic arm trajectory of each group of experiments, and then calculated the offsets of adjacent two points on the X, Y, and Z axes in turn. Then, the distance threshold is set, and the offset error between the trajectory of the manipulator and the trajectory of the hand is smaller than the threshold, which Indicates that the two trajectories are aligned In this study, the distance threshold was set



(a) (b) (c) (d)

Figure 6. Trajectory comparison of human-demonstrated paths and robotic manipulator replication under various conditions. Subfigures (a), (b), and (c) illustrate three distinct trajectories demonstrated by hand movements using MR-based remote control of a robotic manipulator. The trajectories executed by the manipulator (red lines) are compared against the original human-demonstrated trajectories (blue lines) to evaluate precision and alignment accuracy. (d) presents the trajectory comparison during an assembly task, where the manipulator replicates the trajectory of grasping aluminum profiles.

at 0.5mm. The experimental results also show that the remote control scheme has a delay of less than 400 ms, and the main reason for the delay is the network delay, and the coincidence between the hand trajectory and the robot arm trajectory reaches 96%. However, the assembly work in the experiment was too simple and unitary, so more experiments are needed to explore whether the robotic arm controlled by the scheme can help workers complete the assembly work in high-risk or extreme environments.

5.2 Result of Accuracy test

As can be seen in the Figure 6 , the trajectory of the robotic arm essentially coincides with the trajectory of the hand. The table 1 also shows the overlap of the two trajectories in each set of experiments.The higher overlap among the three groups indicates that the robot arm has a certain degree of accuracy. In Experiment ***a***, the overlap between the two trajectories reached 97%, and in Experiments ***b*** and ***c***, the overlap between the two trajectories also reached 92% and 94%, respectively. This means that the manipulator can use this control method for more precise control. From the figure, it can be seen that there is some noise in the hand trajectory, which may be introduced by the MR device when detecting the position and posture of the hand. However, after optimizing the motion control algorithm of the robot arm, the trajectory of the arm became smoother. This indicates that using this control framework, the robot arm also has a certain degree of noise resistance, which can avoid shaking caused by noise.

Experiments	Accuracy
a	97%
b	92%
c	94%
Trajectory in assembling	96%

Table 1. Accuracy of manipulator

6 Conclusion and Future work

In this study, we propose an industrial robot arm teleoperation method based on mixed reality technology for assembly work in prefabricated buildings. Through the integration of TCP connection, data transmission and advanced motion control algorithm, xArm 6 robot arm and Meta Quest 3 are used to carry out the feasibility analysis of this control method, and the aluminum profile assembly experiment and accuracy test are carried out.

The experimental results show that the MR teleoperation frame can effectively and flexibly control the robot arm to complete simple assembly tasks. This direct detection of the position and status of the hand control mode, for the operator to understand and operate the system provides convenience. However, current assembly tasks do not fully demonstrate the complexity of assembly tasks, so further experiments are needed to explore the full potential of robotic arms in supporting and assisting humans with difficult assembly tasks.

In the accuracy test, the robotic arm has a certain accuracy, and its trajectory is very consistent with the trajectory of the hand. Although there is some noise in the hand trajectory, which may be introduced by the MR Device during the hand position and attitude detection process, the motion control algorithm optimization of the robot arm effectively smooths its trajectory. This shows that the control frame makes the manipulator have the ability to resist noise and avoid jitter.

Future work will focus on several key areas. Firstly, our goal is to increase the complexity and diversity of assembly tasks to more comprehensively evaluate whether robotic arms have the potential to help humans in assembly

tasks. This will involve testing a wider range of building components and assembly scenarios. Secondly, in order to improve the accuracy and efficiency of assembly work, efforts will be made to enhance the accuracy and stability of the system. This will require further optimization of motion control algorithms and calibration of MR devices to minimize errors and noise. Finally, we plan to explore the integration of more intelligent features, such as real-time obstacle detection and avoidance, to improve the safety and reliability of robotic arms in building environments [16].

References

[1] J. Jayawardana, M. Sandanayake, J. Jayasinghe, A. Kulatunga, and G. Zhang. A comparative life cycle assessment of prefabricated and traditional construction–a case of a developing country. *Journal of Building Engineering*, 72:106550, 2023.

[2] Y. Xiang, A. Opoku, and L. Florez-Perez. Carbon emissions optimisation in prefabrication construction: A review of current design integrated approaches. Online: `https://www.ucl.ac.uk/bartlett/construction/sites/bartlett_construction/files/carbon_emissions_optimisation_in_prefabrication_construction_a_review_of_current_design_integrated_approaches.pdf`, 2024. Accessed: 01/12/2024.

[3] Y. Tian and S. Spatari. Environmental life cycle evaluation of prefabricated residential construction in china. *Journal of Building Engineering*, 57:104776, 2022.

[4] R. Muchacho, R. Laha, L. Figueredo, and S. Haddadin. A solution to slosh-free robot trajectory optimization. In *International Conference on Intelligent Robots and Systems (IROS)*, pages 223–230, Kyoto, Japan, 2022.

[5] L. Sidenmark, T. Zhang, L. Al Lababidi, J. Li, and T. Grossman. Desk2desk: Optimization-based mixed reality workspace integration for remote side-by-side collaboration. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*, pages 1–15, 2024.

[6] Graintech reduces travel spending and carbon emissions by 68 per cent using microsoft mixed reality technology. Microsoft Australia News Centre. Online: `https://news.microsoft.com/en-au/features/graintech-reduces-travel-spending-and-carbon-emissions-by-68-per-cent-using-microsoft-mixed-reality-technology/`, 2024. Accessed: 24/12/2024.

[7] X. Li, W. Yi, H. Chi, X. Wang, and A. Chan. A critical review of virtual and augmented reality (vr/ar) applications in construction safety. *Automation in construction*, 86:150–162, 2018.

[8] T. Takahashi and H. Ogata. Robotic assembly operation based on task-level teaching in virtual reality. In *Proceedings 1992 IEEE International Conference on Robotics and Automation*, pages 1083–1084. IEEE Computer Society, 1992.

[9] A. Jarecki and K. Lee. Mixed reality-based teleoperation of mobile robotic arm: System apparatus and experimental case study. In *2024 21st International Conference on Ubiquitous Robots (UR)*, pages 198–203. IEEE, 2024.

[10] R. Zhang, X. Liu, J. Shuai, and L. Zheng. Collaborative robot and mixed reality assisted microgravity assembly for large space mechanism. *Procedia Manufacturing*, 51:38–45, 2020.

[11] V. Cerf and R. Kahn. A protocol for packet network intercommunication. *IEEE Transactions on Communications*, 22(5):637–648, 1974. doi:10.1109/TCOM.1974.1092259.

[12] J. Jerald, P. Giokaris, D. Woodall, A. Hartholt, A. Chandak, and S. Kuntz. Developing virtual reality applications with unity. In *2014 IEEE Virtual Reality (VR)*, pages 1–3. IEEE, 2014.

[13] X. Dengxiong, X. Wang, S. Bai, and Y. Zhang. Self-supervised 6-dof robot grasping by demonstration via augmented reality teleoperation system. *arXiv preprint arXiv:2404.03067*, 2024.

[14] Z. Wu, Y. Chu, X. Zhu, Q. Sun, and Y. Yue. Immersive embodied telemanipulation system with a velocity controller. In *2024 29th International Conference on Automation and Computing (ICAC)*, pages 1–6. IEEE, 2024.

[15] J. Arrizabalaga, L. Pries, R. Laha, R. Li, S. Haddadin, and M. Ryll. Geometric slosh-free tracking for robotic manipulators. *arXiv preprint arXiv:2402.05197*, 2024.

[16] H. Wang, Z. Yang, Q. Zhang, Q. Sun, and E. Lim. A digital twin platform integrating process parameter simulation solution for intelligent manufacturing. *Electronics*, 13(4):802, 2024. doi:10.3390/electronics13040802.

Interface Management of Robots in Construction: A Case Study of Overhead Drilling Robot

Zhida Zhang[1], Mi Pan[1], Sujuan Zhang[2,3], Quan Sun[1]

[1]Department of Civil and Environmental Engineering, University of Macau, Macao, SAR, China
[2]School of Project Management, Faculty of Engineering, University of Sydney, Sydney, Australia
[3]John Grill Institute for Project Leadership, University of Sydney, Sydney, Australia
zhidazzd@um.edu.mo, mipan@um.edu.mo, sujuan.zhang@sydney.edu.au, quansun@um.edu.mo

Abstract –

The introduction of innovative robotics to the traditional construction processes involves complex and multi-dimensional interfaces that could hinder effective technology implementation. This paper aims to identify potential interface problems between construction robots and traditional construction processes, and propose effective strategies for interface management. Drawing on a case study of the implementation of an overhead drilling robot in Macao, the findings indicate various issues related to physical, contractual and organizational interfaces and their causes associated with the integration of construction robots. In particular, contractual interface problems were identified as the most critical factors affecting the real-world application of robots in construction. This highlights the importance of developing new business models and fostering collaboration among stakeholders through innovative contract forms. The developed innovative interface management framework for construction robots provides valuable insights for future research and practice to understand and manage complex interfaces between robots and traditional construction processes. The paper should contribute to the better implementation of robots in construction.

Keywords –

Construction robot; Interface management; Drilling robot; Engineering management

1 Introduction

Construction robots, often refer to sophisticated and intelligent machinery designed for construction processes with robotic features, are emerging as a cutting-edge solution that could address many challenges confronting the industry [1]. The first on-site construction robot was developed by the Japanese Shimizu company for spraying surface materials in the 1980s [2], followed by diverse construction robots springing up during the 1990s. Consequently, many prototypes of single-task robots have been developed in the construction industry, such as those for concrete finishing, wall painting, and ceiling board installation, providing evidence of the potential of robots to assist in performing dangerous, dirty, and difficult construction jobs [3]. Recently, research has also embarked on more general-purpose or human-assisted robotics in construction, such as aerial robots, robotic exoskeletons, collaborative robots and humanoid robots [4, 5].

Previous works have contributed to the development of construction robots and demonstrated technological and economic feasibility in specific cases [6]. However, robots have yet to experience large-scale, real-world implementations in construction. There are still doubts regarding how robots will outperform humans in construction and what new problems they might introduce [4, 7]. One major barrier hindering the effective implementation of robots in construction is the complex interfaces between robots and traditional construction processes. However, there is still limited exploration of these interfaces and how they can be effectively managed.

This paper fills the above knowledge gap by achieving three objectives: 1) identifying the potential challenges related to the complex interfaces between robots and traditional construction processes; 2) examining causes of interface problems; and 3) proposing strategies for effective interface management. The research is based on a case study of a pilot trial of an overhead drilling robot in Macao, drawing on the classification of the interfaces into physical, contractual and organizational (PCO) dimensions [9].

2 Background

2.1 Interface Management in Construction

The term ‘interface’ generally refers to the soft or/and hard interaction of entities (e.g., products, people) with boundary limits [8]. In construction, Pavitt and Gibb (2003) [9] argued that interfaces between different elements and components cause most problems and highlighted the importance of interface management.

Interface management is critical to project success, as it could directly or indirectly affect the time, cost and quality performance of a construction project [10]. Wideman [11] defined it as '*managing the problems that often occur among people, departments, and disciplines rather than within the project team itself.*' Construction Industry Institute (CII) [12] defined it as '*management of communications, relationships, and deliverables among two or more interface stakeholders*'. The effective implementation of interface management could enhance construction processes, reduce the occurrence of rework, and shorten the overall construction period [10].

Earlier research has identified interface problems that emerge in traditional construction projects due to the complex and fragmented nature of the construction industry [13]. In particular, poor communication and coordination among project stakeholders have been widely recognized as key issues leading to interface problems like design errors and construction conflicts [14]. Recent studies examined the problematic interfaces in modern construction methods, like prefabricated and modular construction [15, 16].

However, limited empirical research has investigated interface problems from the adoption of construction robots. The introduction of robots as a disruptive technology to the traditional construction processes often involves complex interfaces and triggers new interface problems. To unleash the full potential of robots in construction, it is essential to understand the potential interface problems and develop effective interface management strategies.

2.2 Physical, Contractual and Organizational Interfaces of Robots in Construction

There are different ways to understand interfaces and interface management. From a systems perspective, interface management can be categorized into four types: time, relational, information, and environmental [17]. Additionally, Pavitt and Gibb [9] classified three types of interface management within construction as physical, contractual and organizational. Specifically, physical interfaces refer to the actual physical linkage between two or more building elements or components; contractual interfaces relate to how the work packages involving processes and people have been formed from a contractual basis; organizational interfaces indicate managerial interactions related to individual and/or group relationships. Building upon Pavitt and Gibb's categorization [9], this research examines the complex interfaces from the three dimensions, with a specific focus on the integration of construction robots.

Physically, the implementation of robots on construction sites necessitates careful consideration of spatial constraints, adaptability to diverse and complex work environments, and the potential impact on construction personnel and existing infrastructure.

Contractually, the introduction of robots could affect traditional contracting and subcontracting practices and result in new issues related to obligations, liabilities, and profitability.

Organizationally, the integration of robots will change the existing workflows, roles, and responsibilities of various parties on construction sites, which require organizational adjustments, such as redefining job roles and fostering an innovative culture to embrace change.

3 Methodology

The research design is illustrated in Figure 1. Initially, a critical literature review was conducted using databases including Web of Science, Scopus, and Google Scholar, with keywords related to "construction robotics" and "interface management". The review aims to examine the application challenges of construction robots and explore the potential interface problems, which enables the preliminary analysis of physical, contractual and organizational (PCO) interfaces of robots.

Then, a case study of the implementation of an overhead drilling robot in Macao was carried out to further analyze the PCO interface problems between the drilling robot and traditional construction processes. The case study utilized multiple sources of evidence, including documents, researchers' observations of the actual application of the robot through site visits, and 7 semi-structured interviews with key stakeholder groups covering principal contractor and mechanical, electrical and pumping (MEP) subcontractors. The interviewees were purposefully selected considering their direct involvement in the case of the drilling robot and work experience in the industry, including 1 project director, 2 project managers, 1 site manager and 3 technical staff. The interview questions were semi-structured, combining both open and closed-ended questions to gather comprehensive insights into the interface problems associated with robot integration. In particular, the interviewees were asked to provide their perspectives on what specific interface problems they encountered, why these problems occurred, and their suggestions for addressing these challenges. Each interview lasted for approximately 20-30 minutes. The data collected from the case study were analyzed using the NVivo 15 software.

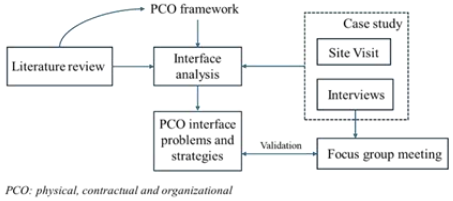


Figure 1. Research design

Finally, the initial findings identified from the case

study were further validated through a focus group meeting with the previous interviewees.

4 Case Study Results and Analysis

4.1 Case Background

The case study was conducted with a pilot trial of an overhead drilling robot applied on the construction site of a residential building project comprising 1,566 residential units. The project is in the form of an Engineering Procurement Construction (EPC) contract in which the MEP work is carried out by various subcontractors. The pilot trial was initiated by the principal contractor, while the robot was responsible for the drilling tasks involving MEP engineering.

The overhead drilling robot utilized in the case study project is a semi-automatic construction robot (Figure 2), with a flexible robotic arm for drilling tasks mounted on a mobile vehicle. It is specifically designed for MEP and interior installation work, which is often strenuous, challenging, and time-consuming for construction workers. In the studied case, the operator was an electrician with one hour of training.

During the construction process, the operator needs to move the robot to the designated work area via a controller and utilize a total station for positioning purposes. Then, within the achievable area, the robotic arm can automatically drill holes for one or multiple MEP subcontractors based on the design data from AutoCAD or Revit.



Figure 2. The overhead drilling robot

Compared with traditional manual-based methods, the drilling robot can offer several distinct advantages. For instance, it can eliminate potential risks associated with overhead installation and conduct vacuuming operations during drilling for improved occupational health and safety. Moreover, the robot can enhance productivity and efficiency by autonomously marking and drilling holes according to digital plans.

4.2 Interface Analysis

Drawing on the case project, the identified PCO interface problems, potential causes, and key stakeholders are shown in Tables 1 to 3, and elaborated in the following subsections.

4.2.1 Physical Interfaces

The implementation of the drilling robot may lead to physical interface problems between robot and operator/worker, site environment, and built structure.

Between robot and operator/worker: During the drilling process, human errors by robot operator (e.g., accidentally touching and displacing the total station) may result in construction deviations. Additionally, cross-work and collaboration between the robot and construction workers from multiple trades are inevitable, potentially leading to physical contact and associated problems.

Between robot and site environment: The dynamic and complex construction site presents significant challenges for the robot performance, particularly regarding their interfaces with site environment. For example, as one interviewee noted, "*due to the high reliance on the total station, when facing a space/room with many obstructions, operators must perform multiple positioning operations for the robot, which greatly reduces its efficiency.*"

Between robot and built structure: The robot's primary function is to drill holes in the built structure to meet the installation requirements of MEP engineering. Consequently, it will inevitably encounter the built structure. The robot may face physical interface problems with the rebar or pre-built piping inside the structure and with the finished paint on the surface.

To summarize, potential causes for physical interface problems mainly include the lack of technological maturity of the robot (such as limited mobility, flexibility, and adaptability) and poor engineering management.

4.2.2 Contractual Interfaces

When work elements are grouped into different work packages to meet the needs of the principal contractor, new contractual interfaces may arise between work packages, typically involving specialist contractors [10]. The introduction of the drilling robot may create contractual interface problems in terms of work package reassignment and liability attribution among stakeholders.

Work package reassignment: The drilling robot is designed for a single task with the aim of integrating all MEP drilling tasks, while in the case project, these tasks were undertaken by separate subcontractors. This necessitates the adjustment of relevant tasks at the contract level, which led to contractual interface problems among the principal contractors and MEP subcontractors. Furthermore, due to the limited coverage, areas that the robot cannot reach need to be drilled by workers, which also requires contractual reassignment of work packages.

Table 1. Identified physical interface problems and potential causes

Category	Physical interface problems	Potential causes	Key stakeholders						
			PC	MEP	P	A	S	O	W
Between robot and operator/ worker	• The robot operator or worker may accidentally touch and displace the total station, which could result in localization errors for the robot, potentially leading to subsequent drilling errors	• The localization of the drilling robot highly relies on the total station; no mechanism to avoid or alarm its displacement	√	√				√	√
	• The drilling robot might introduce new risks (e.g., injury) to co-workers	• Workers' unfamiliarity and incompetence in working with the robot	√				√		√
Between robot and site environment	• The drilling robot requires relocalization for different rooms/spaces with obstacles (e.g., columns, walls), resulting in a long setup time	• There should be no obstructions between the drilling robot and the total station	√	√			√	√	
	• Special arrangements are required for the drilling robot to move in a dynamic site environment with steps	• The drilling robot has limited mobility	√	√			√		
	• The drilling robot is incapable of adjusting drilling to construction errors (e.g., discrepancies between actual construction and drawings)	• The robot can't recognize and address construction errors and uncertainties; poor construction management	√			√	√		
	• Uneven ground can affect the drilling quality and accuracy of the robot	• The drilling robot has limited adaptability	√	√			√		
Between robot and built structure	• The robot is unable to make flexible adjustments of drilling when encountering rebar in the structure to meet subsequent MEP installation requirements	• The drilling robot has limited flexibility	√	√			√		
	• The robot may damage the surface of the structure or pre-built piping during drilling	• The drilling robot has a low intelligence level	√	√			√		
	• The drilling robot may cause damage or pollution to the completed ceiling and floor paint	• Lack of measures to protect finished painting for the previous process	√		√			√	

Note: PC = principal contractor; MEP = MEP subcontractor; P = painting subcontractor; A = architect; S = robot supplier; O = robot operator; and W = Worker.

Table 2. Identified contractual interface problems and potential causes

Category	Contractual interface problems	Potential causes	Key stakeholders						
			PC	MEP	P	A	S	O	W
Work package reassignment	• The drilling robot will integrate all drilling requirements of MEP works that are assigned to different subcontractors under separate contracts, with unclear responsibility	• There is no large subcontractor that can provide an integrated MEP engineering solution in Macao	√	√			√		
	• The drilling robot may have limited coverage and require supplementary work by MEP workers	• The robot is unable to access narrow spaces and corners for drilling	√	√			√		√
Liability attribution	• The drilling robot may cause delays (e.g., robot breakdown, maintenance) in project schedule and result in claims	• Improper planning and scheduling with the use of robots	√	√		√			
	• The liabilities are unclear among stakeholders for the drilling errors of the robot due to improper construction works (e.g., uneven ground, construction deviations, problems with drawings), operator mistakes or robot performance that may incur rework and additional cost	• Immature business models	√	√	√	√	√	√	
	• The introduction of the drilling robot necessitates the implementation of additional measures and agreements (e.g., reserve dedicated channels, coordinate working areas) to guarantee robot operating performance	• The drilling robot is integrated into traditional construction as a new element and role	√	√	√		√		

Note: PC = principal contractor; MEP = MEP subcontractor; P = painting subcontractor; A = architect; S = robot supplier; O = robot operator; and W = Worker.

Liability attribution: As a new role introduced in construction sites, the drilling robot requires additional measures and attention to ensure its functional operation. Moreover, factors such as limited robot performance, poor project management, immature business models and improper construction works could result in the misuse of robot that cause project delays, rework, and even additional costs and contract breaches. These circumstances often involve multiple stakeholders in the attribution of contractual liability.

Based on the case study, many respondents considered contractual interface issues as the most critical factors affecting the practical application of robots. Specifically, these issues involve the complex reassignment of tasks and liability attribution resulting from the integration of robots. This could lead to contractual misunderstandings and disputes, and affect the profitability of stakeholders and the overall efficiency of the construction process. Consequently, these challenges could hinder the widespread adoption of robots in the construction sector.

4.2.3 Organizational Interfaces

The implementation of the drilling robot on construction sites involves multiple stakeholders, with various organizational interfaces forming during the communication and collaboration process at project, government and industry levels.

At the project level, with the introduction of the robot and the integration of drilling requirements, the original workflows and responsibilities within the project have been altered, potentially leading to organizational interface issues, such as difficulties in coordination and communication between organizations. Additionally, poor management may result in worker idleness with the introduction of robots on site.

At the government level, the early stages of robot development are characterized by funding shortages and a lack of relevant regulations and standards. Moreover, respondents have argued that the absence of government support in economic and policy aspects significantly hinders the widespread adoption of robots in construction.

At the industry level, organizational interface problems usually involve all stakeholders. These challenges are primarily reflected in industry resistance to adopting robots, which may be driven by a lack of innovative culture. Potential risks posed by robots, such as employment concerns, present additional challenges in the industry.

Table 3. Identified organizational interface problems and potential causes

Category	Organizational interface problems	Potential causes	Related key stakeholders						
			PC	MEP	P	A	S	O	W	G
Project level	• Robotic integration of MEP drilling involves multiple stakeholders, which has more coordination difficulties and is vulnerable to communication problems	• Lack of management and coordination between multiple stakeholders	√	√	√	√	√			
	• There could be worker idling issues during the operation of the drilling robot	• Inexperience of the management team in work arrangements involving robots	√	√					√	
Government level	• There are new legislation and regulation related issues (e.g., additional time for permit approval) with the introduction of the drilling robot	• The development of robots is still in its early stages and lacks standards and practice	√				√			√
	• The drilling robot involves high initial costs, while its development and adoption lack government support and incentives	• The development of robots is still in its early stages and lacks economic justification	√				√			√
Industry level	• The drilling robot may pose a threat to employment opportunities and salaries of workers	• The introduction of the drilling robot will increase efficiency and reduce the need for manual labor	√	√					√	√
	• Multiple stakeholders may have resistance to the adoption of the drilling robot	• Lack of innovative culture and acceptance of new technology	√	√	√	√	√	√	√	√

Note: PC = principal contractor; MEP = MEP subcontractor; P = painting subcontractor; A = architect; S = robot supplier; O = robot operator; W = Worker; and G = government.

4.3 Strategies for Interface Management

Based on the identified interface issues and their causes associated with the integration of construction robots, several strategies are proposed on the technical, contractual and organizational aspects for different stakeholders.

Technical strategies are mainly focused on the design and development of robots to avoid or address technical interface problems, which are particularly important for robot suppliers, including:

- Improve the adaptability and mobility of robots to better interact with unstructured, dynamic and complex construction site environments.
- Enhance the flexibility of robots to handle uncertainties and a variety of construction tasks.
- Design more user-friendly and collaboration-enabled interfaces to allow robots to effectively work with people in a conflict-free manner.

Contractual strategies primarily concern the need to adjust existing contractual arrangements among principal contractors and subcontractors, including:

- Adjust and optimize contracts to reasonably allocate work packages and improve liability attribution. This includes clarifying the new roles and responsibilities of stakeholders considering the introduction of robots and ensuring that contractual terms align with their work requirements.
- Develop new business models to address the potential conflicts and claims arising from robot-related performance issues, foster collaboration among stakeholders with pain/gain mechanisms and better integration.

Organizational strategies primarily consider the requirement of the government and principal contractors' efforts in project management, policy implementation and innovation culture creation to address organizational interface problems, including:

- Principal contractors should embrace potential robot-incurred issues into existing project management practice and enhance the coordination among various stakeholders.
- The government should accelerate the formulation of relevant legislation, regulations and standards to clarify the roles and responsibilities with the introduction of robots in construction. Government support and incentives should be provided during the initial phases of robot development to encourage research and trial projects.
- The industry should cultivate a culture of innovation to increase the acceptance of robots, and take measures to address the potential risks (e.g., worker unemployment) associated with robot integration.

5 Discussion

This paper conducted a case study on the interfaces between the overhead drilling robot and traditional construction processes from PCO perspectives. By analyzing and summarizing the interface problems, their corresponding causes, and preventive strategies, a PCO interface management framework for construction robots is proposed, as shown in Figure 3.

The research has identified potential interface problems when introducing robots in construction environments. Firstly, limitations in the robot's adaptability, intelligence level and collaborative ability impact its performance in the construction process. This results in numerous physical interface problems, primarily occurring between robots and operators/workers, construction site environments, and existing built structures. Secondly, the contractual interface problems, which are identified as the most critical factors affecting the practical application of robots, mainly involve the reassignment of work packages and the attribution of contractual liabilities. On the one hand, the introduction of the robot integrates traditional tasks in a different way, which incurs changes in work processes and requires the reassignment of subcontractors' work packages. On the other hand, potential issues (e.g., rework, project delays) caused by robots require a determination of contractual liability attribution. Finally, organizational interfaces mainly occur among different organizations or individuals. Factors such as poor communication and collaboration, as well as the lack of clear regulations and standards, may lead to interface problems at the project, government, and industry levels. These findings echo with previous studies to highlight the importance of developing new business strategies and stakeholder networks to promote the implementation of construction robots [18].

The findings further indicate the complex interactions among PCO interfaces. Firstly, physical interface problems can affect the contractual interface. For instance, the mobility limitations of the drilling robot in the studied case necessitated the establishment of a dedicated working channel, requiring supplementary contract clauses. Conversely, modifications to the contractual interface can lead to the regrouping of certain work elements into new work packages, which may involve different subcontractors and result in changes in physical interfaces. For example, integrating the drilling tasks using the robot could create additional physical interfaces between the robot and MEP subcontractors. Secondly, both physical and contractual interfaces can influence organizational interfaces. Physical interface problems (e.g. worker idling) can intuitively reflect issues in communication and collaboration between organizations. Meanwhile, contractual interface issues, such as conflicts and liability attribution, can negatively

affect inter-organizational collaboration and organizational profitability. Finally, effective coordination of organizational interfaces can address problems related to physical and contractual interfaces, such as coordinating work areas and adjusting contract terms.

This study adapts the PCO classification for interface management in the construction field [9, 20] and pioneers the research of interface management for construction robots. While previous studies mainly focus on technical challenges of construction robots in the implementation [19], this research emphasizes the underlying interface problems in the contractual alignment and division of work that could hinder the real-world application of robots in construction. The developed framework enriches the previous studies by specifying the interface elements and illustrating the complex interrelationship among multi-dimensional interfaces, which contributes to both knowledge fields of interface management and construction robots.

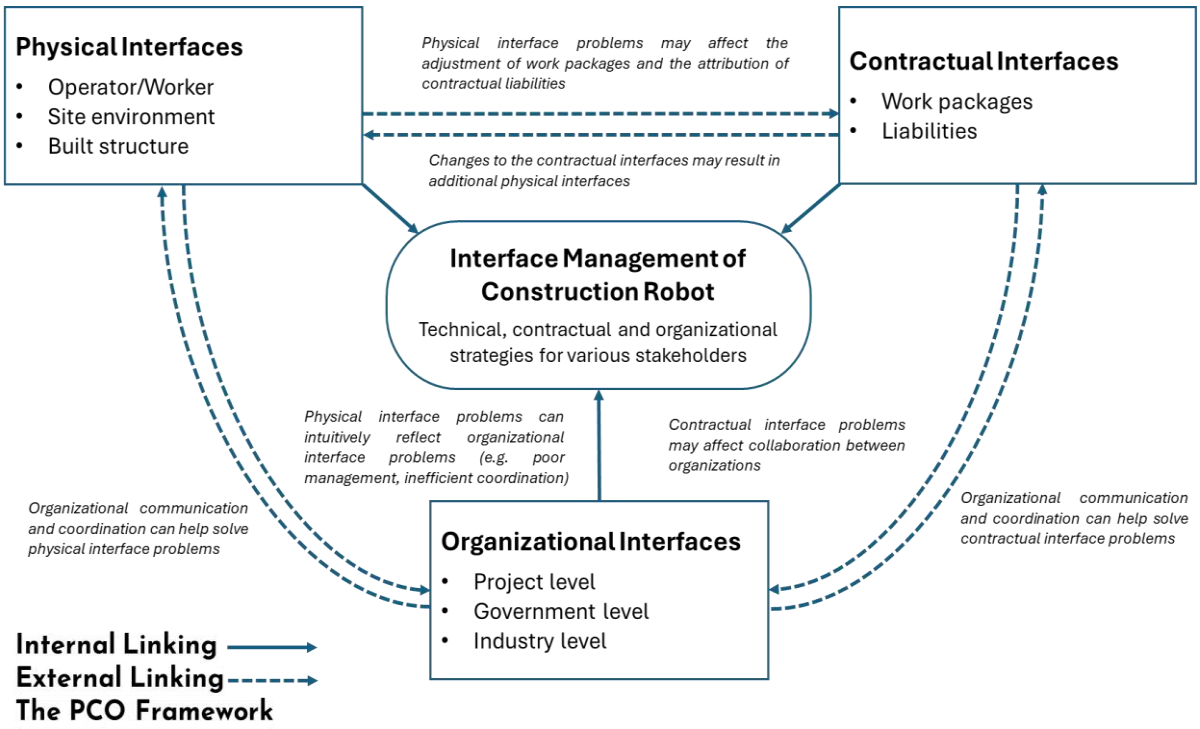


Figure 3. PCO interface management framework for construction robots

6 Conclusion

This paper investigates the interface management of construction robots through a case study of an overhead drilling robot by examining the interface problems, potential causes and preventive strategies. The findings reveal complex interface issues emerging between robots and traditional construction processes, leading to an innovative PCO interface management framework for construction robots. In particular, contractual interface issues are highlighted as the most critical factors hindering the implementation of robots, including work package reassignment and liability attribution. The findings further highlight the importance of developing new business models and stakeholder collaboration schemes with innovative contract forms to address the interface problems involved in construction robot implementation and support effective technology uptake.

This research contributes to the literature on interface management in the construction industry by proposing an innovative framework for interface management specific to construction robots. The PCO framework offers a new theoretical foundation for understanding the complex interfaces with the integration of robots in construction, and provides valuable insights for the industry and academia to address the interface problems derived from the implementation of construction robots. Future research can extend the research by incorporating a broader range of construction robots. Further research can also apply the framework in different institutional contexts to improve its generalizability.

Acknowledgments

We acknowledge funding support from the University of Macau (File no. SRG2023-00006-FST). We also would like to express our sincere gratitude to China Construction Engineering (Macau) Company Limited for their support in the case study.

References

[1] Pan, M., Linner, T., Pan, W., Cheng, H., and Bock, T. Structuring the context for construction robot development through integrated scenario approach. *Automation in construction*, 114, 103174, 2020.

[2] Cousineau, L., and Miura, N. *Construction robots: the search for new building technology*. ASCE Publications, Japan, 1998.

[3] Bock, T., and Linner, T. *Construction robots: Volume 3: Elementary technologies and single-task construction robots*. Cambridge University Press, 2016.

[4] Pan, M., and Pan, W. Stakeholder perceptions of the future application of construction robots for buildings in a dialectical system framework. *Journal of Management in Engineering*, *36*(6), 04020080, 2020

[5] Liang, C. J., Wang, X., Kamat, V. R., and Menassa, C. C. Human–robot collaboration in construction: Classification and research trends. *Journal of Construction Engineering and Management*, *147*(10), 03121006, 2021.

[6] Gamlath, M., Waidyasekara, K. G. A. S., and Pandithawatta, S. Feasibility of robotic technology for the advancement of construction industry in Sri Lanka. In *2020 Moratuwa Engineering Research Conference (MERCon)*, pages 54-59, Moratuwa, Sri Lanka, 2020.

[7] Ojha, A., Habibnezhad, M., Jebelli, H., and Leicht, R. Barrier analysis of effective implementation of robotics in the construction industry. In *Construction Research Congress 2022*, pages 661-669, Virginia, USA, 2022.

[8] Malla, V., and Delhi, V. S. K. Determining interconnectedness of barriers to interface management in large construction projects. *Construction Economics and Building*, *22*(2), 69-101, 2022.

[9] Pavitt, T. C., and Gibb, A. G. F. Interface management within construction: In particular, building façade. *Journal of Construction Engineering and Management*, *129*(1), 8-15, 2003.

[10] Lin, Y. C. Construction network-based interface management system. *Automation in Construction*, *30*, 228-241, 2013.

[11] Wideman, R.M. Wideman Comparative Glossary of Project Management Terms. On-line: http://www.maxwideman.com/pmglossary/PMG_I04.htm, Accessed: 12/10/2024.

[12] Construction Industry Institute. *Interface Management Implementation Guide*; The University of Texas at Austin, Austin, TX, USA, 2014.

[13] Fellows, R., and Liu, A. M. Managing organizational interfaces in engineering construction projects: addressing fragmentation and boundary issues across multiple interfaces. *Construction management and economics*, *30*(8), 653-671, 2012.

[14] Chen, Q., Reichard, G., and Beliveau, Y. Interface management-a facilitator of lean construction and agile project management. *International Group for Lean Construction*, *1*(1), 57-66, 2007.

[15] Zhang, S., Li, Z., Ma, S., Li, L., and Yuan, M. Critical factors influencing interface management of prefabricated building projects: Evidence from China. *Sustainability*, *14*(9), 5418, 2022.

[16] Pan, W., Parker, D., and Pan, M. Problematic interfaces and prevention strategies in modular construction. *Journal of management in engineering*, *39*(2), 05023001, 2023.

[17] Tian, X. (2013). Influencing factors for project interface management. In *ICCREM 2013: Construction and Operation in the Context of Sustainability*, pages: 448-456, Karlsruhe, Germany, 2013.

[18] Linner, T., Pan, W., Hu, R., Zhao, C., Iturralde, K., Taghavi, M., Trummer, J., Schlandt, M., and Bock, T. A technology management system for the development of single-task construction robots. *Construction Innovation*, 20(1), 96-111, 2020.

[19] Gharbia, M., Chang-Richards, A., Lu, Y., Zhong, R. Y., & Li, H. Robotic technologies for on-site building construction: A systematic review. *Journal of Building Engineering*, *32*, 101584, 2020.

[20] Eray, E., Sanchez, B., and Haas, C. Usage of interface management system in adaptive reuse of buildings. *Buildings*, *9*(5), 105, 2019.

Scaled-down Simulation and Analysis of On-site Robotic Fabrication and Construction

Raymond Xu[1], Luka Morita[2], Parmida Khosravian[1] and Anas Itani[1]

[1]Department of Civil and Environmental Engineering, University of Alberta, Canada
[2]Department of Mechanical Engineering, University of Alberta, Canada
hxu6@ualberta, lmorita@ualberta, pkhosrav@ualberta.ca, aitani@ualberta.ca

Abstract –
The Canadian construction industry faces challenges in accessing remote sites, resulting in logistical and environmental issues. Robotic construction, especially in situ or on-site assembly with industrial robots, offers a promising solution. This paper explores the feasibility and impact of on-site fabrication and construction (OSF&C) by performing a scaled down simulation with two UR5e Universal Robots. The study utilizes additive manufacturing (AM) for key construction elements to achieve increased performance. The case study involves the fabrication of a complex pentagonal-cross section bridge structure consisting of foam beams and 3D-printed brackets. The methodology includes detailed description of the fabrication and stacking procedures, material handling optimization, and human-robot interaction. The study proposes a scaled-down robotic simulation approach, which is then used to evaluate production consistency and time, demonstrating the benefits of utilizing additive manufacturing into OSF&C.

Keywords –
On-site robotic construction, additive manufacturing, industrial robots, human-robot collaboration, construction simulation, sustainability.

1 Introduction

In the Canadian construction industry, a key challenge extends beyond just a labour shortage to difficulties in accessing remote construction sites. The conventional construction approach often entails transporting workers and materials to remote or difficult-to-reach locations, resulting in greenhouse gas emissions (GHG), logistical challenges, and potential project delays. Addressing this predicament necessitates innovative solutions.

The realm of robotic construction has undergone significant exploration and adoption in recent decades. A rapidly expanding literature field of robotics in construction has made significant progress in automating field equipment and adopting robots from other industrial sectors for construction purposes (e.g., drones), and robots customised for the construction industry [2].

While robotic construction has been utilised for constructing modular building components off-site, the transportation of these pieces to the construction site contributes to high GHG emissions. An alternative approach involves on-site construction with the assistance of robotic arms, offering the potential to reduce transportation costs and environmental impact. Industrial robots, known for their flexibility, precision, speed, and cost-effectiveness, can navigate challenging areas. This has spurred increased research into novel applications of robots in construction methods, such as TIM proposed by ETH Zurich, wherein a dual robotic arm platform was used to manufacture modular wooden panels on-site, which were then assembled into the elaborate BUGA Wood Pavilion [3].

Robotic construction emerges as a promising alternative to traditional methods, particularly in situations where on-site assembly poses challenges. Off-site construction, involving the assembly of modular components in a factory, can decrease transportation frequency and distance to challenging sites, thus lowering GHG emissions during the construction phase of the project. However, substantial transportation costs and environmental impact persist when moving modules to construction sites [4].

On-site robotic construction, as opposed to off-site construction, employs mobile robotic fabrication platforms to assemble raw construction materials into modules in close proximity to the construction site. This approach circumvents the need to transport prefabricated modules and large teams of workers to remote sites, cutting down on associated GHG emissions and total project time and cost. Industrial robots, known for their flexibility, precision, speed, and cost-effectiveness, can operate in challenging environments, especially in areas with difficult access. This approach aims to facilitate the on-site assembly of modular components using a robotic

arm, which will be assembled to create a larger, complex structure. On-site robotic construction can be even further improved with AM utilization. AM is a collection of manufacturing technologies which deposit materials layer by layer to build parts based on a digital solid model in an entirely computer-controlled manner [1]. AM broadly enables more design flexibility and complexity while reducing material waste, providing two invaluable advantages in a remote construction environment.

This paper presents a method for rapidly evaluating the feasibility and performance of novel on-site robotic construction methods by using a scaled down simulation with two robotic arms. The potential for an AM enabled on-site robotic construction approach is explored, and the advantages and limitations of the scaled down simulation are discussed.

2 Literature Review

The construction industry has changed significantly with the advent of robot-automated construction, and while robotic construction automation is not yet ubiquitous in the field, many high-level studies have predicted that it has significant potential to expand the capabilities of the industry where traditional methods have stagnated [5], [6]. The McKinsey Global Institute survey found that globally, labour productivity growth in construction lagged far behind similar sectors and the overall economy [7]. Robot arms and autonomous vehicles are being employed to perform tasks traditionally carried out by human labour, enhancing precision, efficiency, and safety.

An interesting approach within this paradigm is the exploration of on-site robotic construction, which leverages robotics and automation directly at the construction site. While off-site fabrication is the conventional approach when applying robotic automation to the construction industry, it comes with many challenges such as transportation costs, logistical complexities, and limitations in design flexibility. The need for assembly and installation post-delivery can introduce delays and offset project timelines. Recent comprehensive review studies highlight the benefits of on-site robotic construction, including reduced construction time, lower transportation costs, and the ability to adapt to site-specific challenges such as irregular terrain or constrained spaces [8], [9].

Distributed In-Situ robotics, a rapidly evolving area in construction technology, has been explored through several notable projects at institutions like MIT Cambridge and ETH Zürich [3], [10]. This work demonstrates the advantage of on-site construction, emphasising the elimination of transportation-related issues, reduction in project lead times, and the ability to address site-specific constraints more effectively compared to off-site methods. Its design allows for flexibility and mobility on construction sites, enabling it to perform tasks in various locations without being confined to a fixed setup. This robotic system has been used effectively for tasks like stacking timber elements, demonstrating its capability in handling construction materials with precision [11]. Similarly, the "Digital Construction Platform" from MIT Cambridge is another transportable robotic platform. It's not just mobile but also versatile in its functions. This platform has been used for stacking bricks, among other tasks, showcasing its adaptability in different construction scenarios. The ability of these platforms to move around and perform tasks autonomously marks a significant advancement in construction methodologies, allowing for more efficient, precise, and cost-effective building processes [12].

The incorporation of AM (commonly referred to as 3D printing) technologies in on-site robotic construction presents a novel avenue of research. AM enables rapid fabrication of and design iteration of high complexity structures both in traditional off-site fabrication, as well as directly at the construction site. When applied to the on-site construction automation framework, it can further reduce the need for traditional prefabrication and assembly and improve the flexibility of the automation workflow by allowing for rapid design iterations. Recent work by Paolini et al. [13] highlights the benefits of integrating 3D printing/additive manufacturing in on-site robotic construction, including enhanced design freedom, reduced material waste, and the ability to create complex geometries with ease. A variety of AM technologies including selective laser sintering (SLS), wire arc additive manufacturing (WAAM) and material extrusion, have been applied in case studies to manufacture structural components, with steel and concrete being the most common materials [14].

One compelling reason to integrate 3D printing in on-site construction is the unparalleled design flexibility it offers. Many studies on the adoption of AM in the construction industry [15]–[18] emphasise how AM allows for the creation of bespoke structures, tailored to meet specific project requirements, surpassing the limitations of traditional construction methods. The utilization of AM also aligns with sustainability goals by minimising material waste. Studies by Javid et al. [19] and Ghaffar et al. [20] showcase how 3D printing enables precise material usage, reducing environmental impact and contributing to sustainable construction practices.

Altogether, the benefits of on-site construction compared to off-site fabrication, along with the unique advantages offered by 3D printing, provide a compelling case for combining both technologies in the constructions industry.

3 Research Objectives

Our research designed and executed a scaled down robotic simulation as a case study to evaluate the following aspects of an OSF&C scenario:

- How can a scaled down simulation of an on-site construction environment be used to explore how robots will perform modular fabrication tasks with restricted motion and human collaboration?
- How does utilization of AM affect the OSF&C workflow (complexity, speed, flexibility, etc.)?

4 Methodology

A case study is often essential for examining methods of robotic construction. The simulation employs digital twin software for path planning and utilizes two Universal Robots (UR5e) to construct a complex geometric structure incorporating additively manufactured parts. For this research project, a case study involving a complex geometric structure was developed as a key component. The fabrication and construction processes are iterated multiple times to facilitate further analysis of robotic adaptability, production performance, and human-robot interactions as part of the research effort.

The design structure, as illustrated in Figure 1 below, consists of several pentagon frames connected by lateral beams. The pentagon pattern is specifically chosen to increase the difficulty of the fabrication process. The overall structure shape can be conceptualised as a bridge or a tower. The following sections provide an overview of the simulation sequence, materials used, fabrication platform and details about the fabrication and construction procedures.

4.1 Simulation Sequence

The sequence of the simulation involved fabrication and construction, and each was taken care of mainly by each robot. The overall sequence involves the following five steps as listed below, and as shown in Figure 2 steps:

- Steps I and II are carried out exclusively by the fabrication robot. These two steps assemble a module of the structure. Minor human labour is involved with these two steps.
- Step III is completed through collaboration of two robots. In this step, the fabrication robot will indirectly lift and position the assembled module to be transported by the construction robot.
- Step IV is completed solely by the construction robot. Immediately after completion of Step III, the construction robot will stack the assembled module to another at the designated location. Human labourers will aid the stacking processes. If the stacked module is not the third, then the process will repeat from Step I again.
- Step V is the end of the process where three modules are stacked successfully by the construction robot. Both robots then move back to home position.

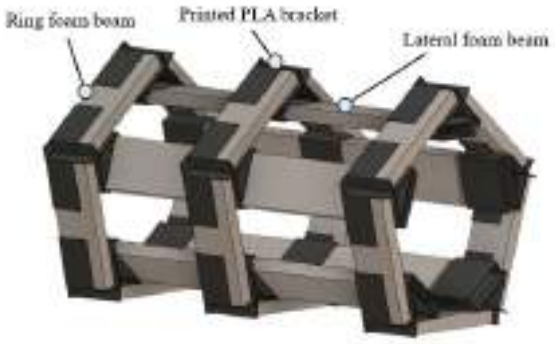


Figure 1: Pentagon Frame

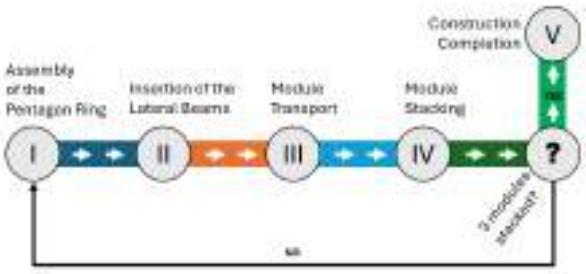


Figure 2: Construction process flow-chart

4.2 Materials Used

4.2.1 Brackets

The main functional component of the scaled-down bridge structure was a set of 3D printed brackets which were designed to hold together the members at each vertex. The brackets use AM because it enables them to be produced on-site with great design flexibility and efficiency. The bridge was held together by 20 of these brackets. Figure 3D Shows various batch sizes which were considered for printing – printing a large batch on the 60x60cm build area Modix Big 60 was avoided because if the batch of 19 failed, it would set the project back significantly. Instead, the brackets were printed out of PLA on an Ender 3 printer and an Ender 5 Plus FDM printer in batches of 2 and 5, respectively. Notable print settings for the brackets included the use of a 0.6mm nozzle and a relatively large layer height of 0.32 mm. The prints took roughly 8.5 hrs for a batch of 2, and about 28 hrs for a batch of 5. The full collection of 20 brackets was printed and post-processed over the course of two weeks. While desktop FDM printing was a suitable stand-in for this relatively small-scale project, more expensive AM

methods could be used in the field to obtain industry grade print performance.

Detailed views of the bracket design are shown in Figure 3A-C. The parts were modelled in Fusion 360 and are easily modified for various beam sizes. The brackets adhered to common principles of Design for Additive Manufacturing (DFAM), which minimised use for support structures, reduced print time, and improved quality control. The brackets have a large first layer with no excessively sharp corners in order to provide a stable first layer that will not delaminate. The brackets have no overhangs above 50-degrees, resulting in no support structures. This reduces both print time and waste. Notches were introduced to the brackets to securely grip the structural members, enhancing the connection. These notches added complexity to the part but did not add any manufacturing time to the print, taking advantage of the unique design benefits of AM. The parts were designed to fit onto an assembly platform in a pentagonal shape before inserting and joining the members. Connections on the bracket were designed so that during this stage, the robot could use force sensing to find the inner surface with the member before inserting it. This utilization of 3D printed brackets into the bridge structure increased design flexibility and reduced material waste.

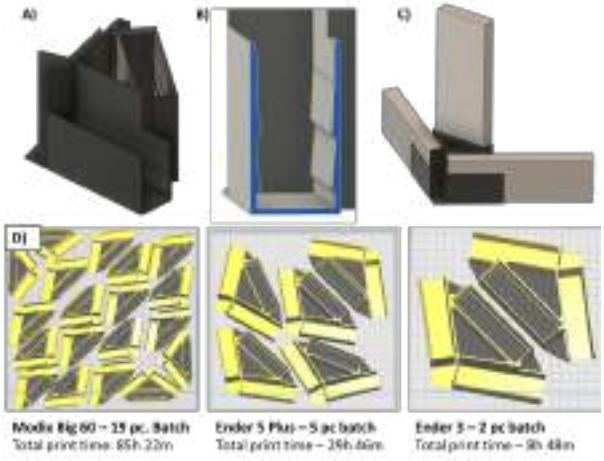


Figure 3: a) isometric view of the AM bracket. b) notch features which hook into the beams. c) A bracket with the beams inserted. d) Various batch sizes that were considered and their associated print times

4.2.2 Beams

For our simulation, all structural members were fabricated from FOAMULAR NGX F-150 2.54cm thickness foam insulation due to their combined price point, strength, and ease of cutting and shaping. The horizontal beams that made up the pentagonal cross sections (Figure 4G) had dimensions of 2.54cm x 9cm x 24cm, while the vertical beams that connected the pentagonal cross sections to each other (Figure 4F) had dimensions of 2.54cm x 10cm x 30cm. Notably, the corners of the beams were chamfered to increase the clearance during insertion, improving the ease of assembly for the robot arm.

4.3 Fabrication Platform

The base robot workspace consisted of one Universal Robots UR5E robot arm and a flat acrylic panel acting as a workbench. A key modification to this setup was the addition of a customised fabrication platform, which fit into the upper right side of the workbench as illustrated in Figure 4.

The assembly process takes place on the assembly base, which was integrated into the assembly platform for the following objectives:

1. Facilitate the assembly by incorporating guides on the surface to assist in placing fabrication materials and securing them once in place (Figure 4d). The 3D printed brackets are designed to fit snugly onto these guides (Figure 4A), improving consistency during assembly.
2. Minimise movement of materials during the assembly. The dimensions of the assembly base were designed to fit one side of the workstation well, restricting movement in three directions during assembly. This, combined with the guides, significantly reduces the potential movement of materials (Figure 4C).
3. Facilitate assembled module transfer. The design allows freedom of movement once the base (Figure 4E) is lifted by the assembly robot (Figure 4B). With an assembled module on top, the base can be lifted and tilted by the assembly robot to a designated transport position (Figure 4H). After transporting the module out of the base, the base can be de-elevated to the assembly position, ready for the next assembly process.

The assembly base represents a scaled-down version of an assembly table with customisable guides, which can be rapidly reconfigured to meet assembly requirements for a diverse range of projects. In practical terms, this assembly base can be mechanised and programmed to optimise the assembly procedure, resulting in higher degree of automation and time savings.



Figure 4: Above) the robotic workstation setup showing the pentagonal fabrication stage and the stacking stage. Below) a closeup of the assembly platform.

4.4 Fabrication Procedure

Fabrication of each module involves Stages I and II: pentagonal cross section assembly and lateral beam insertion. To begin the assembly, the fabrication robot will pick each bracket at the time and place it to the designated location. Once all brackets are placed on the platform, placement of ring beams between each pair of brackets commences, forming the pentagon ring shown in Figure 5A. Lastly, lateral beams are inserted into the designated slot of each bracket. A finished module model is shown in Figure 5C.

The material placement is facilitated by having the guide placed on the assembly base and deploying the universal robot advanced functions Insertion and Finding Surface. These combinations were optimised through iterations to minimise collision potential and trim fabrication time. Section 8 will further discuss this topic.

4.5 Construction Procedure

The on-site construction task is simulated by the second robot. The robot transports the assembled module from the assembly base at the transport position for modular stacking process (Stage III).

The stacking procedure (Stage IV) varies depending on the order of the module. For the first module, the robot places it at the ground of the designated area, and then returns to its neutral position. For the second and third, the robot firstly positions the module directly on top of the other module already placed, at approximately 2 to 3 cm. This is followed by a momentary pause to allow human labour to physically examine whether the precision toleration is met. Once clearance is granted, the robot proceeds to lower the module to a final contact position, while the lateral beams of the module slide into the designated slots of the previous module. Once the final position is reached, the robot pauses while human labour inspects the connection between modules and enhances if necessary. Once connection is verified secure, the robot releases the module and returns to its home position ready for stacking the next module.



Figure 5: Fabrication and construction procedure. a) First robotic arm lifts the assembly platform for easier access. b) the second robot pick the pentagon structure off of the platform. c) the second robot completes a cycle by stacking the pentagonal unit

5 Material Handling Optimization

In assembling the pentagon, beams of two different sizes, crafted from identical materials, are employed. These are paired with 3D-printed joints, their number matching the pentagon's edges, ensuring each beam connects to a joint. The specific spots for situating the pentagon are marked out on a foam base. Accompanying this, a 3D model is crafted, featuring an added female plug to match a corresponding male plug on the foam base, based on these markings. Additionally, for ease of assembly, the longer beam, which is oriented vertically, is shaped to fit more seamlessly into the pentagon structure, aiding the second robot arm in its assembly task. The secure contact between modules is achieved by:

- Engineered connection design: The 3D printed brackets have special features within the slot where the beams slide into making the connections. Notches are designed into the brackets, which bite into the foam material of the beams, securing them in place. This eliminated the need for adhesives or fasteners and reduced prototyping time.
- Additional reinforcement by human labour: a human labourer can choose to insert a thin shim (approximately 5 mm thick) between the bracket slot and the beam upon detection of a loose connection.

6 Path Planning and Collision Avoidance

Collision avoidance is one of the biggest challenges in the on-site construction environment due to the limited working area available and complexity of managing moving equipment, vehicles, and workers in close proximity. For this study, digital twin simulation software RoBIM was explored to plan waypoints and simulate robot maneuverer to achieve collision avoidance. RoBIM uses the modified A* algorithm to find a collision-free path between consecutive waypoints [21]. Manual waypoints were entered into the software to assist the path finding calculations. The planned pathways were examined both in simulation and in real life. Manual adjustments were made to improve overall placement efficiency. During simulation, high collision potentials were found to be associated with the lateral beam placement. Placement sequencing was optimised to avoid collision by prioritising the beams furthest from the material loading area and therefore effectively reducing the number of obstructions for the subsequent beams.

In addition to RoBIM, manual programming was utilised for the module transport phase and stacking phase, because RoBIM was unable to simulate collaboration between two robots and moving part groups. Manual programming involved examination of pick and place pathways, material placement sequencing optimization, addition of intermediate waypoints along the pathway if necessary. Each material pick-and-place manoeuvre was carefully examined to analyse collision potential with the placed material on the assembly as well as the robot body. The stacking process was also found to have high collision potential, and the collision avoidance is mainly achieved by incorporating additional waypoints along the path to maintain sufficient clearance.

7 Human Robot Collaboration (HRC)

The fabrication and construction processes of this study attempted to limit HRC required, because control of the HRC timing can optimise productivity and reduce health and safety risks to human labour. It is also because the scaled-down environment cannot properly simulate full-scaled HRC. The forces exerted by workers in the scaled-down environment cannot be proportionally adjusted for a real-world scenario. Additionally, the inspection time and reinforcement used to connect components do not scale appropriately, as the actual time required in a full-scale setting depends on the workers' skills and the size of the workforce. Moreover, in a real-world environment, additional health and safety protocols would be necessary to ensure worker protection.

The study incorporated a limited extent of human-robotic collaborations. The human-robotic collaboration is limited to just a specific step of each process, while the rest of the process is fully automated. Human labour collaborates at a specific step in each process to inspect and reinforce the connection between the beams and the brackets. Additional manual actions can be taken during collaboration should a defect be detected.

The manual reinforcement in the simulation is analogous to the connection between placed materials done by experienced technicians in the real construction environment. The connection can be in many forms ranging from welding to bolting, depending on the material type and structural requirements. Furthermore, the collaboration of the two robots was controlled by human labour. Human labour needed to initiate the program of the second robot when the assembled material was at the transport position. There remains an opportunity for automating such steps by incorporating collision avoidance sensors and programming.

8 Productivity Evaluation

The fabrication productivity was evaluated. Specifically, the fabrication process was repeated five times with the aim to evaluate production consistency and production time. Table 1 summarises the production time recorded for each run:

The results from the five iterations show acceptable time variance for each task with the standard deviation for each stage being less than 8 seconds. This demonstrates that the fabrication process is stable and can be repeated for a consistent result in terms of production timing. As discussed in Section 8, human labour had a limited collaboration step with the fabrication robot, which took in total 18 seconds to complete and less than 10% of the averaged time for the entire process.

Table 1: Build time evaluation.

Run #	Time Elapsed (second)	
	I. Pentagon Assembly	II. Lateral Beam Insertion
1	103	67
2	94	69
3	97	70
4	89	76
5	110	71
Average	99	71
Standard Dev.	7.3	3.0

In addition, the overall placement prevision was high from these five runs. All brackets were placed at their designated location. Only one ring beam was not able to be placed into its designated slot from the full runs.

9 Discussion

In this study, the down-scaled-down simulation of OSF&C with the UR5e robots to construct a complex bridge structure was successful. Bringing OSF&C to the full-scaled construction environment is expected to have similar benefits as the on-site fabrication with robots explored by Wagner et al. [3]: agile flexibility and mobility of the robot platform, reduction in GHG emission from transportation, and boosted productivity. Throughout this study, several meaningful observations and limitations were noted.

By utilizing AM in the OSF&C workflow, higher complexity, efficiency, and flexibility was observed. The use of AM to print the brackets significantly increased the part complexity available to the project, with features such as notches being added to the design with little to no increase in print time. The use of the assembly platform showed great potential for increasing consistency and quality control of robotic automation in a dynamic on-site construction environment. In future work, AM could be used to rapidly reconfigure assembly platforms to construct multiple different structures with ease. Finally, the brackets produced very little waste - only about a gram of plastic waste per bracket - showing promise for AM to play a part in sustainable construction practices in the future. The tailored bracket features made them easier for the robotic arm to work with and allowed higher overall accuracy.

Despite the successful utilization of AM, many practical limitations were encountered during the case study. Due to time and budget constraints, only one part design was manufactured with AM during the project. This is because the turnaround time for designing, validating, and batch printing multiple AM parts was a bigger bottleneck than expected. The batch printing required roughly a 2-week lead time, which was significantly longer than the construction process itself, so in this case, the brackets would likely be better off being printed at a supply plant and then brought to site alongside the other building materials. On-site printing likely works best when the print time is on the same timescale as the construction process or less. On-site printers would be best used for low production volume parts, such as quick repairs, spare parts, or jigs. Finally, fully AM-integrated construction processes, where robotic arms either directly print parts themselves or interact directly with 3D printers by retrieving and post-processing parts, could further improve OSF&C.

In this case study, a limited extent of collective robotic construction (CRC) was incorporated. It is recognised that CRC at a scaled-down environment cannot be appropriately simulated and evaluated, due to the scaled effect. This signals the need for a full-scale simulation to further evaluate and optimise CRC in OSF&C. Furthermore, collaboration of the two robots highly depended on human control. Future work should explore CRC where several robots are programmed together with appropriate collision avoidance measures including sensors and programming. CRC has seen some success in achieving complex architectural design [22].

The FDM process would likely be replaced by a higher performance type of AM in the field. In this project, the printed parts sometimes displayed unsatisfactory strength in the direction of the layers. This is a well-known issue with FDM printing, and limits its ability to be applied in end-user, load bearing parts. Instead, metal AM technologies, such as SLS or WAAM, could be explored. A benefit of SLS machines is that they can print large batches of parts quite efficiently in addition to having much better part strength. While these larger industrial printers may face their own set of issues when being integrated into OSF&C (especially regarding powder-based processes), they may show promise for shortening the manufacturing supply chain in the field and allowing rapid redesigns and changes to be made on the worksite.

In order to fully capture the benefits of AM in OSF&C, more sophisticated software tools would be needed, such as computer-aided design and manufacturing. Both on-site construction and AM need a robust digital pipeline for the design, tool-pathing and monitoring of their processes. A comprehensive building information modelling software for robotic construction simulation could significantly improve the performance of similar projects in the future.

When compared to the real-life simulation of OSF&C, there are several important features that a scaled-down scenario cannot properly model due to scaling effects. In scaled-down simulations, materials are often lighter and easier to manipulate, like polymer filament for the bracket, and foam for the beams, in place of metal bracket and timber/metal beams. The substitution simplifies the simulation but does not have the same durability or mechanical behaviour of real-life materials. This prevents evaluation of structure integrity of the constructed scaled-down structure. In addition, there are challenges with the HRC simulation in a scaled-down environment, where human labours can't be proportionally scaled up for real-life scenarios. In addition, real-life construction is a dynamic working environment that involves moving vehicles and equipment within the area. In a controlled scaled-down environment, the presence of moving vehicles and human workers is a modelling challenge. Furthermore, real-life construction would require sophisticated collision avoidance systems and protocols to ensure workers' safety, which can't be simulated in a down-scaled

environment. While the scaled-down approach used in this study could prove useful for evaluating modular OSF&C approaches such as that used by Wagner et al. [3], in the future, more rigorous productivity analysis of a simplified construction geometry would be desirable.

Dual robotic arms provided a highly articulated system, whose degrees of freedom offered complex fabrication tasks. While this allowed for manipulation of the assembly platform and efficient passing of parts between arms, capability of dual robotic arm-based construction could be pushed further in a future scaled down case study. Conversely, it would prove valuable to test scaled down OSF&C scenarios with simpler, more accessible technology such as a gantry crane.

10 Conclusion

This study demonstrates the successful evaluation of a novel AM-enabled robotic OSF&C method using a scaled down simulation approach. The down-scaled simulation is a cost-effective way to evaluate feasibility of OSF&C, with a total material cost of less than 70 dollars. The concept of OSF&C has great potential for the future construction industry with several unique benefits from reduction in GHG emissions to flexible adaptation to project needs. The case study was successful and demonstrated the feasibility and benefits of on-site robotic construction with 3D printing. The combination of industrial robots and AM offers a sustainable and efficient alternative to traditional construction methods. Despite practical limitations, the case study proposes a scaled down simulation methodology that can provide valuable insights into the potential of on-site construction automation. In order to continue the development of on-site robotic construction methods, future research is necessary to better understand the limitations of using scaled-down simulations to analyse human-robot interactions, as well as the integration of more advanced construction simulation tools such as BIM and robotic dynamics simulation.

References

[1] Bourell, D. L., Frazier, W., Kuhn, H., Seifi, M., & ASM International (Eds.). (2020). ASM handbook. Volume 24: Additive manufacturing processes / volume editors: David L. Bourell (University of Texas at Austin), William Frazier (Pilgrim Consulting LLC), Howard Kuhn (University of Pittsburgh), Mohsen Seifi (ASTM International). ASM International.

[2] B. Xiao, C. Chen, and X. Yin, “Recent advancements of robotics in construction,” Automation in Construction, vol. 144, p. 104591, Dec. 2022, doi: 10.1016/j.autcon.2022.104591.

[3] H. J. Wagner, M. Alvarez, A. Groenewolt, and A. Menges, “Towards digital automation flexibility in large-scale timber construction: integrative robotic prefabrication and co-design of the BUGA Wood Pavilion,” Constr Robot, vol. 4, no. 3, pp. 187–204, Dec. 2020, doi: 10.1007/s41693-020-00038-5.

[4] Y. Ji, K. Li, G. Liu, A. Shrestha, and J. Jing, “Comparing greenhouse gas emissions of precast in-situ and conventional construction methods,” Journal of Cleaner Production, vol. 173, pp. 124–134, Feb. 2018, doi: 10.1016/j.jclepro.2016.07.143.

[5] T. Bock, “The future of construction automation: Technological disruption and the upcoming ubiquity of robotics,” Automation in Construction, vol. 59, pp. 113–121, Nov. 2015, doi: 10.1016/j.autcon.2015.07.022.

[6] S. Cai, Z. Ma, M. J. Skibniewski, and S. Bao, “Construction automation and robotics for high-rise buildings over the past decades: A comprehensive review,” Advanced Engineering Informatics, vol. 42, p. 100989, Oct. 2019, doi: 10.1016/j.aei.2019.100989.

[7] “Improving construction productivity | McKinsey.” Accessed: Dec. 13, 2023. [Online]. Available: https://www.mckinsey.com/capabilities/operations/our-insights/improving-construction-productivity

[8] N. Melenbrink, J. Werfel, and A. Menges, “On-site autonomous construction robots: Towards unsupervised building,” Automation in Construction, vol. 119, p. 103312, Nov. 2020, doi: 10.1016/j.autcon.2020.103312.

[9] M. Gharbia, A. Chang-Richards, Y. Lu, R. Y. Zhong, and H. Li, “Robotic technologies for on-site building construction: A systematic review,” Journal of Building Engineering, vol. 32, p. 101584, Nov. 2020, doi: 10.1016/j.jobe.2020.101584.

[10] H. Chai, H. J. Wagner, Z. Guo, Y. Qi, A. Menges, and P. F. Yuan, “Computational design and on-site mobile robotic construction of an adaptive reinforcement beam network for cross-laminated timber slab panels,” Automation in Construction, vol. 142, p. 104536, Oct. 2022, doi: 10.1016/j.autcon.2022.104536.

[11] V. Helm, J. Willmann, F. Gramazio, and M. Kohler, “In-Situ Robotic Fabrication: Advanced Digital Manufacturing Beyond the Laboratory,” in Gearing up and accelerating cross-fertilization between academic and industrial robotics research in Europe:, F. Röhrbein, G. Veiga, and C. Natale, Eds., in Springer Tracts in Advanced Robotics. Cham: Springer International Publishing, 2014, pp. 63–83. doi: 10.1007/978-3-319-02934-4_4.

[12] S. J. Keating, J. C. Leland, L. Cai, and N. Oxman, “Toward site-specific and self-sufficient robotic fabrication on architectural scales,” Science Robotics, vol. 2, no. 5, p. eaam8986, Apr. 2017, doi: 10.1126/scirobotics.aam8986.
[13] A. Paolini, S. Kollmannsberger, and E. Rank, “Additive manufacturing in construction: A review on processes, applications, and digital planning methods,” Additive Manufacturing, vol. 30, p. 100894, Dec. 2019, doi: 10.1016/j.addma.2019.100894.
[14] T. Tankova and L. S. da Silva, “Robotics and Additive Manufacturing in the Construction Industry,” Curr Robot Rep, vol. 1, no. 1, pp. 13–18, Mar. 2020, doi: 10.1007/s43154-020-00003-8.
[15] N. Labonnote, A. Rønnquist, B. Manum, and P. Rüther, “Additive construction: State-of-the-art, challenges and opportunities,” Automation in Construction, vol. 72, pp. 347–366, Dec. 2016, doi: 10.1016/j.autcon.2016.08.026.
[16] P. Wu, J. Wang, and X. Wang, “A critical review of the use of 3-D printing in the construction industry,” Automation in Construction, vol. 68, pp. 21–31, Aug. 2016, doi: 10.1016/j.autcon.2016.04.005.
[17] F. Craveiro, J. P. Duarte, H. Bartolo, and P. J. Bartolo, “Additive manufacturing as an enabling technology for digital construction: A perspective on Construction 4.0,” Automation in Construction, vol. 103, pp. 251–267, Jul. 2019, doi: 10.1016/j.autcon.2019.03.011.
[18] D. Delgado Camacho et al., “Applications of additive manufacturing in the construction industry – A forward-looking review,” Automation in Construction, vol. 89, pp. 110–119, May 2018, doi: 10.1016/j.autcon.2017.12.031.
[19] M. Javaid, A. Haleem, R. P. Singh, R. Suman, and S. Rab, “Role of additive manufacturing applications towards environmental sustainability,” Advanced Industrial and Engineering Polymer Research, vol. 4, no. 4, pp. 312–322, Oct. 2021, doi: 10.1016/j.aiepr.2021.07.005.
[20] S. H. Ghaffar, J. Corker, and M. Fan, “Additive manufacturing technology and its implementation in construction as an eco-innovative solution,” Automation in Construction, vol. 93, pp. 1–11, Sep. 2018, doi: 10.1016/j.autcon.2018.05.005.
[21] C. Yang and S. Kang, “Collision avoidance method for robotic modular home prefabrication,” Automation in Construction, vol. 130, p. 103853, Oct. 2021, doi: 10.1016/j.autcon.2021.103853.
[22] S. Leder and A. Menges, “Architectural design in collective robotic construction,” Automation in Construction, vol. 156, p. 105082, Dec. 2023, doi: 10.1016/j.autcon.2023.105082.

A proposed strategy for automating Intermeshed Steel Connection assembly using robotics

Miftahur Rahman[1], Samuel Adebayo[2], David Hester[2], Daniel McPolin[2] and, Karen Raferty[3], Ibukun Awolusi[4], Debra F. Laefer[5,6]

[1]Department of Electrical, Electronic and Robotic Engineering, Kingston University, UK
[2]School of Natural and Built Environment, Queen's University Belfast, UK
[3]School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, UK
[4]School of Civil & Environmental Engineering, and Construction Management, University of Texas at San Antonio, USA
[5]Department of Civil and Urban Engineering, Tandon School of Engineering, New York University, USA
[6]Center for Urban Science and Progress, Tandon School of Engineering, New York University, USA
m.rahman@kingston.ac.uk, samuel.adebayo@qub.ac.uk, d.hester@qub.ac.uk, d.mcpolin@qub.ac.uk, k.rafferty@qub.ac.uk, ibukun.awolusi@utsa.edu, debra.laefer@nyu.edu

Abstract – The assembly of steel structures is a complex and labour-intensive process that poses significant safety risks to workers. The Intermeshed Steel Connection (ISC) is an emerging type of structural steel connection that requires fewer bolts and can benefit from the use of robotics. The assembly process for traditional structural steel connections requires humans to work-at-height, which is unsafe and unergonomic, and the few robotic and automatic steel building assembly systems are not readily adoptable. In response, this paper proposes a novel strategy for automating the ISC assembly using off-the-shelf robotic manipulators. The proposed method aims to enhance safety and efficiency by incorporating a collaborative human-robot approach in the assembly process, which leverages computer vision and robotic manipulation technologies. A detailed workflow for the robotic assembly of the ISC is conceptualized by dividing the assembly process into multiple stages. A small-scale experimental setup has been created for future demonstration. The experimental setup includes a scaled-down ISC column-to-column connection, utilizing a KUKA LBR iiwa14 manipulator equipped with multiple cameras and computer vision algorithms for precise alignment and assembly.

Keywords – Intermeshed Steel Connection, robotic assembly, structural steel connections

Introduction

The process of assembling a steel building is a thoroughly organized sequence of steps that ensure the structure's integrity, functionality, and adherence to safety standards [1], [2]. Recent coupling of robotic arms and advanced manufacturing cutting enables a simpler connection that uses fewer pieces, and, thus, may be better suited for robotic assembly in the field. To demonstrate this concept, this paper uses the recently developed Intermeshed Steel Connection (ISC) to illustrate this concept [1], [3]. This section is divided into multiple sub-sections beginning with the steel structure assembly process in section 1.1. Next, section 1.2 describes the state-of-the-art robotic steel structure assembly process. Then, a brief introduction on ISC is provided in section 1.3. Finally, the limitations of the current SOTA and contribution of this research are described in section 1.4.

1.1 Steel structure assembly process

Steel buildings are comprised of individual beams and columns, which are prefabricated in a manufacturing shop floor. These beams and columns are then transported for on-site assembly. After site preparation and foundation installation, columns and then beams are assembled sequentially according to the construction program. Erection and assembly of structural steel buildings are a critical process comprising many uncertain activities [4] and heavily rely on manual labour [5]. The labour cost for assembly accounts for almost 25% of the total steel building cost [6]. Onsite assembly can be divided into four main steps: 1) lifting the structural steel members from storage area, 2) transporting them to the installation area, 3) aligning each member according to the design and with consideration of what is already assembles, and 4) connecting with bolts and/or welding [7]. There are several risks and safety concerns associated with the current assembly process. Human

workers must work-at-height for significant periods of time during the alignment and connection steps. Another safety risk is the size and weight of the beams and columns. As per British Standard, the depth in the cross-sectional size of a beam can vary from 127.2mm to 1036.3mm, while for column it can vary from 152.4mm to 474.2mm. Moreover, the length of the beams and columns can be any length up to 12m depending on the design. The weight of beams and columns range from 1 tonne to 10 tonnes based on the cross-section and length. Manipulating these large, heavy objects during the alignment step of the assembly pose safety risks to human workers. In many cases, the human worker must conduct the assembly process by standing on a beam, which is unergonomic and has inherent safety risks, as shown in Figure 1 [8]. The greatest risk of overall structural collapse is during the assembly process prior to the building achieving full stability and strength once all connections are made.



Figure 1 Safety risks of a human working at height during steel building construction [5]

1.2 State-of-the-art (SOTA) robotic technologies used in steel structure assembly

The use of robotics to automate assembly operations has grown in prominence in different industries including warehouse, automotive, and electronics [9]. Robotics have the potential to increase assembly processes' efficiency, precision, and safety by substituting intelligent machine systems that can quickly and accurately complete repetitive tasks, as compared to human labour. In an effort to make steel assembly safer and more profitable, researchers on robotics in the construction industry is a growing field.

The Shimizu Manufacturing system by Advanced Robotics Technology (SMART), is a fully automated construction system for high-rise buildings [10]. The system weighed 1,200 tonnes and required over two months to install. It faced issues with low productivity and financial viability. To address the challenges of transporting large and heavy objects and placing it precisely, a rail-boom robotic system and wire-suspended positioning system were proposed but not implemented [8]. In Korea, a robotic beam assembly system, including a robotic transport control system and a robotic bolting device were developed and tested [11], [12]. However, both proposed systems depended upon the creation of a large Construction Factory (CF) and a guide rail mechanism. To date, while industrial manipulators have been used extensively for automating many repetitive tasks, there are very few attempts to use them for the steel structure assembly process. Work by Liang et. al experimented with a scaled-down Robotic Assembly System (RAS) to automate the alignment and bolting process with an industrial manipulator and specially designed rotational device [7]. The manipulator transported both the beam and rotational device prior to correct alignment with the help of a marker and camera. However, how the system would transport a large and heavy beam was not considered.

1.3 Intermeshed Steel Connection

Although bolting requires assembling a large number of pieces, it is favoured over field welding due to easier installation, quicker connection times, better quality control, reusability, and lower inspection demands. However, the typically high number of bolts necessary for structural connections adds complexity to the use of field robotics. Advances in manufacturing techniques have enabled the development of new steel connections that could significantly reduce the number of bolts required, particularly in column-to-column connections. The Intermeshed Steel Connection (ISC) [1], [2], [3] is one such innovation, designed to speed up assembly and offer the possibility of direct reuse. Figure 2 displays an ISC connection in 3D CAD design software and from previous experimental project. The connection leverages advanced manufacturing capabilities to create precision cuts at the ends of the structural members, which allow a quickly assembled, load transferring connection with fewer bolts and auxiliary members to transfer the various load modalities (e.g. compression, shear, bending).

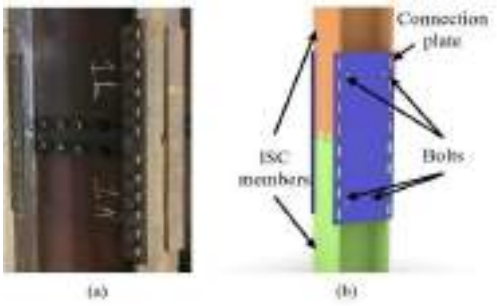


Figure 2 Proposed ISC; a) fabricated ISC with initial design [3] and b) CAD design of ISC with

alternative design with only two connection plates and fewer bolts [1],[2].

1.4 Limitations of SOTA and contribution of this research

Most of the proposed robotic steel structure assembly systems to date require customized robotic devices and structural systems before starting the actual construction. Use of these robotic or automatic assembly systems would require significant changes in the conventional construction method and mindset of the construction company and worker. A new type of structural steel connection, ISC, has been introduced that follows a similar assembly process as the conventional structural steel, but with many fewer bolts, thus lending itself more favorably to robotic assembly. The following research describes the design and experimentation of a robotic assembly process for the ISC using an off-the-shelf manipulator avoiding the need for significant customization. The main contribution of this research will be:

1. Conceptualizing a feasible assembly workflow for ISC using an off-the-shelf manipulator
2. Developing a dummy construction site with small-scale column-to-column ISC for future demonstrations

To place these contributions in context, section 2 of this paper describes more of the ISC assembly process. Section 3 discusses the theoretical framework for robotic assembly of the ISC. Section 4 introduces the proposed workflow of the assembly process, and the final section concludes with the results, the shortcomings, and future recommendations.

2 ISC assembly process

Before conceptualizing and developing the theoretical framework of ISC assembly with robots, it is required to understand the elements of ISC and the manual assembly process. A total of 3 types of elements are required for a complete ISC connection: (1) the ISC member, (2) connection plates, and (3) bolts. Each ISC require 2 structural members and 2 connection plates.

Assembling the members in an ISC requires matching the corresponding grooved sections of the steel members and then locating corresponding cuts that match those on the connection plates' edges. Similar to traditional structural steel members, assembly of ISC require 4 steps (lifting, transporting, aligning, and bolting) [7]. Figure 3 illustrate the steps of manual beam-to-beam ISC assembly inside a lab environment. The assembly process starts with attaching the member with the crane and then transporting it to the designated area as shown in Figure 3(a) and Figure 3(b) respectively. After that, a human worker aligns the member and connection plates (Figure 3(c)) and inserts the bolts (Figure 3(d)). The current ISC assembly process reduces but does not eliminate risks related to human safety and ergonomics. Using robotic technologies to assemble the elements of ISC could further reduce the time and requirement of human working at height and in awkward postures, which will improve safety and ergonomics.

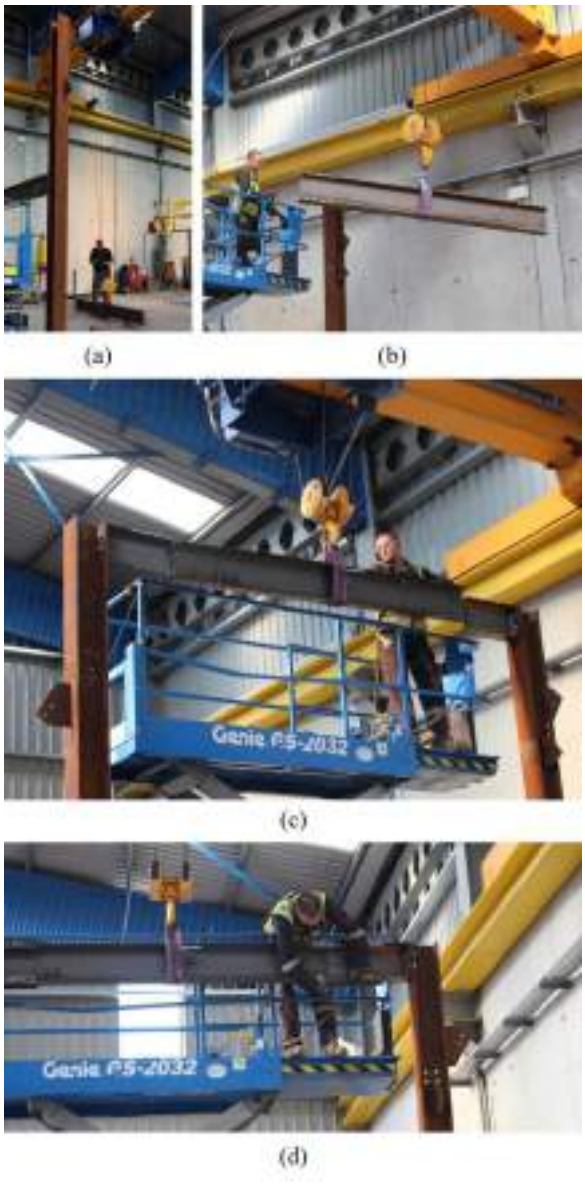


Figure 3 Steps of ISC assembly; a) lifting, b) transporting, c) aligning, d) bolting

3 Theoretical framework for robotic assembly process of ISC

ISC provides advantages over the conventional connection as it requires fewer bolts and improves the possibility of automation [1]. ISC members contain grooved sections, while conventional structural steel members contain flat ends. With the advancement of sensors, computer vision techniques, and accurate manipulation capabilities of industrial robots, robots have been proven to be efficient in repetitive tasks. With the help of computer vision technology and robotic manipulation, humans can potentially be removed from the dangerous task of assembling ISC. This section discusses the theoretical framework for using robotic technologies in the ISC assembly process.

3.1 Robotic assembly process of ISC

Though the manual assembly process of traditional steel connection and ISC, described in Figure 3, has 4 steps, robotic assembly of ISC will have 5 steps. Third step of the assembly process, aligning, will be divided into 2 steps, alignment and connection plate attachment. The whole assembly process can be divided into three main stages: pre-assembly, assembly, and post-assembly stage as shown in Figure 4.

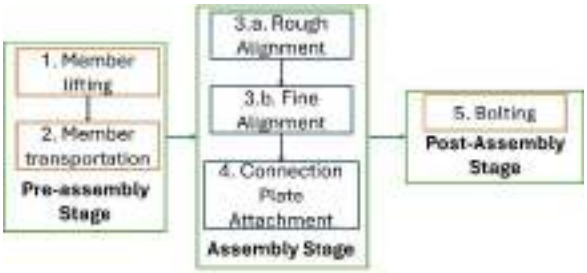


Figure 4 Stages and steps of robotic ISC assembly

Pre-assembly stage consists of lifting and transporting the member to the assembly site with the help of a human operator and crane. With the help of multiple cameras and computer vision algorithms, a robotic manipulator will assemble the member and connection plate during the assembly stage. After the rough alignment, the lateral distance between 2 members of ISC will be around 10-20 cm. but the orientation of the members might be incorrect, as shown in Figure 5(a). However, the rotation, translation correction, and fine placement will be conducted in the fine alignment step, shown in Figure 5(b). Subsequently, the connection plate is attached to the structural members. Finally, a human will put the nuts and bolts in place, which will be considered a post-assemble stage.

3.2 Detail workflow of assembly process

As the steps of assembly and usage of robotic technologies have been determined in Figure 4, this yields a detailed workflow of the whole assembly, as shown in Figure 6. The workflow starts with getting information regarding the member pickup location, assembly location, member size and shape from Building Information Modelling (BIM) through the communication module into the Human Machine Interface (HMI).

This assembly process requires lifting and transporting heavy metals by robot in different stages. Any small error in the robot program or incorrect path planning may lead to catastrophic disaster and cause human deaths. Hence, a strict separation of workspace and task must be maintained for safety reason. The HMI is used to maintain communication between robot and human, initiate tasks, and provide feedback for completion of tasks.

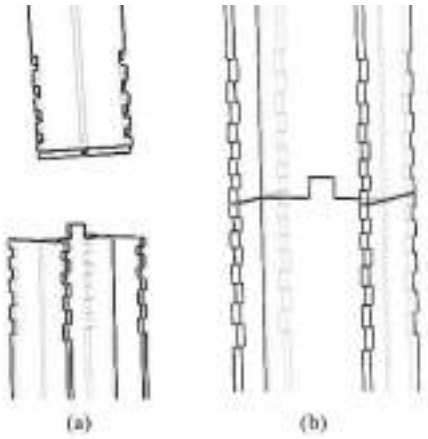


Figure 5 Alignment of members in assembly stage; (a) rough alignment, (b) fine alignment

With the introduction of Industry Foundation Class (IFC), BIM became a popular tool in managing construction projects by representing the information digitally [12]. A human will attach the rope and lifting attachment on the member. Once this is done, the member will be transported to the assembly area from the pickup location with the help of a crane operated by a human. During the whole assembly process, the crane will carry the member's full weight.

In the beginning of rough alignment, a perception module detects the object and determines the pose, with the help of multi-view cameras and trained computer vision algorithms. After receiving pose estimation through the communication module, the robot will move the member to desired location. Control and safety module of the robot will determine the presence of any obstacle in the planned path. If it is unavoidable,

feedback will be sent to the operator for manual intervention. Otherwise, the robot will try to finish rough alignment by trying new paths.

To assisting this step, an identity tag will be attached in the grooved section of the ISC members, which will provide better pose information for the robot perception. The perception module will detect and determine the pose of the identity tag to find the exact placement location. The robot will then correct the rotation and translation of the member, while it hangs from the crane. Once both members are aligned properly, the suspended member will be slowly lowered into the exact connection spot as guided by the robot with the help from the crane operator.

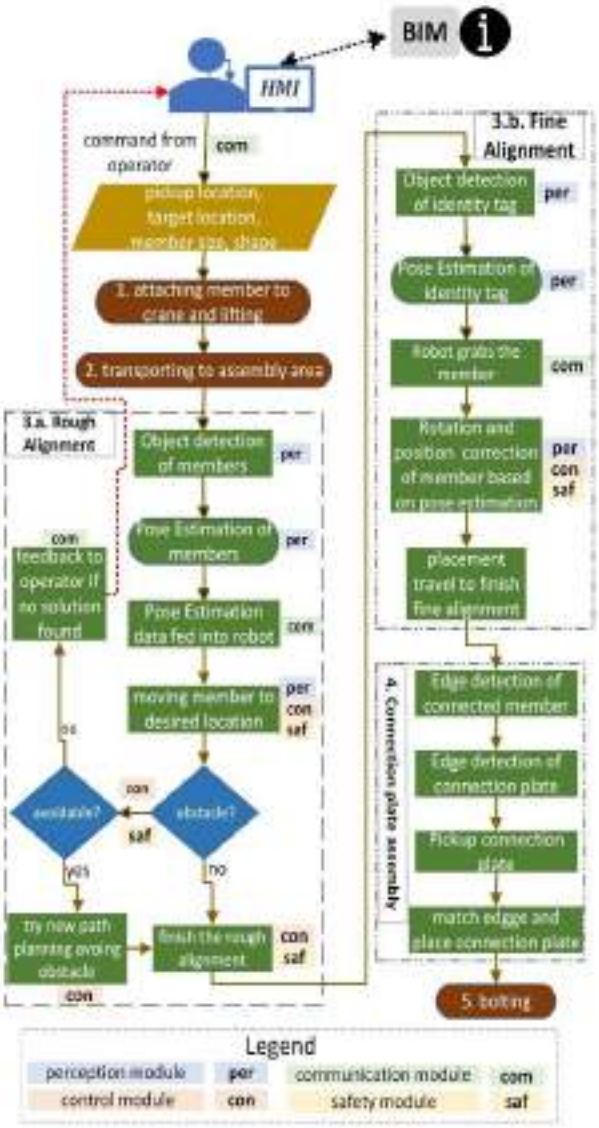


Figure 6 Detail workflow of ISC assembly using robot

The next assembly process step is placing the connection plate. An edge detection algorithm of the perception module will calculate both the edge of the grooved section in the connected members and connection plate. By matching the edges, the robot will align the connection plates on the members. After that, a human will attach the nut and bolts to finish the assembly process.

3.3 Schematic of small-scale demonstration

A scaled construction site along with ISC column members will be developed as a part of the second contribution, which will be used to test the previously described theoretical framework. Figure 7 shows the schematic diagrams of dummy construction site for small-scale ISC assembly with the help of robot. A total of 3 cameras (referred as C1, C2, and C3 in the diagram), will be fitted to create a multi-view perception. One camera will look at the site from top, while the other two will look at the scene from front and back, respectively. Specific areas will be designed for material input and assembly location.

4 Discussion

ISC assembly was divided into multiple stages and a detailed workflow combining manual and robotic steps was proposed.

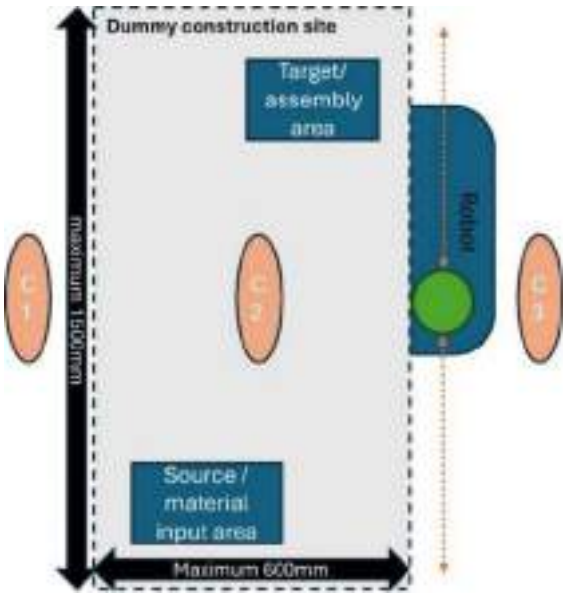


Figure 7 Schematic diagram of small-scale construction site for assembly of ISC using robot

4.1 Manual and robotic process

In this theoretical framework, 2 stages – pre-assemble

and post-assemble stage – were considered as manual processes, while the assembly stage was considered as a robotic process. Based on the size of the beam and column, members can weigh up to 10 tonnes and be up to 10m in length. Manipulating such a long and heavy object with an industrial robot is likely to be unsafe. Hence, the lifting and transporting steps in the pre-assemble (Figure 4) stage will be conducted by a human operating a heavy-duty crane. A special end-effector or gripper will be required to perform the bolting operation as bolting operation requires continuous rotational movement with high torque. In the current theoretical workflow, once all the elements are assembled, a human can quickly tighten the bolts which will significantly reduce the amount of time for working-at-height. In the future bolting will also become a robotic operation.

Aligning the ISC members is dangerous as the workspace is tight and at-height, and the object is heavy. During the assembly stage, machine learning based object detection and pose estimation algorithms can provide accurate results. Moreover, while the members are supported from a crane, a robot can guide the members and connection plate in place.

4.2 Robotic module and their functions

As shown in Figure 6, there are four main robotic modules: perception, control, safety, and communication, in the assembly workflow. These modules and their functionality are required in different steps to ensure the successful assembly of ISC.

Perception of a robot is the ability to sense and interpret the surroundings. For the proposed ISC assembly process, multiple cameras will be used to give the robot to ability to see the environment. Moreover, a trained computer vision algorithm will be used for detection and pose estimation of the assembly components. However, classical edge detection algorithm will be used in the connection plate assembly steps.

As ISC is novel type of connection, there are very few real images of ISC which are inadequate for training any CV algorithm. Hence, a hybrid dataset containing real and synthetic images is currently being generated. Performance of the CV algorithm will be tested in the simulated environment, small-scale experimental setup, and full-scale construction environment and the result will be shared with the research community.

The control module works as the brain of the whole system. It will be responsible for path planning of the manipulator to pick and place the ISC elements while avoiding obstacles with the help of perception and safety module.

The safety module is the part of the software architecture that ensures the safety of the robot and human during the assembly process. By analysing the data from the camera, the safety module will be able to determine any static or dynamic obstacles in the planned path. Finally, the communication module will be responsible for transferring necessary information among different modules and HMI.

4.3 Robot transportation

This proposed workflow herein avoids the current strategy of building another structure to support the robotics. However, the currently proposed workflow assembly of the ISC is divided into stages between manual and robotic process and uses human-robot collaboration to complete the assembly process. As part of this, a crane will be used to transport the robot between places to perform the rough alignment and fine alignment during the assembly stage. Ultimately, these tasks are repetitive in nature and can be replicated with robot and computer vision technology.

4.4 Small-scale assembly scene

To demonstrate the proposed workflow of ISC assembly and find affiliated challenges and shortcomings, a small scale ISC column to column connection was fabricated, and an experimental setup was created with a manipulator and multiple scene cameras.

To fabricate the small-scale column connection, universal column dimensions from the BS EN10056:1999 standards were used. One universal column, 305x305x97, was selected, and the sectional depth and width was scaled down to 103mm and 100mm respectively, with a ratio of 1:0.33 compared to the original dimensions. In this installation, the ISC column connection comprises three elements; bottom part (fixed with base), top part, and connection plates.

In the future demonstrations of the small-scale ISC assembly, a KUKA LBR iiwa14 will be used as a manipulator. It has 7 degree-of-freedoms and can lift a maximum payload of 14kg with a maximum reach of 820mm. That manipulator is also equipped with an Intel RealSense D-435i depth camera for perception and a 2-finger parallel gripper from Robotiq. Furthermore, 2 RGB cameras (one in the front and one in the back and one depth camera for aerial view) are attached in the scene. Figure 8 shows the scene setup for a small-scale column-to-column connection for future demonstrations of the assembly process.

4.5 Performance analysis

Few research had been conducted previously [7], [8], [10], [11], [12] to automate the steel building assembly using conventional steel connections. However, ISC is a novel connection type which has emerged from recent research. There was no research found in the literature

which use robots or completely manual processes for steel building assembly using the ISC connection. This proposed framework reduces the amount of time for work at height as the lifting, transportation and alignment will be carried out by the robot, as shown in Figure 4 and 6. Hence, using robot will improve the safety of the human worker and reduce the number of accidents.

Accuracy of our proposed assembly process depends on the accuracy of CV algorithm in the detection and localisation of ISC member and ISC connection plate and by extension the robot manipulation in fine alignment and connection plate attachment. Moreover, the overall time required for the assembly will depend on the completion time of each step. As this research is currently ongoing, details on the accuracy and time can be calculated only after conducting small-scale and full-scale assembly.

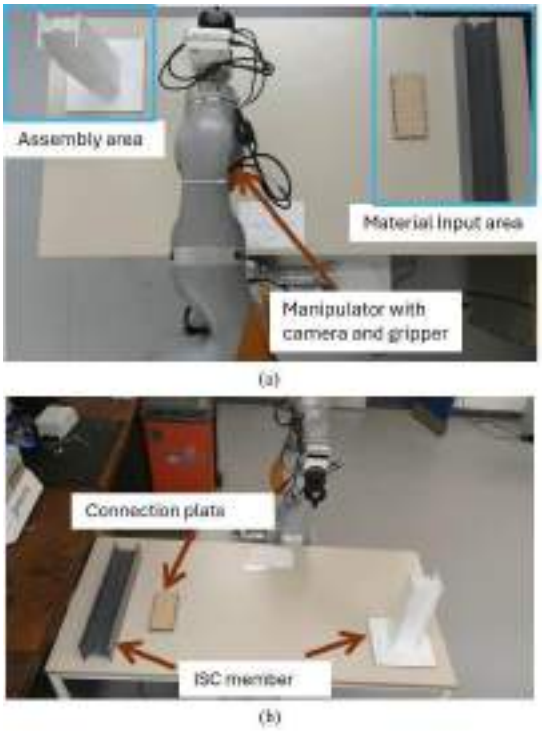


Figure 8 Demonstration scene setup with small scale ISC column to column connection; (a) aerial view, (b) front view

5 Conclusions

This research presents a novel strategy for automating the assembly of Intermeshed Steel Connections (ISC) using off-the-shelf robotic manipulators. The proposed method aims to enhance safety, efficiency, and precision in the assembly process by leveraging computer vision and robotic manipulation technologies. By conceptualizing a detailed workflow and creating a small-scale experimental setup, this study demonstrates the feasibility of integrating robotic technologies into conventional construction workflows without extensive modifications.

The experimental setup, which includes a scaled-down ISC column-to-column connection, highlights the potential of the proposed method to reduce manual labour and improve overall safety and ergonomics. The results indicate that the robotic assembly of an ISC can significantly mitigate the risks associated with human workers handling heavy components at height.

However, the study also identifies several limitations and areas for future research. The current setup is limited to small-scale demonstrations, and further work is needed to optimize the system for large-scale applications. Additionally, the integration of advanced computer vision algorithms and the development of specialized end-effectors for bolting operations are crucial for enhancing the system's performance. Another limitation is the need for a process to synchronise the motion between robot and crane. Future research will focus on refining the robotic algorithms and modification of end-effector for automating the bolting step, improving system integration for robot and crane synchronized motion, design of an HMI for the assembly process command and control centre, and conducting extensive field trials to validate the system's performance in real-world scenarios.

Funding

The authors gratefully acknowledge the financial support provided by National Science Foundation (NSF) award 2222815, by Science Foundation Ireland award 21/US/3797, and by the Department for the Economy (DfE) through grant USI-218.

References

[1] Al-Sabah, S., Laefer, D. F., Truong Hong, L., Phuoc Huynh, M., Le, J.-L., Martin, T., Matis, P., McGetrick, P., Schultz, A., Shemshadian, M. E., & Dizon, R. (2020). Introduction of the Intermeshed Steel Connection—A New Universal Steel Connection. Buildings, 10(3), 37. https://doi.org/10.3390/buildings10030037

[2] Al-Sabah, S., & Laefer, D. F. (2020). Patent No. US 20200378107 A1. United States Patent and Trademark Office.Chi HL, Kang SC. A physics-based simulation approach for cooperative erection activities. Automation in Construction, 19:750–61, 2010. https://doi.org/10.1016/j.autcon.2010.03.004.

[3] Shemshadian, M.E., Labbane, R., Schultz, A.E., Le, JL, Laefer, D.F., Al-Sabah, S., & McGetrick, P.

(2020). Experimental study of intermeshed steel connections manufactured using advanced cutting techniques. Journal of Constructional Steel Research , 172 , 106169.

[4] Chi HL, Kang SC. A physics-based simulation approach for cooperative erection activities. Automation in Construction, 19:750–61, 2010. https://doi.org/10.1016/j.autcon.2010.03.004.

[5] Irizarry J. Application of methods analysis to steel erection tasks: Identification of factors affecting labor performance. *International Journal of Construction Management*, 11:31–47, 2011. https://doi.org/10.1080/15623599.2011.10773177.

[6] Carter CJ, Murray TM, Thornton WA. Cost-effective steel building design. *Progress in Structural Engineering and Materials*, 2:16–25, 2000. https://doi.org/10.1002/(sici)1528-2716(200001/03)2:1<16::aid-pse3>3.0.co;2-q.

[7] Liang CJ, Kang SC, Lee MH. RAS: a robotic assembly system for steel structure erection and assembly. *International Journal of Intelligent Robot Applications*,1:459–76, 2017. https://doi.org/10.1007/s41315-017-0030-x.

[8] Lee S-K, Doh NL, Park G-T, Kang K-I, Lim M-T, Hong D-H, et al. Robotic technologies for the automatic assemble of massive beams in high-rise building. In *Proceedings of the International Conference on Control, Automation and Systems*, p. 1209-1212, Seoul, Korea (South), 2007. doi: 10.1109/ICCAS.2007.4406518. 2007.

[9] Jiang Z, Messner JI, Matts Professor E. Computer vision applications in construction and asset management phases: A literature review. *Journal of Information Technology in Construction (ITcon)*, 28:176–99,2023. https://doi.org/10.36680/j.itcon.2023.009.

[10] Yamazaki Y, Maeda J. The SMART system: An integrated application of automation and information technology in production process. *Computers in Industry*, 35(1), 1998. https://doi.org/10.1016/S0166-3615(97)00086-9.

[11] Jung K, Chu B, Hong D. Robot-based construction automation: An application to steel beam assembly (Part II). *Automation in Construction*, 32:2013. https://doi.org/10.1016/j.autcon.2012.12.011.

[12] Chu B, Jung K, Lim M-T, Hong D. Robot-based construction automation: An application to steel beam assembly (Part I). *Automation in Construction*, 32:2013. https://doi.org/10.1016/j.autcon.2012.12.016.

[13] Sheik NA, Veelaert P, Deruyter G. Exchanging progress information using IFC-based BIM for automated progress monitoring. *Buildings*, 13:2390,2023, https://doi.org/10.3390/BUILDINGS13092390.

Paraphrasing-Based Data Augmentation for Responsible Personnel Classification in Crane Accidents

Deokyeong Kim[1], Leila Kosseim[2], Hongjo Kim[3], and Jong Won Ma[1]

[1]Dept. of Building, Civil and Environmental Engineering, Concordia University, Montreal, QC, Canada H3G 1M8
[2]Dept. of Computer Science and Software Engineering, Concordia University, Montreal, QC, Canada H3G 2W1
[3]School of Civil and Environmental Engineering, Yonsei University, Seoul, South Korea 03722
deokyeong.kim@mail.concordia.ca, leila.kosseim@concordia.ca, hongjo@yonsei.ac.kr, jongwon.ma@concordia.ca

Abstract –

Crane accidents often result in severe injuries and fatalities, necessitating accurate analysis and responsibility prediction to prevent future incidents. However, the scarcity of crane accident-specific datasets and class imbalance pose challenges to developing robust predictive models. This study addresses these challenges by leveraging paraphrase-based data augmentation methodology to expand a dataset of 480 natural language accident descriptions into a balanced dataset of 900 samples. Using Google's Pegasus paraphrasing model, the augmented data improved model performance, as demonstrated by the F1 score increase from 0.29 to 0.5583. Three pre-trained transformer BERT-based models were fine-tuned on the augmented dataset to evaluate their effectiveness in predicting responsible personnel. The results highlight the effectiveness of the paraphrasing technique and transformer-based models in addressing class imbalance and improving classification accuracy. This research demonstrates the potential of natural language text data in enhancing safety analysis and proposes future directions, such as exploring alternative data augmentation methods, leveraging large language models, and experimenting with domain-specific fine-tuning.

Keywords –

Crane Safety, Natural Language Processing, Deep Learning, Data Augmentation.

1 Introduction

Crane accidents frequently occur on construction sites and, due to their heavy nature, often result in severe injuries and fatalities. Previous studies indicate that these accidents are primarily caused by human errors and miscommunication [1-3]. Forensic analysis after crane accidents plays a crucial role in identifying the sequence of events, root cause and pinpointing the responsible personnel. Identifying responsible personnel in crane accidents is critical for improving workplace safety and accurate liability assessment [4]. This research aims to support such processes by providing structured methods to analyze accident reports, making it easier to derive actionable insights for future safety improvements. This research output can potentially be used as an auxiliary tool to provide additional layer of verification and increase overall efficiency in experts' decision-making process. The construction industry generates significant amounts of text-based data, such as incident reports, safety manuals, and communication logs. Text-based analysis has emerged as a powerful tool in safety domains, including construction and manufacturing, for identifying accident processes [5]. Unlike accident statistics that only provide predefined categories and post-incident outcomes [6], natural language-based accident reports capture richer information about conditions, causes, and injuries [7]. For instance, Ma and Chen analyzed 159 text-based reports to identify accident factors [8], while Kumi et al. classified construction accident types using Korean reports [9]. These studies demonstrate the potential of text-based analysis in uncovering critical insights from unstructured data.

However, analyzing crane accident reports poses significant challenges. First, crane accident-specific datasets are scarce, limiting the ability to develop robust models. And this scarcity in labeled data can introduce overfitting during the model training [10]. Second, class imbalance in available datasets—where roles like "Crane Operator" are overrepresented while others like "Owner/User" are underrepresented—hampers the predictive accuracy of machine learning models.

To address these challenges, this study investigates paraphrase-based data augmentation strategies [11-12] for identifying responsible personnel in crane accidents. Using 480 natural language accident descriptions, the dataset was augmented with Google's Pegasus model [13], resulting in a balanced dataset of 900 samples. This augmentation ensures diverse yet semantically consistent training data, mitigating the impact of class imbalance.

Additionally, three pre-trained BERT-based models [14] were fine-tuned on the dataset to evaluate their performance in predicting responsible personnel. This approach aligns with findings by Devlin et al. (2019), which highlight that fine-tuning the entire model on task-specific data, even with limited examples, can significantly enhance performance compared to using fixed embeddings [14].

2 Data

2.1 Data Preprocessing

The original dataset contains 710 crane accident records provided by CRL (Crane Risk Logic, Inc.), where each record includes surveys of the accident descriptions, primary responsible personnel, and more. After removing empty contents and where the target variable (responsible personnel) is not available, a total of 480 crane accident descriptions and their primary responsible personnel were extracted from the original dataset. In this study, we removed or substituted any sensitive information (including personnel names, specific locations, and company identifiers) to protect the privacy of all parties involved. Each crane accident description contains an average of 93 words, ranging from two to three lines to 300 words. There are nine classes for the primary responsible personnel, with 'Crane Operator' being the most frequent class (100 records), while the 'Owner/User' class has the fewest (14 records) (Table 1).

Table 1. Original data per class distribution (total 480)

Class	Name	# of Data
0	Crane Operator	100
1	Rigger	96
2	Lit Director	94
3	Site Supervisor	58
4	Manufacturer	39
5	Other	27
6	Mechanical/Maintenance Issue	27
7	Signal Person	25
8	Owner/User	14

2.2 Input data Preparation

2.2.1 Paraphrase

Due to the imbalanced nature of the original data, this study adopted a paraphrasing technique for input dataset generation. The overall process of data augmentation is shown in Figure 1. A total of five datasets were used in this study. The first dataset contains only the original data, comprising 480 crane accident descriptions. The second dataset is augmented using Google's PEGASUS (Pre-training with Extracted Gap-sentences for Abstractive Summarization) paraphrasing model [13] . This ensures that each class contains 100 balanced samples (Figure 2), resulting in a total of 900 samples for Dataset 2. The third dataset is also generated using paraphrasing; however, the test set was removed in advance before paraphrasing to prevent any potential data leakage during training. For fair comparison, the test dataset size was fixed at 63 samples and remained the same across all three datasets. For the imbalanced original dataset, where the lowest class has only 14 samples, 50% of the smallest class's data count was used as the criterion for constructing the test set. As a result, 7 samples per class were selected, totalling 63 samples for the test set. Table 1 shows the original data distribution per class, and Table 2 data splits for each dataset.

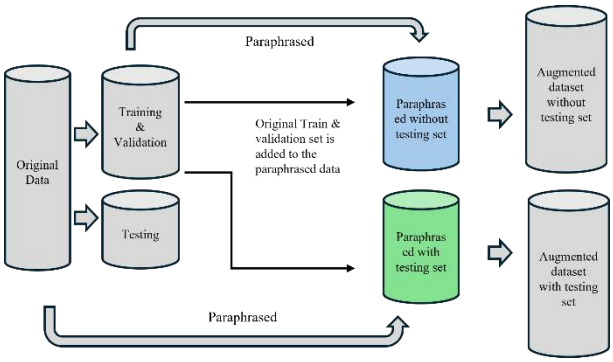


Figure 1. Data split process

Table 2. Description of five dataset used and dataset size

Dataset	Is Test set used for paraphrasing?	Is Test set balanced?	Data size (train-validation, test)
1	No	Yes	(417, 63)
2	No	Yes	(837,63)
3	Yes	Yes	(837,63)
4	No	No	(417, 63)
5	Yes	No	(837,63)

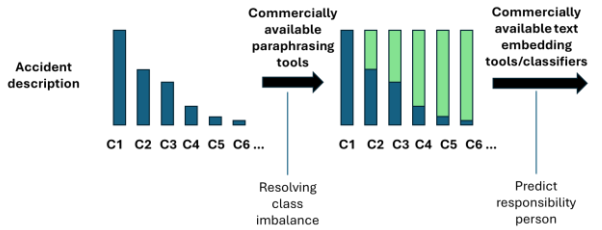


Figure 2. Data augmentation overview

2.2.2 Hyperparameters

The length of the paraphrased text was set to a

minimum of 80% and a maximum of 120% of the original target text. Top-p sampling was used as the sampling method, applied with a cumulative probability threshold (top_p=0.9) to ensure that words were sampled from the most probable token distribution while maintaining linguistic coherence [15]. Only the top 90% of cumulative probability words were considered, and the next word was selected based on normalized probabilities among these options.

In our paraphrase generation task, the temperature parameter was set to 1.5 to control the randomness of token selection during sampling. A temperature value of 1 results in a neutral probability distribution for next-word selection, while values greater than 1 increase randomness by making the probability distribution more uniform. This adjustment enhances output diversity but may occasionally reduce coherence or grammatical accuracy [16].

2.2.3 Paraphrasing quality evaluation

In this study, BLEU (Bilingual Evaluation Understudy) [17] and BERTScore [18] were used to evaluate the quality of paraphrasing. BLEU is a widely used evaluation metric in Natural Language Processing (NLP), which calculates a score based on n-gram overlaps through simple mathematical computations [17]. However, it fails to account for semantic meaning and diversity in paraphrased texts [19]. In contrast, BERTScore leverages embedding vectors to compare the contextual meaning of words, making it well-suited for evaluating semantic similarity [18]. Table 3 provides two paraphrase examples for the 'Owner/User' class. The second paraphrased example in Table 3 scored 0.1163 for BLEU and 0.9162 for BERTScore. For each paraphrased text, both BLEU and BERTScores were calculated. These examples demonstrate the trade-off between semantic similarity and diversity, as reflected in the scores. The resulting paraphrased sentences are compared to its original reference data. The BLEU scores are relatively low due to the diverse rephrasing of the original text, the BERTScores remain high (above 0.8), indicating that the semantic meaning of the text is well-preserved. The average BLEU and BERTScore for 420 paraphrased texts using all the original dataset (excluding 'Crane Operator' class) scored 0.0689 and 0.9181 respectively. This is Calculated to see how diverse the paraphrased result is (lower BLEU) in terms of sentence structure and words choices but still have strong semantic similarity with high BERTScore.

Table 3. Example of two paraphrasing result of 'Owner/User' class with its BLEU, BERTScore

Natural Language Crane Accident Description ('Owner/User' class)	BLEU	BERTScore
Original: A crew was dispatched to an off shore platform to perform prep work for a move to another site. A pedestal crane on site was being used to lift various heavy mechanical equipment onto the deck of the platform. During one lift the load was raised and beginning to swing when the load line broke dropping the load and killing a worker.	-	-
Paraphrased 1: There was work done ahead of a move to the off shore platform. with a team. Heavy mechanical equipment was loaded onto a crane to be put onto adeck the platform. While it was a light Lifting at a construction project one lift had a load line break resulting in a dead worker when it swung.	1.9561E-78	0.8919
Paraphrased 2: A crew was sent to an off shore platform so they could help relocate some structures. It was a very cumbersome lift using a pedestal crane to get heavy mechanical equipment from one location to another. A worker was killed as the loaded load was cut back as it swung by a malfunctioning load line.	0.1163	0.9162

3 Methodology

3.1 Transformers

This study utilized Transformer models, particularly BERT (Bidirectional Encoder Representations from Transformers)-based models, due to its proven effectiveness in NLP tasks [14]. Three types of BERT-based models were used: BERT, RoBERTa (Robustly Optimized BERT Pretraining Approach), and DistilBERT (Distilled Bidirectional Encoder Representations from Transformers). The BERT model, developed by Google, was used as the baseline for

comparison due to its effectiveness in NLP tasks, leveraging bidirectional context to learn semantic meaning [14]. The RoBERTa model, developed by Facebook AI, is an enhanced version of BERT that was trained for a longer duration on larger datasets with improved pretraining techniques [20]. Meanwhile, DistilBERT is a lightweight model that is 40% smaller than BERT but generally achieves competitive performance with reduced computational costs [21].

Each model takes input_ids, which are tokenized text data mapped to integer sequences, along with supplementary information, such as attention_masks, as input. These inputs are passed through the encoder of the transformer model, which generates high-level vector embeddings containing contextual information about the input text using self-attention layers. The self-attention mechanism allows tokens to capture their contextual relationships within the text. For the classification task, a classifier is attached to the final layers of the models [22]. While the encoder component is pretrained, the classifier is initialized randomly. During training, the parameters of both the encoder and classifier are updated to fine-tune the model for the specific task.

3.2 Grid Search

Grid search was conducted to tune hyperparameters, including learning rate, batch size, and the number of epochs. Table 4 summarizes the combinations chosen for each model, for the following hyperparameter ranges explored: learning rate [1e-5, 2e-5, 3e-5], batch size [8, 16], and epochs [5, 10, 15].

3.3 Metrics

Precision, recall, and F1 scores were calculated as evaluation metrics for datasets 1, 2, and 3. F1 score, being the harmonic mean of precision and recall, offers a balanced measure of classification performance. The weighted F1 score is used for imbalanced datasets 4 and 5.

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (3)$$

$$Weighted\ F1\ Score = \sum_{i=1}^{N} w_i \times F1\ Score_i \quad (4)$$

$$where\ w_i = \frac{number\ of\ samples\ in\ class\ i}{total\ number\ of\ samples}$$

3.4 Input data augmentation: Paraphrasing technique

This study conducted two experiments to validate the effectiveness of paraphrasing in augmenting an imbalanced dataset under varying distribution conditions.

3.4.1 Experiment 1: Balanced distribution

For Experiment 1, Datasets 1,2 and 3 are used. Dataset 1 is the original imbalanced dataset containing total of 480 crane accident descriptions. Dataset 2 is augmented using paraphrasing; however, the 63 test is excluded in paraphrasing. Dataset 3 is augmented with all the data used for paraphrasing. Thus, after data augmentation using paraphrasing for dataset 2 and 3, is balanced with each containing 100 data per class.

3.4.2 Experiment 2: Original skewed distribution

For Experiment 2, Datasets 4 and 5 were used. Dataset 4 contains an imbalanced test set with 63 samples. The test set reflects the original dataset's imbalanced nature by using the stratify option in the train_test_split function to preserve the same class distribution as the original data. The remaining data was used for training and validation purposes. Dataset 5 also contains an imbalanced test set; however, the training and validation data were augmented. This augmentation was performed using the training and validation data from Dataset 4, excluding the test set, to create a larger and more diverse dataset for model training.

4 Result

Table 4 shows the performance of the prediction models on five input datasets for three different BERT-based models. For the original imbalanced data (Dataset 1) with a balanced test set, the best performance was achieved with DistilBERT, scoring 0.2898. With the addition of paraphrasing (Dataset 2), the performance improved across all three models, with RoBERTa showing the most significant improvement, increasing from 0.2544 to 0.3261. The results for Dataset 3 showed further improvement, with DistilBERT achieving the highest score of 0.5583. The confusion matrix in Figure 3 revealed that the model performed well in identifying major classes such as 'Crane Operator (class 0),' with high true positive rates, however failed with underrepresented classes (class 4-8). The confusion matrix for dataset 2 (Figure 4), shows improvement compared to dataset 1(Figure 3).

The precision of the base BERT model showed only a slight increase from 0.2672 to 0.2677 after data augmentation. However, the recall improved from

0.2857 to 0.3333, suggesting that the augmentation enriched the positive examples in this class, enabling the model to generalize better and reduce false positive predictions. In contrast, the DistilBERT model experienced a decline in precision from 0.3714 to 0.3089 when trained on the augmented dataset. This result indicates that lightweight models like DistilBERT may have a lower capacity to generalize effectively when exposed to more diverse data.

However, as indicated in Figure 5, there is evidence of overfitting, as lower-data-count classes were classified almost perfectly in the test set. This is likely because of a certain degree of data leakage, given that the training dataset includes paraphrased samples derived from the entire dataset, thus introducing contexts similar to those in the test data. Further experiments are needed to address these issues, as all models exhibited high performance (F1 scores above 0.6) during the validation process. Future research should explore alternative methodologies to enhance dataset quality by determining the most effective data augmentation strategies and their impact on model performance.

Table 4. Result of 5 datasets with each three BERT based transformer models

Data	Model	Precision	Recall	F1	Best Hyperparameters (learning rate, batch size, epochs)
1	BERT	0.2672	0.2857	0.2159	(2e-5, 8, 10)
	RoBERTa	0.2849	0.3175	0.2544	(3e-5, 16, 10)
	DistilBERT	0.3714	0.3492	0.2898	(2e-5, 8, 15)
2	BERT	0.2677	0.3333	0.2502	(2e-5, 8, 10)
	RoBERTa	0.3687	0.3651	**0.3261**	(3e-5, 16, 10)
	DistilBERT	0.3089	0.3651	0.2979	(3e-5, 8, 10)
3	BERT	0.5897	0.5556	0.5423	(2e-5, 8, 10)
	RoBERTa	0.6713	0.4603	0.4473	(2e-5, 8, 15)
	DistilBERT	0.5597	0.5873	**0.5583**	(3e-5, 8, 15)
Data	Model	Precision	Recall	Weighted F1	Best Hyperparameters (learning rate, batch size, epochs)
4	BERT	0.2810	0.2507	0.2445	(3e-5, 8, 5)
	RoBERTa	0.3439	0.2715	0.2804	(2e-5, 8, 15)
	DistilBERT	0.2856	0.2532	0.2484	(2e-5, 16, 10)
5	BERT	0.3726	0.2910	0.2974	(2e-5, 8, 15)
	RoBERTa	0.3183	0.2625	0.2816	(3e-5, 16, 15)
	DistilBERT	0.2761	0.2454	0.2500	(3e-5, 16, 10)

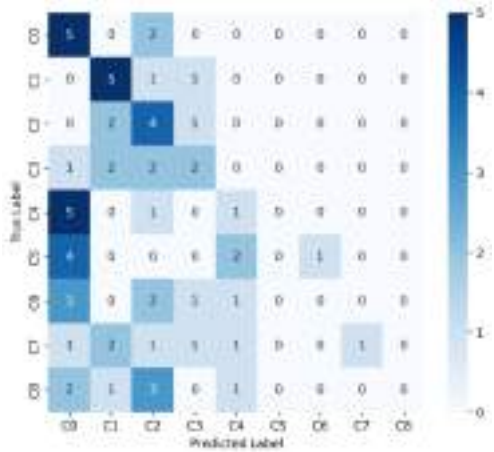


Figure 3. Confusion matrix for best test result using data 1 (DistilBERT)

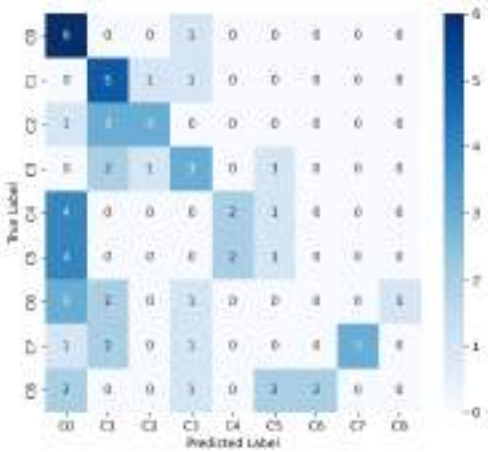


Figure 4. Confusion matrix for best test result using data2 (RoBERTa)

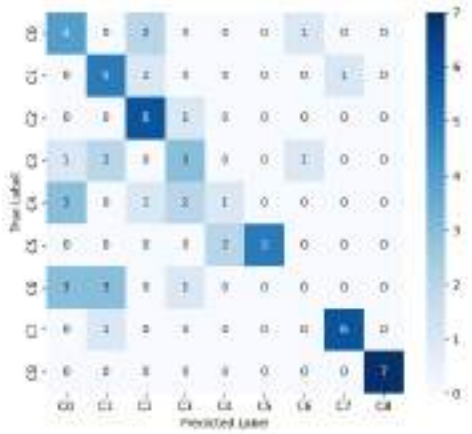


Figure 5. Confusion matrix for best test result using data 3 (DistilBERT)

5 Limitations and Future Work.

Different sentence-level augmentation techniques can be further investigated. Using large language models like Open AI API [23] for generating similar text could also be a potential research methodology. This study focused solely on identifying responsible personnel, but its scope can be expanded to explore other aspects of the crane accident dataset, such as predicting accident-related damages in terms of property and equipment and analyzing root causes of the accidents.

6 Conclusion

In this study, we addressed the challenge of class imbalance in a 9-class classification task for identifying responsible personnel in crane accidents by leveraging paraphrasing techniques for data augmentation. The results across Data 1, 2, and 3 demonstrate that increasing the dataset size through augmentation significantly improved performance. The F1 score of the best-performing model increased from 0.29 (Data 1) to 0.3261 (Data 2) and further to 0.5583 (Data 3).

Through our experiments, we evaluated the effectiveness of paraphrased data in improving the performance of transformer-based models, showcasing the potential utility of widely used paraphrasing tools.

However, this paraphrase-based data augmentation approach has certain limitations. The results may vary depending on the specific settings and methods used to generate the augmented dataset. For future studies, alternative methodologies, such as utilizing large language models for data augmentation, could be explored. Furthermore, experimenting with different hyperparameter settings may yield additional performance improvements.

References

[1] J. D. Wiethorn. An analytical study of critical factors of lift planning to improve crane safety based on forensic case studies of crane accidents. 2018. On-line: http://hdl.handle.net/2152/67528, Accessed: Nov. 13, 2024.

[2] Y. Ji and F. Leite. Automated tower crane planning: leveraging 4-dimensional BIM and rule-based checking. *Automation in Construction*, vol. 93, pp. 78–90, 2018.

[3] Y. Fang, Y. K. Cho, and J. Chen. A framework for real-time pro-active safety assistance for mobile crane lifting operations. *Automation in Construction*, vol. 72, pp. 367–379, Dec. 2016.

[4] V. Herrera-Pérez, F. Salguero-Caparrós, M. del C. Pardo-Ferreira, and J. C. Rubio-Romero. Key Factors in Crane-Related Occupational Accidents in the Spanish Construction Industry (2012 –2021). On-line: https://www.mdpi.com/1660-4601/20/22/7080, Accessed: Mar. 13, 2025.

[5] Y. Suh. Sectoral patterns of accident process for occupational safety using narrative texts of OSHA database. *Safety Science*, vol. 142, p. 105363, 2021.

[6] J. A. Taylor, A. V. Lacovara, G. S. Smith, R. Pandian, and M. Lehto. Near-miss narratives from the fire service: A Bayesian analysis. *Accident Analysis & Prevention*, vol. 62, pp. 119–129, 2014.

[7] J. M. Graves, J. M. Whitehill, B. E. Hagel, and F. P. Rivara. Making the most of injury surveillance data: Using narrative text to identify exposure information in case-control studies. *Injury*, vol. 46, no. 5, pp. 891–897, 2015.

[8] Z. Ma and Z.-S. Chen. Mining construction accident reports via unsupervised NLP and Accimap for systemic risk analysis. *Automation in Construction*, vol. 161, p. 105343, 2024.

[9] L. Kumi, J. Jeong, and J. Jeong. Data-driven automatic classification model for construction accident cases using natural language processing with hyperparameter tuning. *Automation in Construction*, vol. 164, p. 105458, 2024.

[10] J. Chen, D. Tam, C. Raffel, M. Bansal, and D. Yang. An Empirical Survey of Data Augmentation for Limited Data Learning in NLP. *Transactions of the Association for Computational Linguistics*, vol. 11, pp. 191–211, 2023.

[11] J. He, J. Gu, J. Shen, and M. Ranzato. Revisiting Self-Training for Neural Sequence Generation. *arXiv*, 2020. arXiv:1909.13788.

[12] J. Chen, Z. Yang, and D. Yang. MixText: Linguistically-Informed Interpolation of Hidden Space for Semi-Supervised Text Classification. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2147–2157, 2020.

[13] J. Zhang, Y. Zhao, M. Saleh, and P. J. Liu. PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization. *arXiv*, 2020. arXiv:1912.08777.

[14] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv*, 2019. arXiv:1810.04805.

[15] A. Holtzman, J. Buys, L. Du, M. Forbes, and Y. Choi. The Curious Case of Neural Text Degeneration. 2020, *arXiv*: arXiv:1904.09751.

[16] M. Peeperkorn, T. Kouwenhoven, D. Brown, and A. Jordanous. Is Temperature the Creativity Parameter of Large Language Models? *arXiv*, 2024. arXiv:2405.00492.

[17] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu. BLEU: a method for automatic evaluation of machine translation. *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pages 311-318, Philadelphia, Pennsylvania, 2001.

[18] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, and Y. Artzi. BERTScore: Evaluating Text Generation with BERT. *arXiv*, 2020. arXiv:1904.09675.

[19] R. Bawden, B. Zhang, L. Yankovskaya, A. Tättar, and M. Post. A Study in Improving BLEU Reference Coverage with Diverse Automatic Paraphrasing. *arXiv*, 2020. arXiv:2004.14989.

[20] Y. Liu *et al.* RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv*, 2019. arXiv:1907.11692.

[21] V. Sanh, L. Debut, J. Chaumond, and T. Wolf. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv*, 2020. arXiv:1910.01108.

[22] Hugging face. Auto Classes. On-line: https://huggingface.co/docs/transformers/en/model_doc/auto, Accessed: Jan. 15, 2025.

[23] OpenAI. OpenAI API. On-line: https://openai.com/index/openai-api/, Accessed: 03/13/2025.

Structuring Motion-Based Communication as a Language for Robotics: Analysis with Dense Optical Flow

Nicholas Albergo[1], Jeongbin Hwang [2], Doyun Lee [3], and Kevin Han [2]

[1]Department of Computer Science, North Carolina State University, North Carolina State University, Raleigh, NC, USA
[2]Department of Civil, Construction, & Environmental Engineering, North Carolina State University, Raleigh, NC, USA
[3]Department of Civil Engineering & Construction, Georgia Southern University, Statesboro, GA, USA

njalberg@ncsu.edu, jhwang25@ncsu.edu, doyunlee@georgiasouthern.edu, kevin_han@ncsu.edu

Abstract -

The construction industry is currently facing a labor shortage, prompting many researchers and practitioners to explore construction automation through advancements in robotics. For successful construction operations with multiple agents, effective and intuitive communication is essential. To this end, this research proposes a Motion-Based Communication (MBC) framework, which enables collaboration using movement. MBC provides a robust alternative that remains unaffected by various types of disruptions. This work contributes to the field of robotics by developing a motion-based language represented through an ontology. The ontological representation is encoded via directional shifts in motion rather than poses and is decoded by analyzing optical flow vectors. The experimental results in a lab environment using a UR5e robot indicate feasibility, with a shift inference accuracy of 98.8%. These findings can contribute to the development of a multi-agent MBC language to support communication in real time during construction operations.

Keywords -

Motion-Based Communication; Construction Automation via Robotics; Cooperating Robots; Dense Optical Flow

1 Introduction

Traditional robot-to-robot communication consists of physical data transfer via networking. However, in the event of a network failure or disruption, this modality of communication cannot be relied on. Additionally, networked communication efficacy is contingent on condition of the environment. For instance, in the case of large infrastructure construction projects such as dams, roads, and bridges, the installation of Internet networks (i.e., LTE and 5G) may be challenging, making seamless online communication difficult. Additionally, there are wireless networks that do not rely on Internet connections, such as Bluetooth [1] and Wi-Fi [2], which are relatively easy to install and can address some of these shortcomings. However, these networks are limited to specific areas and require manual connection each time they are used. As construction sites physically constantly change over time, wired networks must be reconfigured to consider the site conditions, which requires significant time and cost.

Motion-based communication (MBC) serves as a potential solution for non-networked communication for several reasons. First, it can potentially allow communication between human and computer as motion can be used as a common language channel that is understandable by humans [3]. In addition, compared to other multi-modal robot-to-human communication methods [4], displays and LED-based systems suffer from areas of high contrast or light intensity, while audio communication is evidently ineffective in vacuums or underwater. Lastly, motion is not a unique attribute and has the potential to facilitate a universal language channel in both organic and robotic agents that are capable of some form of actuation.

Therefore, the authors propose implementing MBC as a pseudo-language. This is done in two steps. First, an ontology is created by performing *open coding* [5] on several robot action research papers and corpora and ontologies [6, 7, 8] represented by OWL files (W3C Web Ontology Language [9]). The coding process results in an ontology representing a pseudo-language consisting of **action categories**, **actions types**, **entity types**, and **value types**, and **atomic types** (referred to collectively as **symbols**).

Afterwards, symbols in the ontology are encoded through alternating clockwise and counter-clockwise movement inside a communication window from a UR5e arm. A dense optical flow based motion analysis framework coupled with contour detection and tracking supplemented by Kalman filter is evaluated as the proposed methodology for decoding the motion-based symbols using a Logitech C270 camera with a resolution of 640 by 480. The efficacy of this method is compared to state-of-the-art (SOTA) MBC research. For brevity, this paper focuses on the actual communication framework and decoding rather than the ontology development. An overview of the steps taken can be seen in Figure 1.

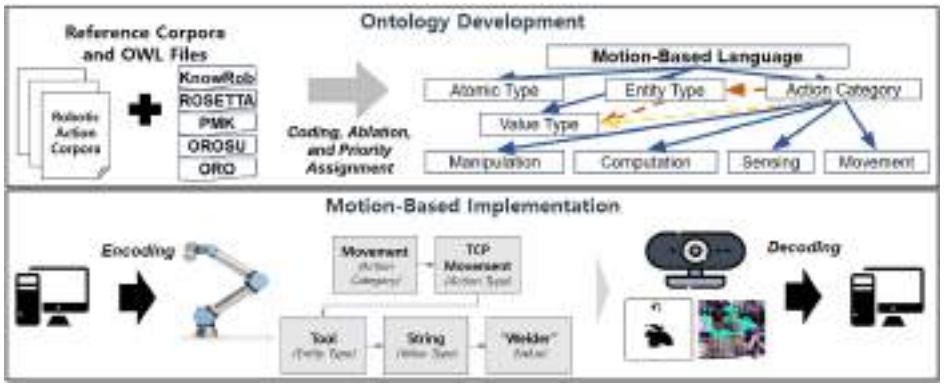


Figure 1. Research overview consisting of an open coding process to construct the language ontology (top) and the communication process with an example message (bottom).

2 Background

2.1 Motion-based communication

MBC deals with encoding a message between multiple agents via pose or motion and interpreting the message using computer vision techniques. These agents can consist of robots, humans, or some mix of the two. The capability to engage in multi-agent (i.e. both human and robot) communication with both human and robotic agents is a crucial benefit of MBC, along with the aforementioned resilience to networking impediments. Modern MBC research, whether dealing with human-to-robot, robot-to-robot, or multi-agent communication, utilizes either gestures [10, 11], poses [4, 3], trajectories [12], or a combination [13] of these to encode information.

2.1.1 Gesture-based

Gestures have been used both in the context of aligning common objects in multi-robot environments [10] as well as a common language channel between humans and robots. Vanc et al. [13] propose a pseudo-language for human-robot interaction (HRI) by implementing a framework using a combination of deictic, mimetic, and quantification gestures in addition to teleoperation. This setup requires Leap Motion Controller [14] and a gesture classification model [11] to recognize the encoded gesture-based signals real-time. Although this framework frames the MBC paradigm as a language, the usage of a model limits the language to communication from a human to a robot.

2.1.2 Pose-based

In the context of this research, the salient components of the pose of an agent are the roll, pitch, and yaw orientations of its body (or end effector, in the case of a robotic manipulator). Encoding from Koreitem et al. [15] is done using a set of ordered lists, called a *code*, where each element is a particular pose configuration (i.e., roll, pitch, yaw) in which the orientation space is binned. Their work focuses on minimizing a cost function across all codes. For each possible pose configuration, the cost is calculated by three factors: probability of occurrence, detection error of their orientation regressor, and time to execute. Decoding is performed via convolutional neural network (CNN) based pose estimation. As the main focus is the optimization of the encoded messages with respect to time, energy, and accuracy, there is less of an emphasis on flexibility. Using a pose estimation model limits generality, and each pose encodes a pre-built, holistic message.

2.1.3 Trajectory-based

Although using pose as a method of encoding messages is popular in modern MBC research, it can be supplemented with trajectory information as well. SOTA research from Fulton et al. [4] and Enan et al. [3] mainly use pose-based encodings, but some of their messages also contain trajectories. For instance, in [3], the message "GO TO LOCATION" is defined by rolling 45-degrees, moving forward and then backward twice, and then proceeding to the desired location. Their research is the first to create a unified HRI and robot-robot interaction (RRI) motion-based framework in which both the human operator and robot can interpret the gestures. In addition, it is the first MBC research to utilize self-attention [16] for decoding. User-defined gestures can also be defined to extend the language via configuration files.

However, as previously mentioned, MBC research has not taken full advantage of the potential generality of motion. That is, there is information contained in the motion itself rather than the pose that can be used as a sort of lexicon. In the example from [3], the individual words *go to* and *location* and their semantics are not encoded. There-

fore, adding another unique message relating to movement, for instance "GO TO AGENT," would require an entirely unique encoding, potentially even unrelated to the original one. Additionally, while the usage of deep-learning models enables potentially high precision, it comes at a high computational cost and can limit the ability to generalize the communication framework to other robots.

3 Methodology

The following sections briefly outline the methodology used in developing the baseline ontology used in this framework as well as covering the decoding process used to interpret messages.

3.1 Ontology coding process

A rough initial ontology is built based upon coding actions, categories, and frames from previous robotic action studies [8, 7, 6] to guide the coding process. The initial ontology consists of mainly action categories and action types with some initial entity types and values. Then, five OWL files (studied in [17]) are inspected using Protégé [18] and all leaves (a class without any subclasses) are extracted. Irrelevant classes, such as those not related to actions, objects, or values, are ablated. Each class is coded as either a type of action, entity, or generic value using the preliminary ontology as a reference.

From largest to smallest, the OWL files used in this research come from KnowRob [19, 20], ROSETTA [21], OROSU [22], ORO [23], and PMK [24]. These five OWL files are selected as they are among the 5 largest, open-source, industrial robotic action ontologies. For each of the 5 OWL files, only classes (i.e. subclasses of **owl:Thing**, not properties or individuals) relating to actions and objects were extracted and coded from the five OWL files.

This entire coding process is repeated for a total of five full runs, refining the codes and resulting in added granularity in the form of introducing action categories and specific value types. The final ontology for the proposed language consists of the following levels: **action categories**, **action types**, **entity types**, **value types**, and **atomic types**. Action categories are broad classifications of actions (*Manipulation*, *Movment*, etc.) while action types are specific varieties of those actions (*Handling*, *Rotation*, etc.). Entities are objects that can be observed or interacted with are children of action types that use them. Value types correspond to the possible arguments that can be used for specific messages. All symbols, aside from literal values, are assigned a priority with respect to other symbols in the same level of the tree based on their frequency in the surveyed literature.

3.2 Optical flow-based decoding

The communication process begins when one robot sends a *ToggleCommunications* symbol which consists of a unique cross pattern (discussed in Section 3.2.5). Then, the communicating robot begins sending a sequence of symbols followed by a final *ToggleCommunications* symbol which terminates communication. Motion detection and tracking for interpreting symbols (encoded as aforementioned **shifts**) is done using Gunnar-Farnebäck optical flow [25] coupled with contour finding [26]. A Kalman Filter [27] is also used to help estimate the center of the tracked objects.

In this language, symbols and literal values are encoded using a discrete quantity of **lexical motion units** (LMU). The magnitude of this motion (i.e. one LMU) is calculated dynamically when an agent initiates communication (see 3.2.5). The motion consists of a circular trajectory inside the communication window and alternates between counterclockwise and clockwise movements for each symbol. This ensures symbols are distinct and identifiable with a definitive **shift** in the direction of movement. In other words, the *direction* of the movement is independent of the symbol. Rather, the direction of motion flips after each symbol such that the overall message consists of a series of directional shifts along a fixed circle. Each symbol's encoding (aside from literal values) is equivalent to the calculated LMU value multiplied by its priority from the open coding process. Literal values have their own encoding rules using a binary representation, which is left out for brevity.

3.2.1 Farnebäck method

Farnebäck's method falls under the category of differential optical flow as it utilizes partial differential equations to model the flow field and minimize an energy function to perform motion estimation. It is chosen as the base motion detection component for several reasons. It does not rely on deep learning or training data and as such allows for better generalization and faster inference time compared to SOTA MBC decoding paradigms that implement pose-based CNNs [15, 4, 3]. This is a result of using a model-free method that only uses motion where there is no need to query any models. Lastly, the OpenCV [28] libraries provide the Farnebäck algorithm, permitting a simpler and more portable framework, which are attributes that are important in resource-limited environments that can often occur in construction.

Results from dense optical flow in environments with inconsistent lighting or a large quantity of background motion can be noisy. Flow vector magnitude thresholding, Gaussian blurring, and morphological operations (such as erosion and dilation) are all applied to yield a cleaner

image before processing continues.

3.2.2 Contour tracking

Dense optical flow yields both a magnitude $m(x, y)$ and direction $\theta(x, y)$ for each (x, y) pixel in the image. However, contours do not inherently contain optical flow information as they only consist of a set of points that define a boundary. Therefore, given a set of optical flow vectors **F** and contours **Q** inside an image **I**, the mapping from pixels in **Q** to **F** is described by:

$$\forall (x, y) \in \mathbf{Q},\ \exists\, (m(x, y), \theta(x, y)) = \mathbf{F}(x, y) \qquad (1)$$

Each contour is then assigned to an **Optical Flow Contour** object (OFC). This is done by comparing the central moment of a contour to each previously found OFC. If the center is within a tracking threshold, T_t (chosen as 22.5 pixels given the experimental setup and camera resolution), the OFC object is updated with the new information from the contour via a Kalman filter. If there are no current $OFCs$ or the current contour's center or area is beyond the thresholds, a new OFC object is created.

3.2.3 Optical flow contour objects

Each OFC contains the current magnitude $\bar{m}_i$, current angle $\bar{\theta}_i$, and center (c_i^x, c_i^y) of the contour it is assigned or updated with. To calculate $\bar{m}_i$ for the OFC, the mean of all *non-zero* magnitude pixels inside the assigned contour is found, filtering small noisy measurements. The angle $\bar{\theta}_i$ is computed by taking the circular mean of the angle vector, which also only includes pixels that are included in the mean magnitude calculation. Therefore, using the remaining n pixels inside the assigned contour, the magnitude and angle are given by:

$$\bar{m}_i = \frac{1}{n}\sum_{k=1}^{n} m_k;\ \bar{\theta}_i = \arg\ \frac{1}{n}\sum_{k=1}^{n} e^{\mathrm{i}\theta_k} \qquad (2)$$

Where i is the imaginary unit and $\arg(z)$ is the argument (angle) of the complex number z. Lastly, the center is found by taking the weighted center of pixel intensity (moment) from the convex hull of the contour. (c_i^x, c_i^y) and $\bar{m}_i$ are passed into the Kalman filter as measurements. This is used to compute the estimated state, or center of the OFC. In this way, the Kalman filter assists in the tracking of an OFC, but does not directly calculate its magnitude or angle. For each frame an arbitrary OFC is detected, $\bar{m}_i$ is added to that OFC's **shift magnitude**, ρ_i. The shift magnitude then contains the sum of all non-zero mean magnitudes, but is reset once a shift is detected for the current OFC. Motion - specifically the motion of the sender in between each shift - is then quantified by using ρ_i.

3.2.4 Shift detection

Shifts are considered changes in the angular component of the optical flow vector of $\frac{\pi}{2}$, also know as the angular threshold T_θ. The decoding process then consists of analyzing the motion segments in between each directional shift and estimating how many discrete LMUs make up each motion segment.

To determine a shift, $\Delta\bar{\theta}_i$ is acquired by finding the angular difference between $\bar{\theta}_i$ and $\bar{\theta}_{i-1}$. A maximum of three angular differences are stored in memory as a deque (double-ended queue) for each OFC to detect shifts; this is known as the **angle history**, Θ, of the OFC. A maximum deque for an arbitrary OFC is then given by:

$$\Theta = \{\Delta\bar{\theta}_{i-2}, \Delta\bar{\theta}_{i-1}, \Delta\bar{\theta}_i\} \qquad (3)$$

Each time an OFC object is matched with a viable contour in the current frame and updated, the angular difference is calculated and added to its angle history. The angles inside Θ are summed and compared to T_θ:

$$f(\Theta) = \sum_{\Delta\bar{\theta}\in\Theta} \Delta\bar{\theta} \qquad (4)$$

If $f(\Theta)$ exceeds T_θ, a shift is detected. $\bar{\theta}_i$, ρ_i, and (c_i^x, c_i^y) are then added to a dequeue known as the **shift history**, S, up to a maximum of seven shifts (the number of shifts required to detect the communication window). When a shift occurs, both Θ_i and the current shift magnitude ρ_i are cleared for the current OFC. Then ρ_i is used to determine the current symbol by rounding it to the nearest whole LMU. For example, given an LMU of 5.5 and final ρ_i of 10.2, the symbol is interpreted as 2 LMUs; the current direction of motion is irrelevant. Figure 2 shows shift detection between two frames.

3.2.5 ROI estimation for communication window

The symbol used to request and end communication, $ToggleCommunications$, serves two main roles. First, to signal the start and end of the motion-based communication chain. Secondly, it establishes the *communication window* and LMU estimation used by the receiver. The purpose of the communication window is also two-fold. It reduces noise from the environment by focusing on a specific region of interest (ROI) within the camera view and it reduces computation by disregarding pixels outside this ROI (see Section 4.1). This has the advantage of increasing frame rate while still avoiding the "aperture problem" caused by local measurements within a small window since the optical flow component continues to track the entire scene.

The sender and receiver are assumed to both be aware of the expected radius of the communication window, r_c,

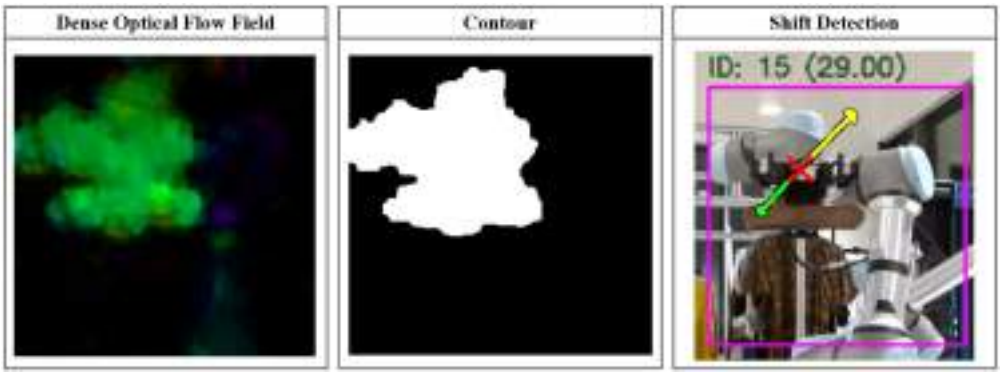


Figure 2. Optical flow field (left), contour (middle) and annotated frame (right) during a detected shift, marked with a red 'X' inside the ROI (purple box). Yellow and green arrows represent magnitude and direction ($\bar{m}$, $\bar{\theta}$) for $i-1$ and i, respectively. The ID of the OFC and current shift magnitude ρ_i are displayed above the ROI.

and an expected **radius ratio**, α, which is used to designate the ratio between ROI_r and r_c. In other words, the $ToggleCommunications$ symbol is meant to be sent within a smaller portion of the overall communicating window. This is used to help establish the ROI and LMU computation between sender and receiver, regardless of distance. The chosen r_c in this case is 10 cm, and α is selected to be 1/3. The $ToggleCommunications$ symbol consists of a series of movements in a cross pattern: $\{down, left, right, down, up, right, left\}$. For each OFC with a fully saturated shift history deque, the receiver actively processes these shifts until this sequence is found. It designates the $ToggleCommunications$ symbol, which sets the OFC that sent the symbol to *active*, meaning the receiver is now actively listening to (or more appropriately, watching for) encoded symbols from this OFC.

The ROI for the communication window is also established during this process. To find the window, the center of the points from the shifts are used to calculate both the center (origin) of the window (ROI_c), its radius (ROI_r), and its average magnitude (ROI_ρ). The center is found by taking the mean (x, y) coordinates and the radius by taking the average distance between each shift s in the $ToggleCommunications$ symbol for an arbitrary OFC:

$$ROI_c = \frac{1}{n}\sum_{k=1}^{n} s_k^{(c_x, c_y)} \tag{5}$$

$$ROI_r = \frac{1}{n-1}\sum_{k=1}^{n-1} \overline{(s_{k+1}^{c_x} - s_k^{c_x})^2 + (s_{k+1}^{c_y} - s_k^{c_y})^2} \tag{6}$$

The initial ROI_r is then multiplied by the reciprocal of the radius ratio $\frac{1}{\alpha}$ to yield the final expected ROI_r. An ROI is created around the origin ROI_c with radius ROI_r in which **only motion inside this window** from the currently **active** OFC is considered for symbol interpretation. Lastly, a weighted average shift magnitude of the radius is used in the computation of acquiring the expected magnitude for a singular lexical motion unit. At any time, the sending robot can signal $ToggleCommunications$ to remove the ROI and active status, thus allowing the receiver robot to resume actively observing for the start of a new $ToggleCommunications$ symbol.

4 Experimental results

To evaluate the overall accuracy of the system, experimental testing is performed on recordings of 13 distinct messages - one from each action type, with arbitrary entity and value parameters, and a message composed of several other messages. Each message is recorded from 1, 2, 3, 4, and 5 meters. Below is a list of example chosen messages along with their meaning:

M05 $Handling \rightarrow Manipulation \rightarrow Tool \rightarrow None \rightarrow None \rightarrow 0$: Pick up the nearest $Tool$.
M08 $Movement \rightarrow TCPMovement \rightarrow None \rightarrow Integer \rightarrow 20$: Move TCP forward 20 cm.
M09 $Movement \rightarrow TCPRotation \rightarrow None \rightarrow IntegerVector \rightarrow [90, 0, 0]$: Rotate TCP by 90 degrees in the x-axis.
M12 $Halt$: Stop all movement/actions and cancel all previous commands.
M13 A message consisting of 3 other messages: pick up the nearest tool, move the TCP to the nearest obstacle, and activate held tool

Results from decoding a message (**M13**) can be seen in Figure 3. Evaluation is performed by comparing the actual ρ_t^a (red dotted line) at each t where a shift occurs against the expected ρ_t^e (green line). Since LMUs are discrete units, ρ_t^a is rounded to the nearest LMU to get the final

decoded result, ρ_t^r (cyan dashed line). The orange dotted lines represent the LMU intervals, with the first (bottom) line equal to one LMU, the next line equal to two LMUs, and so on. The mean residual error is equal to the average difference at each shift ($|\rho_t^a - \rho_t^e|$)

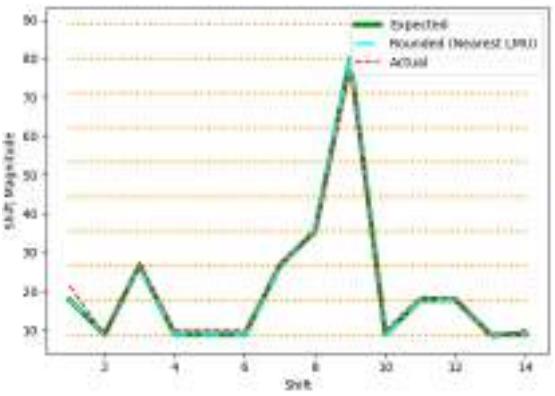


Figure 3. **M13** correctly interpreted at 2m with minimal mean residual error.

Across all 13 messages, there are a total of 167 symbols. Of the distances evaluated, the optimal range of interpreting shifts falls between 1 to 3 meters. From these distances, 98.8% (495 of 501) of symbols are correctly interpreted. Overall, a successful interpretation rate of 94.1% (786 of 835) is achieved from the 1 meter bins between 1 and 5 meters using a 640x480 resolution camera. The accuracy of each messages can be seen in Table 1. Each message along with the number of shifts required for its encoding is provided, along with the shift accuracy for both the optimal range (1, 2, and 3 meters) and all ranges (1, 2, 3, 4, and 5 meters). Figure 4 also shows the performance of messages from 1-3 meters, which is the weighted (0.95) accuracy minus the weighted (0.05) mean residual error that results from each shift.

4.1 Performance

Without downsampling, multi-threading, or hardware acceleration, the decoding algorithm runs at around 18-20 FPS on a AMD Ryzen 7 3700X 8-Core Processor. After the ROI is selected, the FPS reaches between 70 (for larger ROIs, in the case of close proximity to the communicating robot) and 330 FPS. As LMUs are dynamically calculated with respect to the relative motion of the robot, the algorithm is theoretically independent on speed assuming that the velocity of the robot remains relatively constant throughout the signaling.

Dense optical flow can struggle in environments with poor lighting or noisy backgrounds, such as flickering lights or numerous rapidly moving objects. The presence of heavy background noise or complex scene dynamics can be a limitation to the proposed dense optical flow approach for MBC, reducing both accuracy and performance. Although preprocessing and ROI construction help to mitigate noise under the evaluated conditions, the effects of environmental illuminance and highly disruptive noise on the proposed system are currently being studied.

Table 1. Results of the 13 messages evaluated from the optimal range (1 to 3 meters) and all ranges (1 to 5 meters) using a UR5e robotic arm. The values for shift inference accuracy represent the percentage of correctly decoded shifts using a 640x480 resolution camera at optimal range (left, 1-3m) and all evaluated ranges (right, 1-5m).

Message	Shifts	Shift Inference Accuracy (%)
M01	25	98.7/98.4
M02	36	100/94.4
M03	4	100/100
M04	22	100/100
M05	6	100/86.7
M06	18	96.3/93.3
M07	4	100/100
M08	10	96.7/88
M09	17	98.0/96.5
M10	5	100/100
M11	4	100/90
M12	2	100/100
M13	14	97.6/77.14

4.2 Comparison with SOTA

The SOTA in motion-based communication [3] achieves a recognition accuracy of 94.67% in simulation using their RRCommNet model, with an average inference speed of 10 FPS and 12.5 FPS (for RRCommNet-Skip, which averages 88% accuracy as it uses every other frame). Real-world accuracy for the two models is 83.33% and 73.81%, respectively for the regular and frame-skipping versions. By using optical flow vectors without machine learning, the MBC language runs faster and does not incur additional computational costs stemming from querying a model.

Since the language is fundamentally different than SOTA works, the comparison between shift inference accuracy and message inference accuracy is not necessarily one-to-one. However, the individual shift inference accuracy from this research outperforms SOTA real-world and simulation message inference accuracy at 1 through 5 meters using a 640x480 resolution camera without the usage of any type of model. The performance is also higher, with a minimum of about 18 FPS, reaching up to 330 FPS.

5 Conclusions

The authors have presented a novel framework for motion-based communication in which it is framed as a

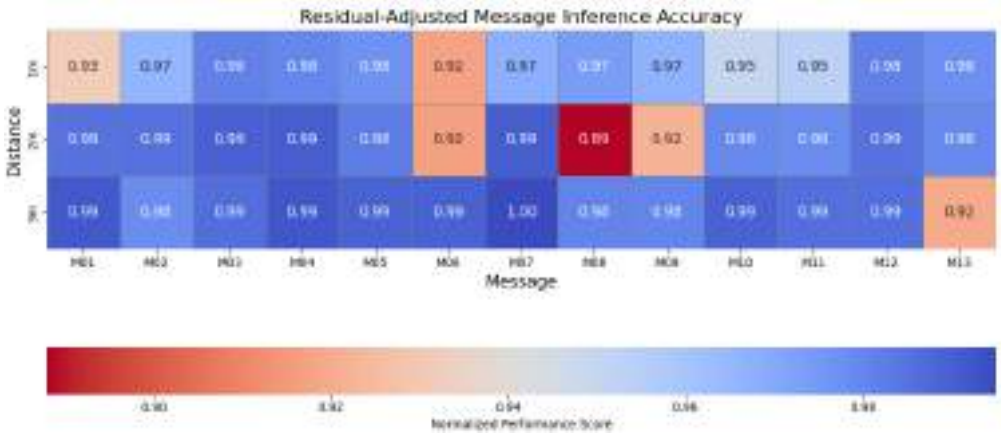


Figure 4. Accuracy across each message from 1-3 meters, accounting for average shift magnitude residuals. Lower values indicate higher mean residual error; orange/red indicates an incorrect shift detection.

language rather than discretely mapping poses to set messages. This is done by developing an ontological structure and encoding its nodes as units of motion along a circle. Experimental testing on a UR5e arm in a real-world lab environment demonstrates the feasibility of implementing this approach with robotic agents. An overall shift inference accuracy of 98.8% between 1 to 3 meters is attained with a 640x480 resolution Logitech C270. Due to the model-free approach, inference is done in real-time without the need for model computation overhead. Additionally, the modularity gained from framing the MBC paradigm as a natural language should allow for enhanced communication flexibility.

Evaluation of the proposed network-free framework in dynamic and complex construction sites where there may be elevated levels of network interference is planned as future work. Since the symbols are encoded and interpreted solely through motion, an additional area of future work is evaluation on other robotic and human agents to verify generality. Due to the extendability of the language and potential for decoding generality, we believe this MBC framework helps lay the groundwork for flexible and robust multi-agent communication going forward.

References

[1] Michael Abner, Peter Kok-Yiu Wong, and Jack C.P. Cheng. Battery lifespan enhancement strategies for edge computing-enabled wireless bluetooth mesh sensor network for structural health monitoring. *Automation in Construction*, 140:104355, 2022. ISSN 0926-5805. doi:https://doi.org/10.1016/j.autcon.2022.104355.

[2] Payam R. Zekavat, Sungkon Moon, and Leonhard E. Bernold. Performance of short and long range wireless communication technologies in construction. *Automation in Construction*, 47:50–61, 2014. ISSN 0926-5805. doi:https://doi.org/10.1016/j.autcon.2014.07.008.

[3] Sadman Sakib Enan, Michael Fulton, and Junaed Sattar. Robotic detection of a human-comprehensible gestural language for underwater multi-human-robot collaboration. In *IEEE International Conference on Intelligent Robots and Systems*, volume 2022-October, pages 3085–3092. Institute of Electrical and Electronics Engineers Inc., 2022. ISBN 9781665479271. doi:10.1109/IROS47612.2022.9981450.

[4] Michael Fulton, Chelsey Edge, and Junaed Sattar. Robot communication via motion: A study on modalities for robot-to-human communication in the field. *ACM Transactions on Human-Robot Interaction*, 11:1–40, June 2022. ISSN 25739522. doi:10.1145/3495245.

[5] John W. Creswell and Cheryl N. Poth. *Qualitative inquiry and research design: Choosing among five approaches*. SAGE Publications, 4 edition, February 2016.

[6] Matteo Pantano, Thomas Eiband, and Dongheui Lee. Capability-based frameworks for industrial robot skills: a survey. In *2022 IEEE 18th International Conference on Automation Science and Engineering (CASE)*, pages 2355–2362. IEEE Press, 2022. doi:10.1109/CASE49997.2022.9926648.

[7] Andrea Vanzo, Danilo Croce, Emanuele Bastianelli, Roberto Basili, and Daniele Nardi. Grounded language interpretation of robotic commands through structured learning. *Artificial Intelligence*, 278, January 2020. ISSN 00043702. doi:10.1016/j.artint.2019.103181.

[8] Irena Markievicz, Jurgita Kapočiūtė-Dzikienė, Minija Tamošiūnaitė, and Daiva Vitkutė-Adžgauskienė. Action classification in action ontology building using robot-specific texts. *Information Technology and Control*, 44:155–164, July 2015. ISSN 2335884X. doi:10.5755/j01.itc.44.2.7322.

[9] Michael Dean, Guus Schreiber, Sean Bechhofer, Frank Van Harmelen, Jim Hendler, Ian Horrocks, Deborah L. McGuinness, Peter F. Patel-Schneider, and Lynn A. Stein. Owl web ontology language overview, 2004. Available at: `https://www.w3.org/TR/owl-features/`.

[10] Polychronis Kondaxakis and Khurram Gulzar. Robot–robot gesturing for anchoring representations. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 107–114. IEEE, 2018. doi:10.1109/IROS.2018.8502843. URL `https://ieeexplore.ieee.org/document/8502843`.

[11] P. Vanc. Gesture teleoperation toolbox v.0.1, 2022. URL `https://github.com/imitrob/teleopgesturetoolbox/`.

[12] Haruki Nishimura and Mac Schwager. Active motion-based communication for robots with monocular vision. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2948–2955, 2018. ISBN 9781538630815. doi:10.1109/ICRA.2018.8463152.

[13] Petr Vanč, Jan Kristof Behrens, Karla Štěpánová, and Václav Hlaváč. Communicating human intent to a robotic companion by multi-type gesture sentences. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 9839–9845, 2023. doi:10.1109/IROS55552.2023.10341944. URL `https://doi.org/10.1109/IROS55552.2023.10341944`.

[14] F. Weichert, D. Bachmann, B. Rudak, and D. Fisseler. Analysis of the accuracy and robustness of the leap motion controller. *Sensors*, 13(55):6380–6393, May 2013.

[15] Karim Koreitem, Jimmy Li, Ian Karp, Travis Manderson, and Gregory Dudek. Underwater communication using full-body gestures and optimal variable-length prefix codes. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 8018–8025, 2019. ISBN 9781538660270. doi:10.1109/ICRA.2019.8793644.

[16] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2023.

[17] Alberto Olivares Alacros. A review and comparison of ontology-based approaches to robot autonomy. *The Knowledge Engineering Review*, 34, 2019. doi:10.1017/S0269888920000090.

[18] Mark A Musen. The protégé project: a look back and a look forward. *AI Matters*, 1:4–12, June 2015. doi:10.1145/2757001.2757003.

[19] Moritz Tenorth and Michael Beetz. Knowrob — knowledge processing for autonomous personal robots. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4261–4266, 2009. ISBN 9781424438044. doi:10.1109/IROS.2009.5354602.

[20] Moritz Tenorth and Michael Beetz. Knowrob: A knowledge processing infrastructure for cognition-enabled robots. *International Journal of Robotics Research*, 32:566–590, April 2013. ISSN 02783649. doi:10.1177/0278364913481635.

[21] Maj Stenmark and Jacek Malec. Knowledge-based industrial robotics. In *Twelfth Scandinavian Conference on Artificial Intelligence*, pages 265–274. IOS Press, 2013. doi:10.3233/978-1-61499-330-8-265.

[22] Paulo J.S. Goncalves and Pedro M.B. Torres. Knowledge representation applied to robotic orthopedic surgery. *Robotics and Computer-Integrated Manufacturing*, 33:90–99, 2015. ISSN 07365845. doi:10.1016/j.rcim.2014.08.014.

[23] Séverin Lemaignan, Raquel Ros, Lorenz Mösenlechner, Rachid Alami, and Michael Beetz. Oro, a knowledge management platform for cognitive architectures in robotics. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3548–3553, 2010. ISBN 9781424466764. doi:10.1109/IROS.2010.5649547.

[24] Mohammed Diab, Aliakbar Akbari, Muhayy Ud Din, and Jan Rosell. Pmk—a knowledge processing framework for autonomous robotics perception and manipulation. *Sensors*, 19, March 2019. ISSN 14248220. doi:10.3390/s19051166.

[25] Gunnar Farnebäck. Two-frame motion estimation based on polynomial expansion. In Josef Bigun and Tomas Gustavsson, editors, *Image Analysis*, pages 363–370. Springer Berlin Heidelberg, 2003. URL `http://www.isy.liu.se/cvl/`.

[26] Satoshi Suzuki and KeiichiA be. Topological structural analysis of digitized binary images by border following. *Computer Vision, Graphics, and Image Processing*, 30:32–46, 1985. ISSN 0734-189X. doi:https://doi.org/10.1016/0734-189X(85)90016-7.

[27] R E Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82:35–45, March 1960. ISSN 0021-9223. doi:10.1115/1.3662552.

[28] Gary Bradski and Adrian Kaehler. Opencv, 2000. Accessed: 2024-07-16.

A Validation Study of the Four-Pick Point Design Method of a Mobile Crane Rigging Assembly with Over Six Pick Points

Ramy Kheir[1], Sang Hyeok Han[1], Ashutosh Bagchi[1]
[1]Department of Building, Civil and Environmental Engineering, Concordia University, Canada
E-mail: Ramy.kheir@mail.concordia.ca, sanghyeok.han@concordia.ca, ashutosh.bagchi@concordia.ca

Abstract

In the field of modular construction, mobile cranes are an essential tool for lifting large and heavy modules vertically. These modules often need to be lifted without rotation, meaning they must be kept in a flat position during the lifting process to avoid potential structural damage. One of the most important aspects of the lifting process is the rigging design, which consists of components like slings, shackles, and spreader bars. These elements play a vital role in distributing the load during the lifting operation. While considerable progress has been made to automate the rigging design process, the existing methods are often limited in their ability to handle larger lifting configurations involving more than six lifting points. This study addresses these limitations by presenting an in-depth structural analysis verification of the current rigging design method that can be applied to modules with more than six lifting points. The paper provides a detailed step-by-step calculation procedure for all the required components necessary for the design and selection of rigging components. The results presented in this paper represent an important step forward in the development of more advanced methodologies for lifting larger modules with complex rigging requirements, ensuring safety and efficiency.

Keywords –

Rigging assembly; Mobile crane; Pick-points

1 Introduction

In most present heavy lift projects, modularization is adopted as a fast and efficient process where modules are required to be transported to the job sites, lifted from their delivered locations, and set in their final positions. Mobile cranes, set up on the job site, are considered the centerpiece equipment during the on-site module installation phase. Disruptions occurring during this phase can easily take away the benefits of adopting modular construction.

Generally, the modules can weigh up to 160 tons and can come in various sizes of up to 24 ft (7.3 m) in width, 138 ft (42 m) in length, and 27.5 ft (8.4 m) in depth [1]. Modules are categorized into three main types: cable trays, pipe racks, and building modules. These modules may have up to five levels and up to 16 lifting points. The lifted modules should remain in a level position to avoid deflection in the module and/or uneven distribution of the carried load. Rigging failure can result in safety risks, cost overruns, and schedule delays. Cho et al. [2]; MinayHashemi et al. [3] reported that 28.1 % of fatal accidents associated with cranes in the United States were due to the failure of slings, wire ropes, and hoist lines between 2002 and 2012. As rigging assemblies connect the crane's hook to the modules, their slinging arrangement plays a key role in ensuring proper load distribution from the lifting lugs up to the crane's hook. For this reason, a sustainable design method for rigging assemblies with different configurations and sizes is crucial for a safe and efficient module lifting process.

Previously, the common slinging arrangement consisted of a 4-point pick configuration, where four diagonal slings were directly attached between the crane's hook and the pick points. In rare cases when the center of gravity (COG) coincides with the center of mass (COM), an identical length of the four slings transferring the loads from the pick points of the module to the crane's hook was used. In practice, however, sling length adjustments are often required to ensure balanced loads and keep the module horizontal during lifting [3]. Available standards, such as the American Petroleum Institute's API RP2A guidelines [4] and the Det Norske Veritas (DnV) design criteria

[5], can be used for the analysis and design of such rigging assembly configurations. While the former computed the sling loads in a suspended equilibrium condition with a single lump sum safety factor of 4 accounting implicitly for dynamic effects; the latter recommended a load sharing between diagonal slings in a 4-point configuration of 2/3 to 1/3 ratios with a 1.5 Dynamic Amplification Factor (DAF) [6]. Later, Sam [7] developed a spreadsheet solution to find precise load distributions between the four slings, accounting for COG position and variations in sling lengths. [8] performed a non-linear analysis in case when sling lengths tolerances were exceeded due to misfit for computing sling load distributions. Despite the development of these tools, many rigging engineers continue to rely on the trial-and-error method based on their experience and guesswork. While a few commercial design software solutions have been developed to automate the design process, they are not widely adopted due to economic, practical, and technical limitations [3].

A traditional rigging assembly is shown in Figure 1. In these arrangements, triangular slings transfer the load from the lifting points (pick points) to the crane's hook through a series of spreader bars, slings, and shackles found at different levels. The number of lifting points on the module dictates the number of levels in a traditional rigging assembly. In an advanced lift scenario (with 16 points), the total number of inventory components including spreader bars, slings, and shackles can exceed 270. The total rigging configuration height can range between 75 to 165 feet, requiring five assembly levels [3], [9], [10].

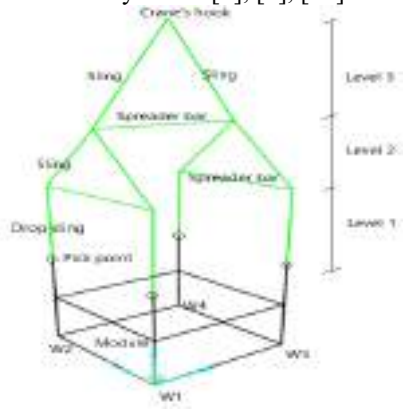


Figure 1. 3D view of a 4-point traditional rigging

Despite the research development presented in MinayHashemi et al. [3] and Hashemi et al. [9], which limited the design process to 6-pick point rigging assemblies, further verification is still required for the current methods used in designing larger rigging configurations with more than six lifting points. This study aims to address that gap and present a more comprehensive verification for the design of rigging systems used in lifting larger modules with up to 16 lifting points.

2 Methodology

2.1 Limitations of the Existing Method (Static Equilibrium)

In 2020, MinayHashemi et al. [3] proposed an automated design framework for analyzing and selecting rigging components in systems with up to four lifting points. This method, while efficient and beneficial for smaller modules, faces significant limitations when applied to larger configurations. The core of this methodology relies on the concept of static equilibrium, which assumes that all forces in the system balance out. This concept works well for smaller systems with simple geometries, but as the number of lifting points increases, the complexity of balancing these forces grows significantly.

2.1.1 First and second levels of 4-pick point configuration

For a system with a 4-point configuration (shown in Figures 1 and 2), the calculations are relatively straightforward. The weight of the module is distributed equally across all lifting points, and the design of the rigging components is based on this assumption. However, when the COG of the module is offset or when the configuration involves more than four lifting points, the situation becomes more complicated. For instance, if a module's COG is not centered, the slings may need to be adjusted in length to balance the forces applied at each lifting point, which may lead to uneven load distribution or the risk of tipping during the lift. For this scenario, the study [3] derived 4 equations (1), (2), (3), and (4) to analyze left sling forces at the first and second levels (FL_1, and FL_2) and the right sling forces at the first and second levels (FR_1 and FR_2) as shown in Figure 2.

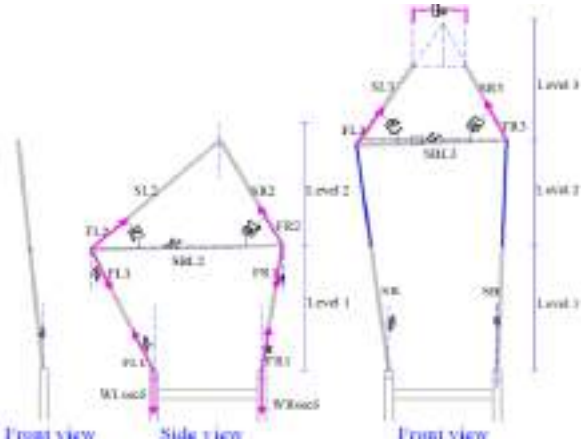


Figure 2. Front and side views of a 4-pick point rigging arrangement with sling internal forces

$$FL_1 = (WL + WR)\sec\delta\csc(\alpha+\beta)\,sin\alpha \quad (1)$$
$$FR_1 = (WL + WR)\sec\delta\csc(\alpha+\beta)\,sin\beta \quad (2)$$
$$FL_2 = (WL + WR)\sec\delta\cos(\theta R-\gamma)\csc(\theta L+\theta R) \quad (3)$$
$$FR_2 = (WL + WR)\sec\delta\cos(\theta L+\gamma)\csc(\theta L+\theta R) \quad (4)$$

For all the equations in this study, the letters *L* and *R* represent respectively left and right sides of the assembly. *WL* and *WR* are respectively the left and right weight load forces applied at the 2 pick points in the plane containing the drop slings and the spreader bars in the global Y-direction. α and β are the respective right and left drop sling angles with the vertical. θR and θL are respectively the right and left sling angles at the second level. δ is the angle between the plane containing the spreader bar and the 2 above slings in the Y-direction at either the left or the right side of the assembly with the vertical plane. γ is the slope of the spreader bar at the second level in either side plane.

2.1.2 Third level of 4-pick point configuration

Following the sling force computations at the first and second levels, the analysis of the third-level components was conducted in a 2D plane (XZ) containing the spreader bar and the two diagonal slings acting above it using the same method [3] . However, these computations were not shown in that same paper. Here, a detailed analysis of the two sling forces using the free-body diagram of the spreader bar at the third level was performed. Figure 3 presents the free-body diagram of the spreader bar at level 3. Equilibrium equations in both directions were applied to find the required capacity of the two slings and the mounted shackles at the third level. The results are presented in Equations (5) and (6).

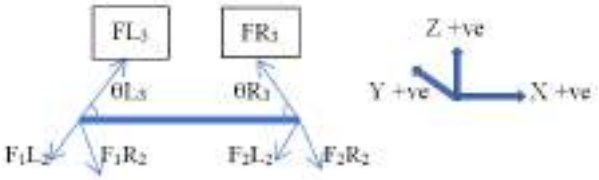


Figure 3. Free body diagram of spreader bar at level 3 with applied forces

$$FL_3\cos(\theta L_3+\gamma_3) - FR_3\cos(\theta R_3-\gamma_3) + \sin\delta_{12}[F_1R_2\sin(\theta_1R_2-\gamma_{12}) + F_1L_2\sin(\theta_1L_2+\gamma_{12})] - \sin\delta_{22}[F_2R_2\sin(\theta_2R_2-\gamma_{22}) + F_2L_2\sin(\theta_2L_2+\gamma_{22})] = 0 \quad (5)$$

$$FL_3\sin(\theta L_3+\gamma_3) + FR_3\sin(\theta R_3-\gamma_3) - \cos\delta_{12}[F_1R_2\sin(\theta_1R_2-\gamma_{12}) + F_1L_2\sin(\theta_1L_2+\gamma_{12})] - \cos\delta_{22}[F_2R_2\sin(\theta_2R_2-\gamma_{22}) + F_2L_2\sin(\theta_2L_2+\gamma_{22})] = 0 \quad (6)$$

Where θL_3 and θR_3 are the sling angles with the spreader bar SBL_3 at level 3; θ_1R_2 and θ_1L_2 are respectively the right and left sling angles at the second level in the left side plane; θ_2R_2 and θ_2L_2 are respectively the right and left sling angles at the second level in the right-side plane. The left and right-side planes are supported respectively by the left/right ends of the spreader bar at the 3rd level; δ_{12} and δ_{22} are the angles of the 2 side planes with the vertical plane. γ_{12}/ γ_{22} and γ_3 are the slopes of the spreader bars at the second and third levels respectively. FL_3 and FR_3 are the left and right sling forces at the third level; F_1R_2/F_1L_2 and F_2R_2/F_2L_2 are respectively the right/left sling forces at the second level in respectively the left and right-side planes.

In Equation (5), the two terms $\sin\delta_{12}[F_1R_2\sin(\theta_1R_2-\gamma_{12}) + F_1L_2\sin(\theta_1L_2+\gamma_{12})]$ and $\sin\delta_{22}[F_2R_2\sin(\theta_2R_2-\gamma_{22}) + F_2L_2\sin(\theta_2L_2+\gamma_{22})]$ represent the horizontal projections of the diagonal sling forces at the second level acting on the horizontal axis at the third level. These components are very small because the angles δ_{12} and δ_{22} are nearly zero, making their sines almost negligible. Additionally, these forces act in opposite directions on the spreader bar at level 3. Since keeping them in Equation (5) would complicate the calculation of FL_3 and FR_3 without significantly changing their final values, they were removed from the equation. So, Equation (5) will be expressed as (5').

$$FL_3 = FR_3\cos(\theta R_3-\gamma_3)/\cos(\theta L_3+\gamma_3) \quad (5')$$

By substituting FL_3 in Equation (6) to solve for FR_3 then solving for FL_3 in Equation (5'), Equations (7) and (8) will be obtained.

$$FR_3 = \frac{\{cos\delta_{12}*[F_1R_2*sin(\theta_1R_2-\gamma_{12})+F_1L_2*sin(\theta_1L_2+\gamma_{12}]+cos\delta_{22}*[F_2R_2*sin(\theta_2R_2-\gamma_{22})+F_2L_2*sin(\theta_2L_2+\gamma_{22})]\}*cos(\theta L3+\gamma_3)}{cos(\theta R_3-\gamma_3)+cos(\theta L_3+\gamma_3)*sin(\theta R_3-\gamma_3)} \quad (7)$$

$$FL_3 = \frac{\{cos\delta_{12}*[F_1R_2*sin(\theta_1R_2-\gamma_{12})+F_1L_2*sin(\theta_1L_2+\gamma_{12})]+cos\delta_{22}*[F_2R_2*sin(\theta_2R_2-\gamma_{22})+F_2L_2*sin(\theta_2L_2+\gamma_{22})]\}}{1+cos(\theta L_3+\gamma_3)*tan(\theta R_3-\gamma_3)} \quad (8)$$

In more complex systems with more than six lifting points, the traditional static equilibrium model becomes insufficient. The complexity of these systems, which involve multiple sling lengths, shackles, and spreader bars, requires more advanced methods to ensure that all forces are balanced and that the module remains stable throughout the lift. While static equilibrium analysis can still be applied to smaller systems, it becomes less reliable as the number of lifting points increases, particularly when the rigging system is subjected to dynamic forces ignored within the current study.

2.2 Designing Rigging Assemblies for Larger Configurations

For larger modular lifting configurations, such as those involving 8 or 16 lifting points, the traditional approach can still be extended by breaking down the larger system into smaller, more manageable sub-systems. However, with those employing a total of 12 or 14 lifting points composed of respectively a pair of 6-pick point segments or a pair of 6-pick point segments and a 2-pick point segment, the traditional model wouldn't be computationally accessible. In other words, it would fail to analyze the sling forces in the 12 or 14-pick point segments. Lastly, for the 10-pick point configuration composed of an 8-pick point and a 2-pick point segments, the method could still analyze the sling forces in the former segment. However, based on the design methodology adopted in MinayHashemi et al. [3], this step would be insufficient for selecting the 2-pick point segment components.

2.2.1 8-pick point configuration

An 8-point configuration can be divided into two 4-point rigging assemblies connected by a spreader bar at a higher lifting level. Each 4-point system is analyzed separately, and the total load is transferred to the spreader bar at level 4 (Figure 4).

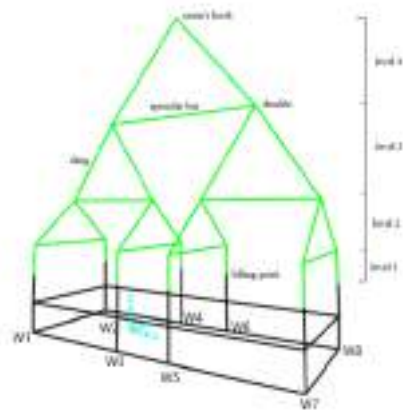


Figure 4. 3D schematic of an 8-pick point rigging assembly

The left segment (W_1 to W_4) and the right segment (W_5 to W_8) each transfer their respective forces to the spreader bar, which then distributes the total load up to the crane's hook. Using the same methodology and equations from the previous section, Equations (9), (10), (11), and (12) can be derived to calculate the forces in respectively the left- and right-side assembly. Equilibrium equations applied to the free-body diagram of the spreader bar at level 4 shown in Figure 5 will yield the sling forces F_{R4} and F_{L4} expressed in Equations (13) and (14) respectively.

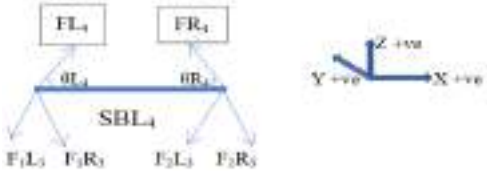


Figure 5. Free body diagram of spreader bar at level 4

$$F_1R_3 = \frac{\{cos\delta_{12}*[F_1R_2*sin(\theta_1R_2-\gamma_{12})+F_1L_2*sin(\theta_1L_2+\gamma_{12})]+cos\delta_{22}*[F_2R_2*sin(\theta_2R_2-\gamma_{22})+F_2L_2*sin(\theta_2L_2+\gamma_{22})]\}*cos(\theta_1L_3+\gamma_{13})}{cos(\theta_1R_3-\gamma_{13})+cos(\theta_1L_3+\gamma_{13})*sin(\theta_1R_3-\gamma_{13})} \quad (9)$$

$$F_1L_3 = \frac{\{cos\delta_{12}*[F_1R_2*sin(\theta_1R_2-\gamma_{12})+F_1L_2*sin(\theta_1L_2+\gamma_{12})]+cos\delta_{22}*[F_2R_2*sin(\theta_2R_2-\gamma_{22})+F_2L_2*sin(\theta_2L_2+\gamma_{22})]\}}{1+cos(\theta_1L_3+\gamma_{13})*tan(\theta_1R_3-\gamma_{13})} \quad (10)$$

$$F_2R_3 = \frac{\{cos\delta_{32}*[F_3R_2*sin(\theta_3R_2-\gamma_{32})+F_3L_2*sin(\theta_3L_2+\gamma_{32})]+cos\delta_{42}*[F_4R_2*sin(\theta_4R_2-\gamma_{42})+F_4L_2*sin(\theta_4L_2+\gamma_{42})]\}*cos(\theta_2L_3+\gamma_{23})}{cos(\theta_2R_3-\gamma_{23})+cos(\theta_2L_3+\gamma_{23})*sin(\theta_2R_3-\gamma_{23})} \quad (11)$$

$$F_2L_3 = \frac{\{cos\delta_{32}*[F_3R_2*sin(\theta_3R_2-\gamma_{32})+F_3L_2*sin(\theta_3L_2+\gamma_{32})]+cos\delta_{42}*[F_4R_2*sin(\theta_4R_2-\gamma_{42})+F_4L_2*sin(\theta_4L_2+\gamma_{42})]\}}{1+cos(\theta_2L_3+\gamma_{23})*tan(\theta_2R_3-\gamma_{23})} \quad (12)$$

$$FR_4 = \{(cos\delta_{12}\ [F_1R_2sin\ (\theta_1R_2 - \gamma_{12}) + F_1L_2sin(\theta_1L_2 + \gamma_{12})] + cos\delta_{22}\left[F_2R_2sin(\theta_2R_2 - \gamma_{22}) + F_2L_2sin(\theta_2L_2 + \gamma_{22})\right]\left(\frac{sin(\theta_1R_3+\theta_1L_3)}{[1+cos(\theta_1L_3+\gamma_{13})sin(\theta_1R_3-\gamma_{13})]}\right) + (cos\delta_{32}\ [F_3R_2sin\ (\theta_3R_2 - \gamma_{32}) + F_3L_2sin(\theta_3L_2 + \gamma_{32})\right] + cos\delta_{42}[F_4R_2sin(\theta_4R_2 - \gamma_{42}) + F_4L_2sin(\theta_4L_2 + \gamma_{42})])\left(\frac{sin(\theta_2R_3+\theta_2L_3)}{[1+cos(\theta_2L_3+\gamma_{23})tan(\theta_2R_3-\gamma_{23})]\,cos(\theta_2R_3-\gamma_{23})}\right)\}\frac{cos(\theta L_4+\gamma_4)}{sin(\theta L_4+\theta R_4)} \quad (13)$$

$$FL_4 = \{(cos\delta_{12}\ [F_1R_2sin\ (\theta_1R_2 - \gamma_{12}) + F_1L_2sin(\theta_1L_2 + \gamma_{12})] + cos\delta_{22}\left[F_2R_2sin(\theta_2R_2 - \gamma_{22}) + F_2L_2sin(\theta_2L_2 + \gamma_{22})\right]\left(\frac{sin(\theta_1R_3+\theta_1L_3)}{[1+cos(\theta_1L_3+\gamma_{13})sin(\theta_1R_3-\gamma_{13})]}\right) + (cos\delta_{32}\ [F_3R_2sin\ (\theta_3R_2 - \gamma_{32}) + F_3L_2sin(\theta_3L_2 + \gamma_{32})\right] + cos\delta_{42}[F_4R_2sin(\theta_4R_2 - \gamma_{42}) + F_4L_2sin(\theta_4L_2 + \gamma_{42})])\left(\frac{sin(\theta_2R_3+\theta_2L_3)}{[1+cos(\theta_2L_3+\gamma_{23})tan(\theta_2R_3-\gamma_{23})]\,cos(\theta_2R_3-\gamma_{23})}\right)\}\frac{cos(\theta R_4-\gamma_4)}{sin(\theta L_4-\theta R_4+2\gamma_4)} \quad (14)$$

2.2.2 10-pick point configuration

A 10-point system can be analyzed by combining an 8-point assembly with a 2-point segment (Figure 6).

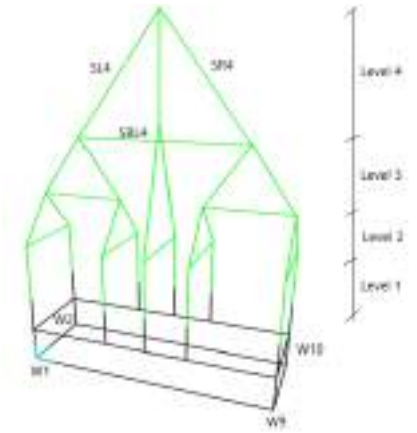


Figure 6. 3D schematic of a 10-pick point rigging assembly

The 2-point system, located in the center of the rigging setup, handles a smaller portion of the load. The analysis of the 8-point system is carried out as described in the previous section, but the 2-point system, which has a different structural design, requires additional consideration. Unlike the 8-point system, the 2-point segment is less complex, consisting of fewer lifting points and levels. The challenge lies in ensuring that the 2-point system, which has fewer lifting points, does not lead to uneven distribution of forces when combined with the larger 8-point system. Based on that, its structural design was conducted following that of the 8-point segment on a level-to-level basis.

Previous work by MinayHashemi et al. [3] did not include an in-depth structural analysis of the 2-point segment and assumed identical sling lengths across both the 2-point and 4-point segments in a 6-pick point configuration. However, this assumption becomes problematic when the 2-point segment differs in terms of the number of levels compared to the 8-point segment. For accurate rigging design, it is essential to carefully analyze the 2-point configuration and adjust for the different levels in the system.

2.2.3 12-pick point configuration

A 12-point system consists of two 6-point assemblies; each designed using the principles applied in the 4-point configuration. The total weight of the module is transferred to a spreader bar located at the fourth level, which ensures an even distribution of forces across all the lifting points (Figure 7).

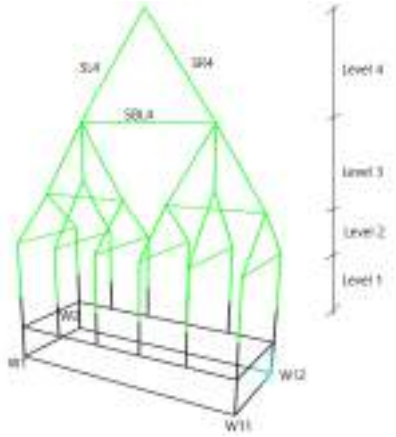


Figure 7. 3D schematic of a 12-pick point rigging assembly

Each 6-point assembly is analyzed independently, with the forces from the left and right segments being transferred to the spreader bar. In this case, when analyzing the forces acting on the spreader bar at level 4 in the XZ plane, four unknown force components need to be determined for each end of the spreader bar as shown in Figure 8. FL_4 and FR_4 are respectively the left and right sling forces at the fourth level of the rigging assembly; θL_4 and θR_4 are respectively the left and right sling angles with the spreader bar SBL_4 at the

fourth level; F_1M_3 and F_2M_3 are respectively the left- and right-middle sling forces at the third level supported by the left and right spreader bar ends at level 4 of the rigging assembly. By applying equilibrium, Equations (15) and (16) are obtained.

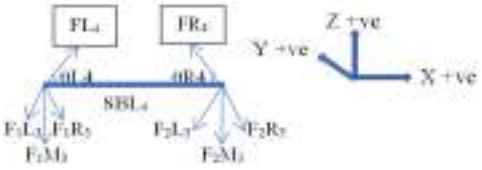


Figure 8. Free body diagram of spreader bar at level 4 for a 12-pick point configuration

$$FL_{4X} - FR_{4X} - F_1L_{3x} + F_1R_{3x} - F_2L_{3x} + F_2R_{3x} = 0 \quad (15)$$

$$FL_{4Z} + FR_{4Z} - F_1L_{3z} - F_1R_{3z} - F_2L_{3z} - F_2R_{3z} - F_1M_3 - F_2M_3 = 0 \quad (16)$$

F_1L_3 and F_1R_3 are the sling forces on the left side, and F_2L_3 and F_2R_3 are the sling forces on the right side of the spreader bar SBL_4 were already computed in the 8-pick point rigging assembly analysis section in this paper. F_1M_3 and F_2M_3 are the 2 additional forces applying vertically downwards due to the presence of the 2-pick point segments. This will lead to a system of 2 variables (FL_{4x} and FR_{4x}) in the X direction and 4 variables (FL_{4z}, FR_{4z}, F_1M_3, and F_2M_3) in the Z direction. Having only 2 equations to solve such a system will jeopardize reaching a solution.

While static equilibrium can be used to analyze the 6-point assemblies, the complexity increases at the fourth level, where the spreader bar introduces additional unknowns. These unknowns can't be easily solved using static equilibrium methods alone, and therefore, finite element analysis (FEA) could be employed to accurately model the forces in the system. The additional level of complexity requires a more robust approach to ensure stability and accurate force distribution calculations.

2.2.4 14 pick-point segments

A 14-point system includes both a 12-point assembly and a 2-point assembly (Figure 9). The design for the 12-point segment follows the same principles outlined above, but the 2-point segment introduces a new challenge. The 2-point system, located in the middle of the rigging assembly, consists of fewer lifting points and levels, which complicates the design process. MinayHashemi et al. [3] assumed that the rigging configuration of the 2-point segment matched the number of levels in the 4-point system, but this assumption is no longer valid in this case.

To design the 2-pick point system, a more advanced method involving its analysis should be applied. A new method for selecting rigging components is necessary. This method should account for the differing number of levels in the system compared to the 12-point segment and ensure that the forces are balanced and stable. The 2-point system must be analyzed independently of the 12-point segment and integrated into the overall rigging design.

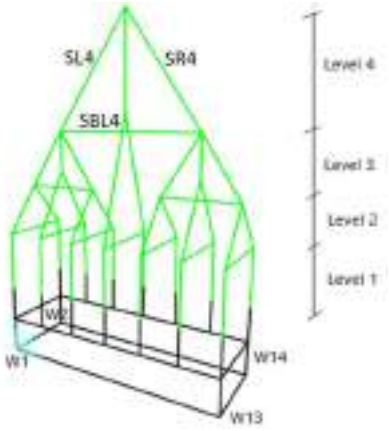


Figure 9. 3D schematic of a 14-pick point rigging assembly

2.2.5 16 pick-point segments

A 16-point configuration is one of the most complex systems used for large modular lifts. It consists of two 8-point assemblies, with a spreader bar and two diagonal slings at the fifth level (Figure 10).

Similar to the 8-point configuration, each 8-point segment is analyzed independently using the methods from Section 2.2.1. However, with the addition of a fifth level of rigging components, the spreader bar introduces new forces that must be carefully analyzed. In this case, the forces at level 5 must be determined through a free-body diagram of the spreader bar in the XZ plane, which will have unknowns in both the vertical and horizontal directions (Figure 11). Considering the equations already derived from section **2.2.1** for F_1R_4 (Equation 17) and F_1L_4 (Equation 18) in the left 8-pick point segment; and similarly those for F_2R_4 (Equation 19) and F_2L_4 (Equation 20) in the right 8-pick point segment; and by applying equilibrium, Equations (21) and (22) were obtained. And by solving the system of these 2 equations, FL_5 and FR_5 were

determined in Equations (23) and (24) respectively with the integration of Equations (25) and (26) for simplicity.

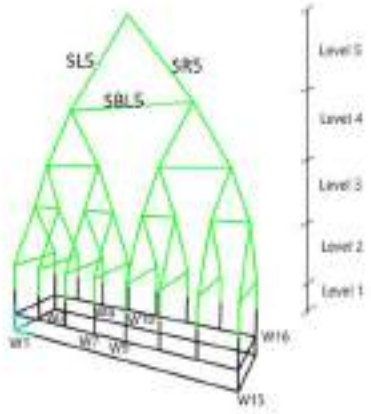


Figure 10. 3D schematic of a 16-pick point rigging assembly

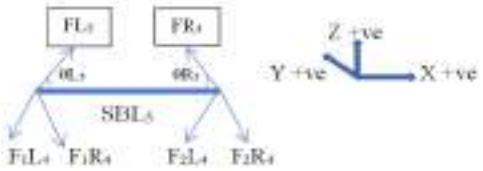


Figure 11. Free body diagram of spreader bar at level 5 with applied forces

The analysis for the 16-point system remains similar to the 8-point and 12-point configurations, but the additional level introduces complexity. To ensure the accuracy of the design, finite element analysis (FEA) should be used to model the forces acting on the system at this level. This advanced analysis will ensure that all components are properly designed and the forces are balanced throughout the system.

$$F_1R_4=\{(\cos\delta_{12}[F_1R_2\sin(\theta_1R_2-\gamma_{12})+F_1L_2\sin(\theta_1L_2+\gamma_{12})]+\cos\delta_{22}[F_2R_2\sin(\theta_2R_2-\gamma_{22})+F_2L_2\sin(\theta_2L_2+\gamma_{22})])((\sin(\theta_1R_3+\theta_1L_3))/([1+\cos(\theta_1L_3+\gamma_{13})\sin(\theta_1R_3-\gamma_{13})]))+(\cos\delta_{32}[F_3R_2\sin(\theta_3R_2-\gamma_{32})+F_3L_2\sin(\theta_3L_2+\gamma_{32})]+\cos\delta_{42}[F_4R_2\sin(\theta_4R_2-\gamma_{42})+F_4L_2\sin(\theta_4L_2+\gamma_{42})])((\sin(\theta_2R_3+\theta_2L_3))/(\cos(\theta_2R_3-\gamma_{23})[1+\cos(\theta_2L_3+\gamma_{23})\tan(\theta_2R_3-\gamma_{23})]))\}\cos(\theta_1L_4+\gamma_{14})/\sin(\theta_1L_4+\theta_1R_4) \quad (17)$$

$$F_1L_4=\{(\cos\delta_{12}[F_1R_2\sin(\theta_1R_2-\gamma_{12})+F_1L_2\sin(\theta_1L_2+\gamma_{12})]+\cos\delta_{22}[F_2R_2\sin(\theta_2R_2-\gamma_{22})+F_2L_2\sin(\theta_2L_2+\gamma_{22})])((\sin(\theta_1R_3+\theta_1L_3))/([1+\cos(\theta_1L_3+\gamma_{13})\sin(\theta_1R_3-\gamma_{13})]))+(\cos\delta_{32}[F_3R_2\sin(\theta_3R_2-\gamma_{32})+F_3L_2\sin(\theta_3L_2+\gamma_{32})]+\cos\delta_{42}[F_4R_2\sin(\theta_4R_2-\gamma_{42})+F_4L_2\sin(\theta_4L_2+\gamma_{42})])((\sin(\theta_2R_3+\theta_2L_3))/(\cos(\theta_2R_3-\gamma_{23})[1+\cos(\theta_2L_3+\gamma_{23})\tan(\theta_2R_3-\gamma_{23})]))\}\cos(\theta_1R_4-\gamma_{14})/\sin(\theta_1L_4-\theta_1R_4+2\gamma_{14}) \quad (18)$$

$$F_2R_4=\{(\cos\delta_{12}[F_1R_2\sin(\theta_1R_2-\gamma_{12})+F_1L_2\sin(\theta_1L_2+\gamma_{12})]+\cos\delta_{22}[F_2R_2\sin(\theta_2R_2-\gamma_{22})+F_2L_2\sin(\theta_2L_2+\gamma_{22})])((\sin(\theta_1R_3+\theta_1L_3))/([1+\cos(\theta_1L_3+\gamma_{13})\sin(\theta_1R_3-\gamma_{13})]))+(\cos\delta_{32}[F_3R_2\sin(\theta_3R_2-\gamma_{32})+F_3L_2\sin(\theta_3L_2+\gamma_{32})]+\cos\delta_{42}[F_4R_2\sin(\theta_4R_2-\gamma_{42})+F_4L_2\sin(\theta_4L_2+\gamma_{42})])((\sin(\theta_2R_3+\theta_2L_3))/(\cos(\theta_2R_3-\gamma_{23})[1+\cos(\theta_2L_3+\gamma_{23})\tan(\theta_2R_3-\gamma_{23})]))\}\cos(\theta_2L_4+\gamma_{24})/\sin(\theta_2L_4+\theta_2R_4) \quad (19)$$

$$F_2L_4=\{(\cos\delta_{12}[F_1R_2\sin(\theta_1R_2-\gamma_{12})+F_1L_2\sin(\theta_1L_2+\gamma_{12})]+\cos\delta_{22}[F_2R_2\sin(\theta_2R_2-\gamma_{22})+F_2L_2\sin(\theta_2L_2+\gamma_{22})])((\sin(\theta_1R_3+\theta_1L_3))/([1+\cos(\theta_1L_3+\gamma_{13})\sin(\theta_1R_3-\gamma_{13})]))+(\cos\delta_{32}[F_3R_2\sin(\theta_3R_2-\gamma_{32})+F_3L_2\sin(\theta_3L_2+\gamma_{32})]+\cos\delta_{42}[F_4R_2\sin(\theta_4R_2-\gamma_{42})+F_4L_2\sin(\theta_4L_2+\gamma_{42})])((\sin(\theta_2R_3+\theta_2L_3))/(\cos(\theta_2R_3-\gamma_{23})[1+\cos(\theta_2L_3+\gamma_{23})\tan(\theta_2R_3-\gamma_{23})]))\}\cos(\theta_2R_4-\gamma_{24})/\sin(\theta_2L_4-\theta_2R_4+2\gamma_{24}) \quad (20)$$

$$FL_5\cos(\theta L_5+\gamma_5)-FR_5\cos(\theta R_5-\gamma_5)+F_1R_4\cos(\theta_1R_4-\gamma_{14})-F_1L_4\cos(\theta_1L_4+\gamma_{14})+F_2R_4\cos(\theta_2R_4-\gamma_{24})-F_2L_4\cos(\theta_2L_4+\gamma_{24})=0 \quad (21)$$

$$FL_5\sin(\theta L_5+\gamma_5)+FR_5\sin(\theta R_5-\gamma_5)-(F_1L_4\sin(\theta_1L_4-\gamma_{14})-F_1R_4\sin(\theta_1R_4-\gamma_{14})-F_2R_4\sin(\theta_2R_4-\gamma_{24})-F_2L_4\sin(\theta_2L_4+\gamma_{24})=0 \quad (22)$$

$$FL_5=\frac{c_2\cos(\theta R_5-\gamma_5)+c_1\sin(\theta R_5-\gamma_5)}{\sin(\theta L_5+\theta R_5)} \quad (23)$$

$$FR_5=\frac{c_2\cos(\theta R_5-\gamma 5)\cos(\theta L_5+\gamma_5)+c_1\sin(\theta R_5-\gamma_5)\cos(\theta L_5+\gamma_5)-C1\sin(\theta L_5+\theta R_5)}{\cos(\theta R_5-\gamma_5)\sin(\theta L_5+\theta R_5)} \quad (24)$$

FL_5 and FR_5 are respectively the left and right sling forces at the fifth level of the rigging assembly; θL_5 and θR_5 are respectively the left and right sling angles with the spreader bar SBL_5 at the fifth level; γ_5 is the slope of the spreader bar at the fifth level. c_1 and c_2 are combinations of constants given in Equations (25) and (26).

$$c_1 = -(F_1R_4 \cos(\theta_1R_4 - \gamma_{14}) - F_1L_4 \cos(\theta_1L_4 + \gamma_{14}) + F_2R_4 \cos(\theta_2R_4 - \gamma_{24}) - F_2L_4 \cos(\theta_2L_4 + \gamma_{24}) \quad (25)$$

$$c_2 = F_1L_4 \sin(\theta_1L_4 + \gamma_{14}) + (F_1R_4 \sin(\theta_1R_4 - \gamma_{14}) + F_2L_4 \sin(\theta_2L_4 + \gamma_{24}) + F_2R_4 \sin(\theta_2R_4 - \gamma_{24}) \quad (26)$$

3 Conclusion

This study has explored the limitations of the static equilibrium method in designing rigging systems for modular lifting configurations with more than 6 lifting points ignored in previous studies. While static equilibrium is effective for smaller configurations, it becomes less reliable for systems with more than six lifting points, where additional techniques, such as finite element analysis (FEA), could be required. The findings provide a framework for designing rigging assemblies for larger modular systems, offering insights into how to precisely and efficiently balance forces and ensure stability during lifting operations.

Despite not being the ideal method for analyzing large rigging configurations, the verification process confirmed, through a detailed mathematical computation process, that static equilibrium methods are still valid for configurations like 8-point and 16-point systems despite the need for successive analyses. However, this process is: *(i)* tedious because it takes time to perform analysis calculations for each level and proceed to the next higher one(s); *(ii)* error-prone since any error in the earlier steps of calculation would be carried forward to higher levels which would lead to design and selection inaccuracies; and *(iii)* limited to sling force computations only excluding those for spreader bars and other rigging components. On the other hand, for other configurations, particularly the 10-point and 14-point systems, the method becomes less effective due to the complexity of the force distribution and the non-matching number of levels. Moreover, the method becomes invalid with the 12-pick point system due to having more unknowns than equations.

This study presents a hands-on insight to introduce a new comprehensive rigging design methodology that: (i) computes the sling angles and sling length justifications to establish equilibrium of rigging arrangements with 6 or more lifting points. (ii) simplifies the structural analysis by applying simultaneous solutions (with the help of structural analysis software preferably applying the FEA method), instead of systematic or successive analyses based on level-to-level free-body diagrams which are time-wasting, error-prone, and invalid for all n-pick point configurations. (iii) has the potential to account for not only static loads as in the existing studies, but also address dynamic effects (i.e., forces and stresses) associated with the lifted modules such as wind loading and dynamic vibrations, in addition to the module's deformations.

References

[1] H. Taghaddos, U. Hermann, and A. Abbasi. Automated crane planning and optimization for modular construction. *Automation in Construction, 95*, 219-232, 2018.

[2] J. Cho, K. Lee, and H. Lee. Impact analysis of the new OSHA cranes and derricks regulations on crane operation safety. *KSCE Journal of Civil Engineering, 21*(1), 54-66, 2017.

[3] S. MinayHashemi, S. Han, J. Olearczyk, A. Bouferguene, M. Al-Hussein, and J. Kosa. Automated rigging design for heavy industrial lifts. *Automation in Construction, 112*, 103083, 2020.

[4] Institute, A. P. (2000). API RP2A: Recommended Practice for Planning, Designing, and Constructing Fixed Offshore Platforms-Working Stress Design: Upstream Segment. API Recommended Practice 2A-WSD (RP 2A-WSD): Errata and Supplement 1, December 2002: *American Petroleum Institute*.

[5] Veritas, D. N. Guideline for offshore structural reliability analysis: application to jacket platforms. *Det Norske Veritas,* 1996.

[6] J. W. Bunce and T. A. Wyatt. Development of unified design criteria for heavy lift operations offshore. In *Proceedings of the Offshore Technology Conference*, OTC-4190-MS, Houston, Texas, 1982.

[7] K. Sam. Offshore heavy lift rigging analysis using spreadsheet. *Practice Periodical on Structural Design and Construction*, 14(2), 63-69, 2009.

[8] C. Crull. Direct determination of sling tensions in heavy rigging. In *Proceedings of the Offshore Technology Conference*, OTC-6395-MS, Houston, Texas, 1990.

[9] S. Hashemi, S. Han, J. Olearczyk, A. Bouferguene, M. Al-Hussein and J. Kosa. Automated mathematical-based design framework for the selection of rigging configuration. In *Proceedings of the 36th ISARC*, pages 172-178, Banff, Canada, 2019.

[10] J. Olearczyka, M. Kainata, J. Kosab, S. Adeeba, and M. Al-Husseina. Heavy lift rigging algorithm for construction operation (*iRigging),* 2015.

Evaluation Framework for Indoor Localization Systems in the AEC Environment

Lukas Fritzsche[#, 1], Rolf Starke[#, 1], Benjamin Felbrich[2], Ilija Vukorep[1]

[1] Faculty of Architecture, Civil Engineering and Urban Planning, BTU Cottbus-Senftenberg, Germany
[2] Felbrich Holding UG
[#] contributed equally
lukas.fritzsche@b-tu.de, starke@b-tu.de, benjamin@felbrich.com, ilija.vukorep@b-tu.de

Abstract –

Accurate spatial localization is critical in the construction industry, indoor robotics, and other applications involving actor movement within buildings. This paper introduces a combined software and hardware framework designed to evaluate indoor localization systems leveraging optical tracking integrated with a ROS communication system. It supports recording location data, aligning trajectories and evaluating the recorded samples against the ground truth allowing for real world comparison within a custom setup. While it is designed to evaluate any localization system, a presented case study compares two established SLAM algorithms regarding their suitability for an AEC application.

Keywords –

localization evaluation, indoor localization, mobile robotic, SLAM, ROS

1 Introduction

The rapid advancements in indoor localization technologies have led to their increasing integration across diverse industries, including construction, robotics, and indoor navigation [1]. Accurate spatial localization is pivotal for Architecture, Engineering and Construction (AEC) applications ranging from autonomous robots[2] operating in confined spaces to real-time monitoring of dynamic processes (Figure 1). However, the multitude of available localization methods — each with unique strengths, limitations, and precision levels — poses a challenge for selecting the most suitable system for specific applications [3].

The proposed framework does not aim to create a new benchmark for evaluating SLAM algorithms. Instead, it builds on existing benchmarks, metrics, and computer vision algorithms to help industries adopt and improve localization systems, advancing robotic development in AEC. The system supports a context-aware comparison of suitable localization methods in a realistic environment, rather than validating their performance under laboratory conditions. This novel evaluation framework combines optical tracking with a modular ROS-based software system. Designed to measure and compare the precision of localization systems, the framework provides a robust method for evaluating different approaches, allowing for informed decisions in research and development. While the framework is adaptable to various technologies, this work demonstrates the workflow in a case study focusing on two widely used localization algorithms.

2 Background and Related Work

Precise indoor localization of actors in AEC — like robots, machines, and sensing devices — is crucial for their operation. Indoor localization has the problem of being in the “shadow” of the building and thus cannot utilize satellite localization.

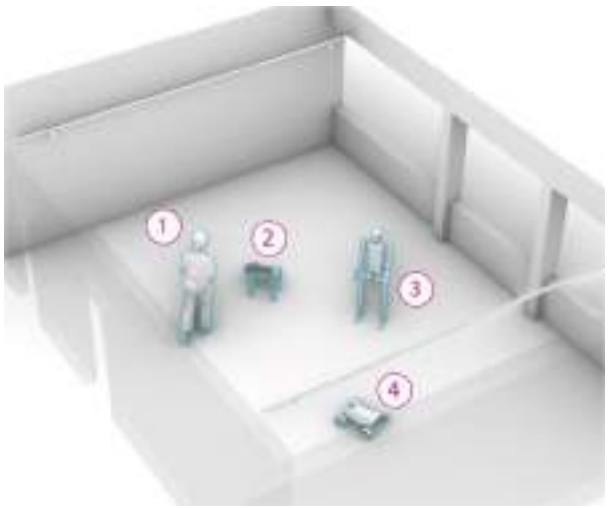


Figure 1. indoor mobile robots or devices that use localization: 1) handheld device; 2) quadruped robot; 3) humanoid robot; 4) four-wheeled robot.

For precise localization inside buildings there are many technologies like Inertial Measurement Units (IMU), Visual Odometry, Simultaneous Localization and Mapping (SLAM) with cameras or Lidar. These different methods are often unified in sensor fusion systems to increase precision [4].

To compare and evaluate the precision of different localization methods common databases and benchmark suites are used that include high-quality sensor data collected from real-world driving scenarios, including stereo camera images, 3D point clouds from LiDAR, and GPS/IMU data, enabling researchers to test and compare algorithms under consistent conditions. While there is an abundance of outdoor datasets like Aachen Day-Night Dataset [5], CrowdDriven Dataset [6], Extended CMU Seasons Dataset, RobotCar Seasons Dataset, Symphony Seasons Dataset some indoor datasets like InLoc Dataset [7], Gangnam Station and Hyundai Department Store Datasets, ETH-Microsoft Dataset [8] or the TUM RGB-D Dataset [9] can be found as well. With the help of these datasets researchers can develop their best localization algorithm or hardware system.

Once a localization algorithm for an application is chosen, the next steps require testing it in the context of the application in order to validate the promised precision. This is where the presented frameworks evaluation and calibration commences. To our knowledge there are no out of the box solutions allowing to determine and compare the suitability of different localization algorithms for a custom environment.

Technical descriptions of applied evaluation and calibration methods in previously mentioned localization software solutions differ widely. For example, in [9] handheld Kinect RGBD cameras were tracked by 8 motion tracking high-speed cameras to accurately track the pose of the localization device. In [10] IMUs were added to camera sensors and a motion tracking system was installed in the starting position where also the scanning ended (loop closure) providing ground truth for the TUM VI dataset. A similar setup was also used for the creation of the TUM VIE dataset [11] where the sensorial layers were enriched with stereo-event data. Other SLAM evaluation frameworks [12], [13] compare different algorithms in one benchmark environment, while the presented framework can be used to compare one algorithm in different environments. Existing frameworks also concentrate on calculating error metrics for specified trajectories and are not equipped to provide a comprehensive workflow for recording the positional data.

3 System description

3.1 Hardware Setup

A crucial part of the proposed system is to provide precise ground truth positional data for the localization actor which is achieved by an outside-in tracking system based on high-speed IR cameras. This section provides a detailed overview of the hardware components used, their specifications, and their integration.



Figure 2. 1) Hardware setup in the case-studies lab environment (7x10m) consisting of 2) four OptiTrack Prime X 13 cameras (ground truth) and 3) a depth camera for localization (e.g. RealSense D455)

Four OptiTrack Prime X 13 cameras are installed on the ceiling to observe a designated tracking area with the size 7 x 10 m (Figure 2). These cameras, chosen for their high-speed capture rates and submillimetre precision of ±0.20 mm [14], are positioned strategically to provide comprehensive coverage of the monitored space. The cameras are connected to the central computer via an 8-port Ethernet PoE/PoE+ switch, specifically the TRENDnet TPE-TG80G model. This switch enables stable data transmission and supplies power to the cameras through Ethernet cables, simplifying the overall cabling requirements and ensuring uninterrupted operation. A workstation running Windows 11 and

OptiTrack's Motive 3.1.4 motion capture software [15] transforms the raw tracking data into user-defined frames. A calibration workflow is applied to align the coordinate spaces of the localization algorithm and the OptiTrack.

For the calibration, as a critical aspect of the setup, marker spheres with a diameter of 9.5 mm are employed. These markers are specifically designed to attach to the camera used for localization, facilitating accurate calibration by ensuring they are easily detectable by the OptiTrack cameras.

3.2 Software Setup

The software architecture supporting the optical tracking and localization system is designed to operate efficiently within a modular, containerized environment. This approach ensures scalability, ease of deployment, and isolation of various system components.

In order to achieve comparable results and to allow offline runs of computationally intensive localization systems, the framework supports a workflow where the localization sensor data is first recorded as *rosbags* [16] and the evaluation is performed at a later stage.

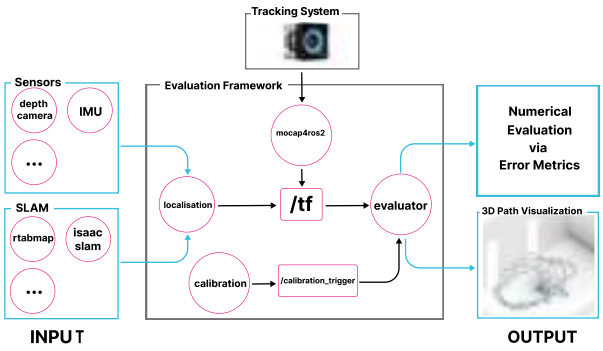


Figure 3. Overview of the proposed framework. Application specific components like sensors and localization algorithms (left) are utilized and extended by the proposed framework's components (centre) which record, align and evaluate positional data. Results are presented in the form of well-established error metrics as well as a visualization of the recorded trajectories (right).

3.2.1 Containerized ROS2 Instances

The foundation of the software system consists of ROS2 instances deployed as individual Docker containers. Each container encapsulates a specific functionality (Figure 3), ensuring a clean separation of responsibilities and enabling seamless communication between components via ROS2 nodes. This architecture leverages Docker's containerization capabilities to isolate dependencies, simplify system updates, and allow independent testing of each component.

3.2.2 OptiTrack Integration in ROS2 Docker

The motion capture system recording the ground truth localization data is wrapped in a ROS2 node hosted within a Docker container, which acts as a receiving counterpart to the Motive Software as NatNet client [17]. This node handles the optical tracking processes, transforming the tracking data captured from the OptiTrack Prime X 13 cameras and publishing the calculated positions of tracked objects.

The component is based on the mocap4ros2 package [18] and extended to send transform data in a format that is compatible with the overall setup. The container communicates with other system components by broadcasting positional data in real time, ensuring high accuracy and low latency.

3.2.3 Localization system in ROS2 Docker

A separate Docker environment encapsulates any given localization system, thereby ensuring its isolation and enabling seamless integration with diverse sources of localization. Within this environment the localization system needs to be configured to publish the measured pose via a ROS2 node. The localization component operates in parallel with the OptiTrack container, enabling the system to combine optical tracking and visual localization seamlessly.

3.2.4 Calibration and Evaluation System in ROS2 Docker

In a third dockerized environment the evaluation system runs as a custom ROS2 node. By subscribing to localization data published by the other system components and reacting to user input, as described in the calibration section, this unit handles the core evaluation logic. It provides a real time visualization of the different poses via the default 3D visualizer for ROS2, captures the data frames and provides logic for evaluating the captured data.

3.2.5 Inter-Component Communication

The ROS2 ecosystem facilitates communication between the Docker containers through ROS2 topics and nodes. Each container publishes data, such as positional information or error metrics, which other containers or scripts can subscribe to for further processing. The modularity provided by ROS2 ensures that each component remains loosely coupled while maintaining synchronization and data consistency across the system. Additionally, ROS2's design for distributed systems enables low-latency configurations across various computing environments, making the integration of a new localization component particularly straightforward due to the minimal setup required.

In order to maximize compatibility with different localization systems, the framework utilizes the tf2

package as a well-established standard for capturing coordinate frames over time [19]. The ground truth as well as the localization system component publish the measured poses as transform. The evaluation component is continuously consuming those transforms and processes them to be used for evaluation.

3.3 Initialization, Calibration and Trajectory Alignment

To evaluate the accuracy of the localization system, the deviation between the poses as measured respectively by the OptiTrack system and by the localisation system must be determined.

Since these originate from different systems their coordinate spaces will differ. The evaluation framework provides a registration workflow as a systematic approach to abstract from the substantial complexity inherent in aligning these poses and transforming them into a shared coordinate space.

3.3.1 Registering the Test Environment and Calibrating Ground Truth

In the initial step the test environment needs to be prepared. The motion capture system serving as ground truth must be calibrated by traversing the tracking volume with a standardized calibration device.

As part of this procedure, the origin position and orientation of the global coordinate space must be defined in the physical world by placing a specific marker on the ground floor. The alignment of the origin with a physical feature e.g. a wall or corner is recommended to greatly simplify subsequent evaluation, as it enables the alignment of a 3D model with minimal effort.

3.3.2 Measuring and Specifying Physical Configuration of the Localization Device

To track the localization device with OptiTrack, a sufficient number (3-5pcs) of markers must be fixed statically to the device.

Prior to the OptiTrack system tracking the pose of the device, it is necessary to register the marker configuration within the Motive Software. As part of this process a *rigid body* is created, which captures the static offsets between the markers as one physically connected object. After the rigid body is created, the corresponding coordinate frame should be adjusted to match ROS coordinate space definition as defined in as defined in REP103/105: x → forward, y → left, z → up [20], [21]

To facilitate the alignment of poses calculated by both localization and ground truth, the physical configuration of the tracking device must be specified and loaded into the evaluation framework. This can be achieved by measuring the offsets between the origin of the rigid body and the origin of the used tracking solutions e.g. the camera centre.

Placing the origin at the position of one of the used markers will simplify measuring the offsets.

The framework utilizes the well-established *universal robot description format (urdf)* [22] for reading the devices physical configuration.

3.3.3 Aligning Coordinate Frames

After the ground truth has been calibrated and the device has been set up for being localized in OptiTrack, an evaluation run can be conducted.

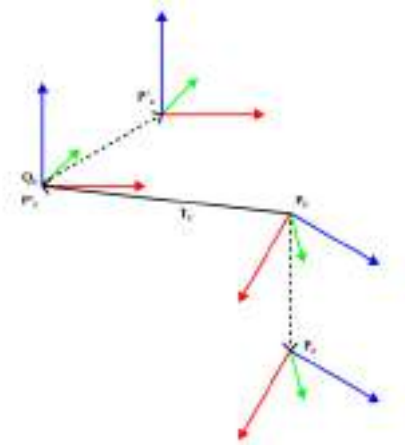


Figure 4. Alignment of the different coordinate systems of ground truth Q and the localization system P at calibration data sample t=0 and some subsequent sample t=n. Using the calibration transform T_C the calibrated transform P' can be obtained at any point in time.

As soon as both localization systems are running and the tracking device's markers are fixed, the alignment procedure of the evaluation system must be triggered. By pressing a key, the user starts the evaluation run. The evaluation system stores the tracked transform P_0 as well as the ground truth transform Q_0 at this timestamp *index* = 0. The recording of poses is started.

From now on the system calculates a new pose P'_0 representing the calibrated tracked pose (i.e. the pose measured by the localization system transformed into OptiTrack's coordinate space). In every subsequent sample with *index n* the poses Q_n and P'_n can be utilized to calculate the deviation as shown in Figure 4.

3.3.4 Timestamp Alignment

The data packages sent via NatNet streaming protocol are enriched with the timestamp when this data was recorded, measured in turn with the OptiTrack PC's clock. The OptiTrack ROS2 node publishes the corresponding transform with the original timestamp, compensating for latency.

Any pose evaluation needs to be calculated with exact matching timestamps for it to allow conclusions about the trajectory. Possible timestamp mismatches between the ground truth and the localization system can arise due to differences in sensor frequencies, outliers through sensor measurement errors, tracking losses in either the ground truth or the localization system, and delays introduced by the processing times of raw image data recording, the SLAM algorithm, and the transform alignment calculation. To compensate for those mismatches the delay in the localization system's timestamps, which is assumed to be static, needs to be determined and corrected for a static value or the timestamps need to be upsampled, downsampled or interpolated before the evaluation.

3.4 Means of Evaluation

To support a wide range of localization applications the proposed framework implements several means of evaluating the chosen localization system's precision. With the availability of the first RGB-D image datasets, which were infused with ground truth camera poses [9] the absolute trajectory error (ATE) as well as the relative error or relative pose error (RPE) became the predominant error metrics for localization systems. Adaptations of ATE and RPE calculations from public repositories [12] are included in the proposed framework.

3.4.1 Numerical Evaluation

RPE describes the difference between the estimated motion and the true motion for each pose, and is used to calculate drift, which describes the accumulation of sensor measurement errors during visual odometry [23]. While computationally expensive and difficult to interpret, RPE is less sensitive to estimation errors and allows for a detailed evaluation for any chosen segment size of the trajectory. Since it relies on motion, it is used in scenarios where accurate short term motion estimation is more important than precise global positioning e.g. wearable VR or AR devices.

ATE is a simple yet comprehensive evaluation metric. It calculates the absolute pose difference from position error, rotation error and velocity error at the time step $_i$ between two trajectories. Specifically, the localization systems trajectory P is aligned with the ground truth trajectory Q using the rigid-body transformation S

.

$$F_i := Q_i^{-1} S P_i \quad (1)$$

The resulting error F is commonly used to calculate the Root Mean Squared Error which reduces the errors over all time steps to a single number metric. The requirement for a ground truth trajectory and the lack of compensation for drift, which disproportionately emphasizes errors in the initial segment of the localizations system's trajectory compared to those occurring later, are recognized as limitations of the ATE metric. Nonetheless, it is a prominent metric for SLAM evaluation due to its simplicity, ease of comparison, and intuitive visual interpretability [23].

3.4.2 Graphical Evaluation

The proposed framework is specifically designed to utilize localization algorithms among others for construction site monitoring, 3D scanning, surveying, autonomous equipment navigation, indoor inspections.

Therefore it includes a custom ROS2 node which records every frame update, providing a trajectory or path visualization in *rviz* [24], as well as a grasshopper 3d [25] script, which translates the recorded path into control points of a polyline for a seamless integration with commercial CAD software (see Figure 5), allowing non-experts to quickly assess if a localization system is operating within desired tolerances for a chosen application.

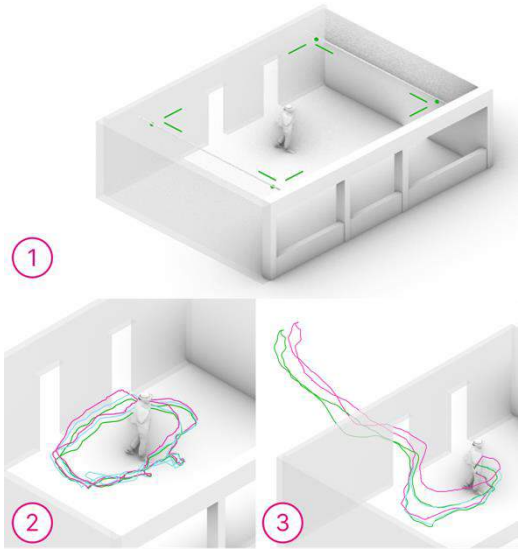


Figure 5. Visualization of two exemplary runs using the SLAM algorithms within a 1) lab environment. 2) Bottom shows 2) a circular path through a single room and 3) a run through multiple rooms in Rhino 3d. The ground truth is shown in *cyan*, results from RTAB-Map SLAM in *magenta* and ISAAC SLAM's trajectory in green.

4 Application of the Evaluation Framework

4.1 Description of a Case Study

To demonstrate and validate the proposed framework, an exemplary scenario is examined where a handheld device based on the NVIDIA Jetson AGX Orin platform equipped with a Stereolabs ZED X stereo camera is used for semi-automated building material and component inspection as part of so-called pre-demolition audits. In this scenario, machine learning based object recognition is combined with a visual inertial localization system to create an accurate material inventory. To achieve reliable results in this application, it is critical that the localization component of the system provides an accurate location of the actor. With RTAB-Map SLAM [26] and the Isaac Visual SLAM [27] two state-of-the-art visual-inertial SLAM systems were selected for their high on-paper accuracy as well as their high performance and flexibility within the ROS ecosystem. To determine which algorithm to use, an in-depth evaluation was performed in a realistic environment using the proposed evaluation framework.

4.2 Results

The results of the localization were then evaluated for the test run with a length of 28.4m. The ATE and RPE were calculated between the aligned trajectories of the SLAM systems and the OptiTrack serving as the ground truth. (see Figure 5, Table 1 and Table 2)

Table 1 Absolute Trajectory Error (ATE)

error metric	RTAB-Map	ISAAC
RMSE	0.194m	0.256 m
MEAN	0.159 m	0.224 m
STD	0.111 m	0.123 m
MIN	0.000 m	0.000 m
MAX	0.586 m	0.686 m

With an ATE Root Mean Squared Error *RMSE* of 19.4 cm RTAB-MAP outperforms the ISAAC SLAM.

Table 2 Relative Pose Error (RPE)

error metric	RTAB-Map	ISAAC
RMSE	0.058 m	0.047 m
MEAN	0.028 m	0.026 m
STD	0.051 m	0.039 m
MIN	0.000 m	0.000 m
MAX	0.458 m	0.337 m

For the Relative Pose Error *RPE*, when calculating drift per frame at $\Delta = 1$, the ISAAC SLAM outperforms the RTAB-Map SLAM, achieving a Root Mean Square Error *RMSE* of 4.7 cm.

Based on the presented Evaluations (Table 1 and Table 2), a logical inference is to prioritize RTAB-Map SLAM for tasks requiring high stability due to its low ATE, while selecting ISAAC SLAM for tasks demanding high accuracy, given its superior performance in RPE.

In the above-mentioned application of semi-automated building inventory, the prevention of double counting objects in consecutive frames is pivotal. Only with a low RPE the same object can be safely identified in different frames and correctly positioned in the localization algorithms map. Therefore, ISAAC SLAM should be selected for this application due to its superior RPE in the tested environment.

5 Conclusion, Limitations and Future Work

In order to validate the correctness of the results produced by the framework in the given case study, the error metrics are compared with the accuracy claimed in the literature.

Of all the datasets on which RTAB-Map has been run in [28], the TUM room sequence is the most comparable to the single room setup presented, as it is the only indoor handheld scenario. Within the benchmark, an ATE (RMSE) of 6.6 cm was achieved, which is approximately one third of the accuracy (19.4 cm) determined within the case study. Preliminary investigations have indicated that the observed discrepancy in accuracy is most likely due to the limited processing power of the embedded device and the significantly longer path length.

For ISAAC SLAM, it should be noted that available benchmarks are exclusively for the KITTI dataset. Despite the considerable differences between this scenario and the case study (long-distance outdoor, high-speed vehicle data), a comparison is nevertheless made to validate the feasibility. The translation error with respect to the trajectory length in [29] is given as 0.94%. The ATE (RMSE) of 25.6 cm corresponds to a translation error of 0.9% on the case study trajectory.

A notable difference between the case study results and previously published results highlights the need for individual evaluations to align with specific applications and environments. The proposed framework has been demonstrated to achieve comparable and practical results for this application.

One limitation of the proposed framework is the OptiTrack maximum tracking area size, which effectively limits the availability of ground truth data to

one room and makes it difficult to track across different spaces. By having a trajectory start and end within the tracking volume – as shown in 3) in Figure 5 – some fundamental assessments can still be made. For a detailed evaluation in multi-room (or even outdoor) environments another method needs to be chosen.

The accuracy of the OptiTrack system, as stated by the manufacturer, was neither independently verified nor incorporated into the ground truth trajectory of the proposed framework. However, given that the translational error rates of state-of-the-art SLAMs are an order of magnitude larger than OptiTrack's precision (even for shorter trajectories, e.g. 93 mm error with 10 m trajectory for ISAAC SLAM) the influence on the evaluation results is negligible.

Future work could evaluate different localization systems like LIDAR, Bluetooth, Wi-Fi systems and others to develop the proposed framework into a localization system primer for other applications.

Acknowledgments

We gratefully acknowledge the funding from Forschungsinitiative Zukunft Bau des Bundesministeriums für Wohnen, Stadtentwicklung und Bauwesen (reference number SWD-10.08.18.7-24.46).

References

[1] T. Aziz and I. Koo, "A Comprehensive Review of Indoor Localization Techniques and Applications in Various Sectors," *Appl. Sci.*, vol. 15, no. 3, p. 1544, Feb. 2025, doi: 10.3390/app15031544.

[2] J. M. Davila Delgado *et al.*, "Robotics and automated systems in construction: Understanding industry-specific challenges for adoption," *J. Build. Eng.*, vol. 26, p. 100868, Nov. 2019, doi: 10.1016/j.jobe.2019.100868.

[3] J. Li, Y. Song, Z. Ma, Y. Liu, and C. Chen, "A Review of Indoor Localization Methods Leveraging Smartphone Sensors and Spatial Context," *Sensors*, vol. 24, no. 21, p. 6956, Oct. 2024, doi: 10.3390/s24216956.

[4] J. Zhu, H. Li, and T. Zhang, "Camera, LiDAR, and IMU Based Multi-Sensor Fusion SLAM: A Survey," *Tsinghua Sci. Technol.*, vol. 29, no. 2, pp. 415–429, Apr. 2024, doi: 10.26599/TST.2023.9010010.

[5] T. Sattler *et al.*, "Benchmarking 6dof outdoor visual localization in changing conditions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8601–8610. Accessed: Jan. 14, 2025. [Online]. Available: http://openaccess.thecvf.com/content_cvpr_2018/html/Sattler_Benchmarking_6DOF_Outdoor_CVPR_2018_paper.html

[6] A. Jafarzadeh *et al.*, "Crowddriven: A new challenging dataset for outdoor visual localization," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 9845–9855. Accessed: Jan. 14, 2025. [Online]. Available: http://openaccess.thecvf.com/content/ICCV2021/html/Jafarzadeh_CrowdDriven_A_New_Challenging_Dataset_for_Outdoor_Visual_Localization_ICCV_2021_paper.html

[7] H. Taira *et al.*, "InLoc: Indoor Visual Localization with Dense Matching and View Synthesis," in *CVPR*, 2018.

[8] ETH Zurich Computer Vision Group and Microsoft Mixed Reality & AI Lab Zurich, "The ETH-Microsoft Localization Dataset." 2021. [Online]. Available: https://github.com/cvg/visloc-iccv2021

[9] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of RGB-D SLAM systems," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Vilamoura-Algarve, Portugal: IEEE, Oct. 2012, pp. 573–580. doi: 10.1109/IROS.2012.6385773.

[10] D. Schubert, T. Goll, N. Demmel, V. Usenko, J. Stueckler, and D. Cremers, "The TUM VI Benchmark for Evaluating Visual-Inertial Odometry," in *International Conference on Intelligent Robots and Systems (IROS)*, 2018.

[11] S. Klenk, J. Chui, N. Demmel, and D. Cremers, "TUM-VIE: The TUM Stereo Visual-Inertial Event Dataset," in *International Conference on Intelligent Robots and Systems (IROS)*, 2021.

[12] M. Grupp, "evo: Python package for the evaluation of odometry and SLAM." 2017. Accessed: Jan. 14, 2025. [Online]. Available: https://github.com/MichaelGrupp/evo

[13] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, Providence, RI: IEEE, Jun. 2012, pp. 3354–3361. doi: 10.1109/CVPR.2012.6248074.

[14] OptiTrack, "Primex 13 Datasheet." Accessed: Jan. 14, 2025. [Online]. Available: https://optitrack.com/cameras/primex-13/specs.html

[15] OptiTrack, "Motive – Optical motion capture software." Accessed: Jan. 14, 2025. [Online]. Available: https://optitrack.com/software/motive/

[16] "rosbag." Accessed: Jan. 14, 2025. [Online]. Available: https://github.com/ros2/rosbag2

[17] OptiTrack, “NatNet SDK.” Accessed: Jan. 14, 2025. [Online]. Available: https://optitrack.com/software/natnet-sdk/

[18] F. Martin, J. M. Guerrero, A. A. Garcia, F. Rodriguez, and V. Matellan, “MOCAP4ROS2: An Open Source Framework for Motion Capture Systems in Robotics,” in *Proceedings of the 18th International Symposium on Open Collaboration*, in OpenSym ’22. New York, NY, USA: Association for Computing Machinery, Sep. 2022, pp. 1–3. doi: 10.1145/3555051.3555076.

[19] T. Foote, “tf: The transform library,” in *2013 IEEE Conference on Technologies for Practical Robot Applications (TePRA)*, Apr. 2013, pp. 1–6. doi: 10.1109/TePRA.2013.6556373.

[20] “REP 105 -- Coordinate Frames for Mobile Platforms (ROS.org).” Accessed: Dec. 20, 2024. [Online]. Available: https://www.ros.org/reps/rep-0105.html

[21] T. Foote and M. Purvis, “REP 103 -- Standard Units of Measure and Coordinate Conventions (ROS.org),” REP 103 -- Standard Units of Measure and Coordinate Conventions (ROS.org). Accessed: Jan. 14, 2025. [Online]. Available: https://www.ros.org/reps/rep-0103.html

[22] “urdf - ROS Wiki.” Accessed: Jan. 14, 2025. [Online]. Available: https://wiki.ros.org/urdf

[23] Z. Zhang and D. Scaramuzza, “A Tutorial on Quantitative Trajectory Evaluation for Visual(-Inertial) Odometry,” in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Madrid: IEEE, Oct. 2018, pp. 7244–7251. doi: 10.1109/IROS.2018.8593941.

[24] Dave Hershberger, “3D visualization tool for ROS.” Accessed: Jan. 14, 2025. [Online]. Available: http://wiki.ros.org/rviz

[25] McNeel, “Grasshopper.” Accessed: Jan. 15, 2025. [Online]. Available: https://www.grasshopper3d.com/

[26] M. Labbé and F. Michaud, “RTAB-Map as an open-source lidar and visual simultaneous localization and mapping library for large-scale and long-term online operation,” *J. Field Robot.*, vol. 36, no. 2, pp. 416–446, 2019, doi: 10.1002/rob.21831.

[27] NVIDIA, “NVIDIA Isaac SLAM.” Accessed: Jan. 14, 2025. [Online]. Available: https://github.com/NVIDIA-ISAAC-ROS/isaac_ros_visual_slam

[28] M. Labbé and F. Michaud, “RTAB-Map as an Open-Source Lidar and Visual SLAM Library for Large-Scale and Long-Term Online Operation,” *J. Field Robot.*, vol. 36, no. 2, pp. 416–446, Mar. 2019, doi: 10.1002/rob.21831.

[29] A. Korovko, D. Robustov, D. Slepichev, E. Vendrovsky, and S. Volodarskiy, “Realtime Stereo Visual Odometry,” 2018.

Ultra Low Delay and Resilient Multi-channel Video Encoder for Efficient Remote Construction

Yasuyuki Nakajima[1], Yasukazu Hontama[1], Kazuhiko Shibairi[1], Yasuyuki Jitsuta[2], Daisuke Endo[3], Genki Yamauchi[3], Takeshi Hashimoto[3], and Hitoshi Itoh[4]

[1]Hytec Inter Co. Ltd., Japan
[2]Jitsuta Chugoku Co. Ltd., Japan
[3]Public Works Research Institute, Japan
[4]Chudenko Corporation, Japan

y-nakajima@hytec.co.jp, hontama@hytec.co.jp, shibairi@hytec.co.jp, y.jitsuta@jtch.co.jp, endou-d177cl@pwri.go.jp, yamauchi-g573bs@pwri.go.jp, t-hashimoto@pwri.go.jp, h.itoh@chudenko.co.jp

Abstract –

Efficient remote operation of construction machines equipped with multiple video cameras requires smooth and real-time video transmission to simulate the experience of operating machinery in person. Due to the diverse network environments at construction sites, including satellite, Wi-Fi, and both public and private LTE/5G networks, an ultra-low delay video codec with adaptive bitrate control is crucial to maintaining low end-to-end latency in these environments. In this paper, the development of a multi-channel video encoder for the above requirements is described demonstrating the codec's performance across five types of wireless networks at a remote construction site.

Keywords –

Unmanned construction, Remote operation, Low delay video transmission, Network resilience

1 Introduction

In 2024, Japan's Ministry of Land, Infrastructure, Transport and Tourism (MILT) introduced the 'i-Construction 2.0' action plan, aiming to automate construction sites by 2040 [1]. In this plan, it is reported that Japan's working-age population is expected to be 20% smaller in 2040 than in 2023. Therefore, the plan focuses on labor reduction via increase of productivity through automation and autonomy in construction. For this purpose, remote construction technology is expected to help to solve labor issues as a step toward automated and autonomous construction.

In the remote construction, it is necessary to observe the situation of the remote site via the captured video, but there is a possibility of decrease of construction efficiency due to transmission delay and packet loss when using wireless networks. In this paper, firstly issues related to the remote construction using wireless networks such as private 5G and LEO (Low Earth Orbit) satellite networks are discussed. Next, we detail the development of a multi-channel, low-delay video codec to address these challenges.

Through experiments conducted across five types of wireless networks, we demonstrated that the ultra-low delay codec supports adaptable remote construction architectures tailored to each network.

2 Issues and countermeasures for remote construction

Figure 1 shows a typical system configuration of a remote construction and its issues (#1-#4) for efficient operation. In the following sections, we describe each issue and its countermeasures.

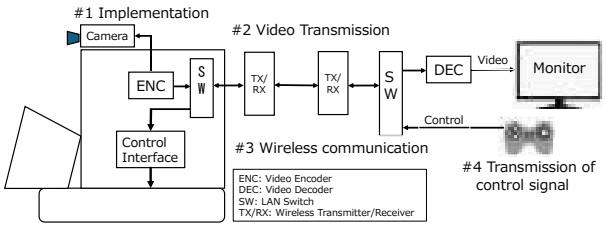


Figure 1. System configuration of remote construction

2.1 Issues and countermeasures for implementation

First of all, the issues with remote, automated, and autonomous operations of construction machines lay in the fact that it requires high expertise for installation and setup in such fields as video transmission, machine control, and communication networks since these factors are involved for the total systems. For example, four

video cameras facing forward, backward, left, and right would be installed on a construction machine to grasp the field situation and spot any danger. Since each camera requires a separate video encoder, a multi-channel encoder is preferable to reduce both cost and space. In addition, all the setup processes for the total system should be as simple as possible.

2.2 Issues and countermeasures for video transmission

As for the issues in remote operation using video, (1) end-to-end delay, (2) coding bitrate fluctuation, and (3) communication network quality should be pointed out. For example, if the advanced video coding architecture with temporally separated frames is employed to enhance the prediction performance, the coding efficiency can be improved. However, this might affect coding delays since the encoder has to wait for the following frames to encode the current frame. In the following subsections, each issue and countermeasures are described.

2.2.1 End-to-End delay

In this subsection, methods for end-to-end low delay architecture and target delay are discussed.

Remote construction efficiency would be directly affected by end-to-end delays [2]. As shown in Figure 1, each component such as video camera, video encoder, or communication network involves a certain amount of delay. Therefore, it is required to decrease each delay as much as possible. For example, 60fps video camera should be used instead of 30fps since the delay becomes 16.6msec instead of 33msec. On the other hand, delays of video codec and communication network may vary from tens to hundreds of msec, and suitable components should be chosen to realize very low end-to-end delay.

As for the human sensitivity on delay, it is reported that a delay of 100msec or over can be perceived [3]. For the electric rear-view mirrors for automobiles, the UN regulated the end-to-end delay not to take more than 200msec [4]. As for the remote construction, it is reported that the work efficiency was almost the same as the onsite remote operation when the end-to-end delay was less than 300msec [2]. Therefore, this figure can be recognized as one of the targets for the efficient remote construction.

2.2.2 Video coding delay and bitrate fluctuation

Here, delay and coding bit rate fluctuation mechanism caused by video coding scheme and countermeasures are addressed.

The international standard for video coding, MPEG (ranging from MPEG-1 and MPEG-2 to recent standards HEVC and VVC), combines intraframe coding (I frames) with interframe coding (P frames and B frames) to achieve a high compression ratio [5]. The intraframe coding reduces redundant information within the frame, whereas the interframe coding reduces redundant information between frames located temporally earlier and/or later. Here, P frames only use the earlier frames as a forward prediction and B frames can use both the earlier and the later frames as a bi-directional prediction.

In terms of coding efficiency, I frames are usually inserted every 0.5sec or 1sec to be used as random-access points, and the other frames are coded by interframe coding (P or B frames). Normally, considering the high correlation of frames in the temporal domain, P frames are inserted every few frames and the rest of the frames are coded with B frames to enhance the coding efficiency as shown in Figure 2(a).

#1 #2 #3 #4 #5 #6….
(a) I B B P B B P … I B B P…
(b) I P P P P P P … I P P…
(c) I P P P P P P … P P P…
Figure 2. Examples of coding frame layouts

Although this IBP frame structure can achieve a high compression ratio, there are two issues in terms of delay and coding bit amount for remote construction using network.

The first issue arises from variations in coding delay based on the video codec architecture. As stated earlier, MPEG uses interframe coding predicted from the earlier and/or later frames. Therefore, both encoder and decoder must wait for the later frames to encode/decode the current frame, which results in additional delay. In the case of Figure 2(a), for example, the frame #2 is coded as a B frame, but it can be coded after the frame #4 is coded as a P frame.

The second issue evolves from the fact that coding bit amount varies depending on the coding frame types. In the IBP coding structure, I frames have the largest bit amounts and B frames have the smallest bit amounts. For example, there is a case where the ratio of bit amounts for B:P:I frames becomes 1:10:25 [6]. Therefore, there is a possibility where the coding bit amounts of I frames are very large and it may take some time to send I frame data for limited network bandwidth. Then in the decoder side, the decoding timing of I frames becomes late for playback timing which may result in a frame drop.

In order to achieve low delay coding, the use of only earlier frames (i.e. IP frame structure) for the interframe coding as shown in Figure 2(b) is preferable in exchange for lower coding efficiency than that of IBP frame structure. As for the coding bit amount fluctuation, the introduction of intra-slice architecture within an IP frame structure as shown in Figure 2(c) which is further described in Section 3 may alleviate the bit amount variation.

2.3 Issues and countermeasures for communication networks

2.3.1 Communication network quality

In both wired and wireless networks, it is necessary to pay attention to four types of quality degradation causes (transmission delay, jitter, packet loss, and bandwidth fluctuation) for video transmission over IP networks. In this section, these issues and countermeasures are described.

As for the delay, low delay transmission can be realized when local networks such as Wi-Fi and private 4G(LTE)/5G. However, when public networks such as 4G(LTE) and 5G are employed for remote operation, a few dozen msec is usually involved due to transmission over the backbone and the access networks. In the case of satellite networks, GEO (Geostationary Earth Orbit) satellite communication has a few hundred msec roundtrip delay, whereas less than 100msec delay can be expected in the LEO satellite communication. To choose the desired network, network delay should be investigated before designing the remote construction system.

Regarding the jitter, since the video decoder is required to output the decoded frames to a video monitor within the playback timing, there is a possibility of frame drops when the packet data arrival is delayed due to the jitter. To solve this issue, the decoder buffer size should be designed to absorb the maximum jitter which can be determined by the actual amount of jitter.

As for the packet loss, obstacles and/or noise may cause transmission losses, especially in a wireless environment. The former packet loss is caused when the connection between a construction machine and a base station is interrupted by the backhoe movement, for example. The latter packet loss is caused by heavy rain and/or interference where packet data cannot be recovered due to noisy signals, for example. To solve the former packet loss issue, ARQ (Automatic Repeat reQuest) technique should be used to retransmit the lost packets while it involves additional delay. To solve the latter issue, FEC(Forward Error Correction) should be used, although it requires additional overhead data for error correction.

With regard to the bandwidth(throughput), the upload bandwidth from a construction machine is a very important parameter which determines transmission video quality. In the case of public networks such as 4G and 5G, more than a few Mbps throughputs can be expected for upload link. Recently, satellite networks using LEO became widely available and users can get up to 10Mbps for upload via Starlink [7], whereas only a few hundred kbps bandwidth can be used in satellite networks such as Inmarsat on the GEO. However, there is an issue of bandwidth variation of public networks since it may change dynamically according to the amount of network usage. To solve this, the video encoder should adapt to the available bandwidth.

2.3.2 Wireless networks

Although the history of remote construction in Japan is about 30years, the wireless network environment has been greatly advanced in this period. The unmanned construction using video cameras installed on the construction machines first adopted and applied for the whole disaster recovery work after the eruption of Mt.Unzen from 2km distance under wireless and relay environment at 2Mbps in 1993 [8].

In the damaged Fukushima nuclear power plant after the East Japan Great Earthquake in 2011 in Japan, ten construction machines (two 600-ton class large crawler tower cranes and eight construction machines) with video cameras were remotely operated from 500m distance using five mesh Wi-Fi base stations at 5GHz [8]. Recently, high speed wireless networks such as 5G network and LEO satellite network became widely available and a wide choice of wireless networks including Wi-Fi became possible. However, as stated earlier it is necessary to transmit low delay, stable, and high-quality video over the wireless network to realize an efficient remote construction.

In addition, resilient wireless networks should be constructed using mesh type network and/or error protection techniques such as ARQ for such cases as disconnection due to movements of construction machines, for example. It should be noted that preparation of a backup network would also be considered for network resiliency and safe remote operation. Another issue we should consider is that adaptive bitrate control for stable transmission is also very important since the bandwidth of wireless networks may fluctuate due to network congestion and/or interference, for example. In section 5, we show the solutions toward the above network related issues employing five types of broadband wireless networks and ARQ to realize efficient remote operations of construction machines.

2.4 Control signal transmission

In order to remotely control construction machines, control signals should be transmitted between the construction machines and the control sites. The amount of coded video data will become very large when the construction machines move or rotate since a large amount of coded video data is expected due to large motion or scene changes. Therefore, there are many cases where video is transmitted over Wi-Fi, while control signal is transmitted over the other network such as specified low power radio as in the case of Fukushima [8], which may result in complicated system

configuration. In order to transmit both the video and the control data in the same network, it is necessary to strictly control the total amount of coded video data below the bandwidth limit under any circumstances.

3 Development of multi-channel ultra low delay video codec

In this section, the development of video codec to solve the issues of low delay optimization, bit rate control mechanism and multi-channel support is described. Table 1 summarizes the main issues and solutions.

Table 1. Main issues and solutions

Issues	Solutions
Codec implementation	4ch HD encoder
Coding delays	IP frame structure
I-frame bursts	Intra slice coding
Bandwidth fluctuations	Strict rate control

For situation awareness and safe operations, a typical construction machine is equipped with four video cameras positioned at the front, rear, and both sides. In addition, at least HD quality video cameras would be required for the efficient remote operation [9]. Considering the above resolution requirement and overall network bandwidth, an all-in-one type four channel 2K low delay video encoder(LVRC-4000) based on the 4K HEVC video encoder (it can only support IP frame structure) [10] was developed, which can be installed on a construction machine (Figure 3 and Figure 4).

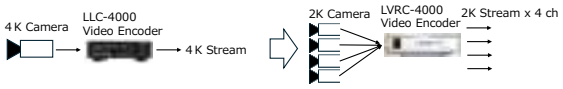


Figure 3. Development of 4ch HD video encoder

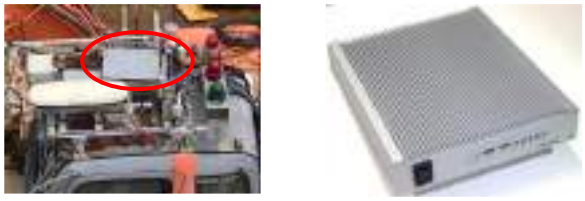


Figure 4. Ultra-low latency 4ch HD video encoder

As for the video coding structure, we use IP frame structure shown in Figure 2(b). Unlike the IBP structure, which uses both forward and backward predictions, the IP structure relies solely on forward prediction. While this reduces coding efficiency, it significantly lowers codec delay by eliminating the need to wait for future frames during encoding and decoding.

To address bandwidth fluctuations, the codec employs strict bitrate control, ensuring that the data rate remains within the available bandwidth at all times. In the case of IP frame structure shown in Figure 2(b), periodic insertion of I frames can cause bitrate spikes that exceed available bandwidth, leading to potential transmission delays or frame drops. Therefore, as an alternative structure, we have adopted an intra slice scheme within IP frame structure, which can also be used in HEVC coding. In this case, as shown in Figure 2(c), the first frame is encoded as I frame, but all the subsequent frames are coded by P frames where a slice of P frame is intra coded periodically. This intra slice coding can improve bitrate stability by periodically encoding small portions of each frame independently, preventing sudden spikes in data transmission.

For example, as shown in Figure 5, shadowed blocks (blocks on the third line in the figure) which construct a slice of blocks are intra coded. The position of the intra slice is moved at each frame. Then, in the next P frame, the slice of blocks below the current slice is intra coded. Since each P frame contains both intra- and inter-coded blocks, maintaining consistent picture quality and frame-to-frame bitrate stability is crucial. We implemented a strict rate control mechanism that regulates the bitrate of each frame, taking into account the differing data requirements of inter- and intra-coded blocks.

This approach ensures stable picture quality and bitrate while minimizing coding delay.

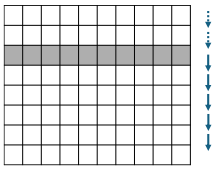


Figure 5. Intra slice coding

4 Video encoder performance

The video encoder performance was evaluated based on two key metrics: coding delay and bitrate fluctuation. Tests were conducted with four asynchronous video inputs, including high-motion scenes, to assess encoder stability under varying conditions. All the measurements in Section 4 and Section 5 were carried out for 3min. As shown in Figure 6, this experiment was carried out indoors since video codec performance test without wireless network can only be done in the indoor system.



Figure 6. Example of high-motion video input used for encoder evaluation

Figure 7 and Figure 8 show coding bitrate variations

of IP frame structure for one and four channels, respectively. Table 2 shows statistics (average bitrate, fluctuation range, standard deviations and 95% confidence intervals) of coding bitrate variations. Here, the coding bitrate is set to 3Mbps and the intra frame insertion period is set to 0.5sec which is generally used. In the case of one channel encoding, the coding bitrate changes rapidly according to motion and texture change, and periodical peaks at every 0.5sec can be seen in Figure 7. Since relatively large fluctuation is observed as shown in the table, it is necessary to set the coding bitrate to much lower than the available bandwidth in order to correspond to the high peak bit rate, which may result in lower video quality than expected.

Table 2 Statistics of coding bitrate variations

Coding Mode, #ch	AVG(kbps)	Fluctuation Range	SD(kbps)	95% CI(Mbps)
IP structure, 1ch	2681	+28% to -38%	217	2.71 to 2.65
IP structure, 4ch	11209	+3.5% to -14%	366	11.26 to 11.16
Intra slice, 1ch	2896	+1.5% to -3%	27	2.90 to 2.89
Intra slice, 4ch	10462	+3% to -8%	160	10.49 to 10.43

On the other hand, in the case of four channels, although the periodic peaks at every 0.5sec can also be observed in Figure 8, the bitrate fluctuation range becomes smaller than that of one channel as shown in the table. This would be caused by the statistical multiplexing effect for four asynchronous video channel encoding, which absorbs bit rate fluctuation of each channel to some extent. In statistical multiplexing, it distributes bitrate peaks across multiple asynchronous streams, reducing overall fluctuation.

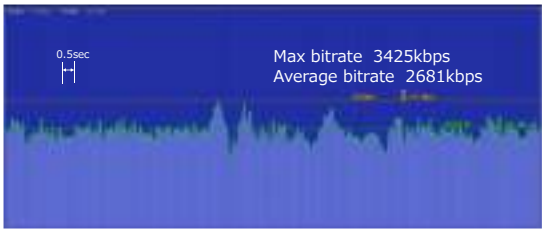


Figure 7. Bitrate variation for one-channel video encoding with IP structure (3Mbps x 1ch, Intra@0.5sec)

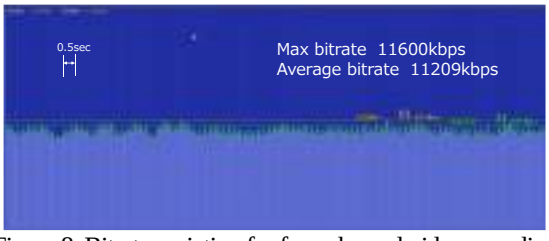


Figure 8. Bitrate variation for four-channel video encoding with IP structure (3Mbps x 4ch, Intra@0.5sec)

Figure 9 and Figure 10 show the bitrate variations for one and four channels, respectively when intra slice structure is employed. In this case, the bitrate fluctuations to the average are greatly reduced when compared with IP structure, which can also be observed in Table 2. As for the average bitrate, it was 2681kbps for the IP structure, whereas it was 2896kbps for the intra slice structure. Since the bitrate was set to 3Mbps for both structures, it can be said that the latter structure can use the available bandwidth more efficiently than the former structure.

When the amount of the bitrate fluctuation of four channels and one channel in intra slice coding is compared, the former is larger than the latter. Since the fluctuation range for each channel is very small due to intra slice coding, each small amount of fluctuation may accumulate when four channel bit streams are integrated.

As for the codec (from the encoder input to the decoded monitor) latency, 114msec delay was observed for the IP structure, whereas 66msec delay was obtained for the intra slice structure. This result means that the proposed intra slice structure can achieve both ultra-low latency while strictly keeping the coding bitrate and making the most of available bandwidth.



Figure 9. Bitrate variation for one-channel video encoding with intra-slice structure (3Mbps x 1ch)

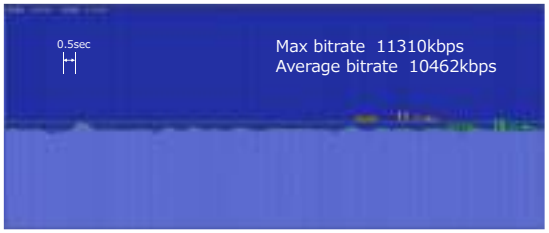


Figure 10. Bitrate variation for four-channel video encoding with intra-slice structure (3Mbps x 4ch)

5 Video transmission experiment

5.1 Experimental setup

We have conducted an end-to-end video transmission experiment using video coding setup with intra slice structure described in Section 4. The experiment aimed to evaluate transmission performance across various wireless networks using a remote construction machine setup, as shown in Figure 11. Four video cameras (OMRON SENTECH video cameras (STC-HD853HDMI) for full HD@60fps) and the newly

developed encoder (LVRC-4000) were installed on the construction machine (Hitachi Construction Machinery ZX-120).

Regarding the wireless networks, we used the following commercially available networks to evaluate various types of wireless environments.

1) Wireless for robots (5.7GHz, Doodle Labs RM-5700, point to point, Antenna 2.4m high)
2) Mesh Wi-Fi (Rajant ES1-2450R@const. machine, FE1-2450x2+ES1-2450R@AP)
3) Public LTE (au, Hytec Inter HW5G-3200x2)
4) LEO satellite network (Starlink)
5) Private 5G (Fujitsu@AP, Hytec Inter HW5G-3200 x 2)

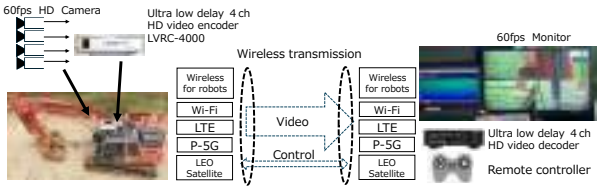


Figure 11. Experimental setup

As for 1) wireless for robots, it became available under license for image communications of unmanned systems such as robots and drones at 5.7GHz and 2.4GHz in Japan in 2016 [11]. As for 4) LEO satellite network, considering the situation of public mobile and optical fiber network failure due to a large natural disaster, we had an experiment connecting the construction machine to the LEO satellite network via mesh Wi-Fi, and video data was received at the remote control center via the Internet and public mobile and optical fiber network.

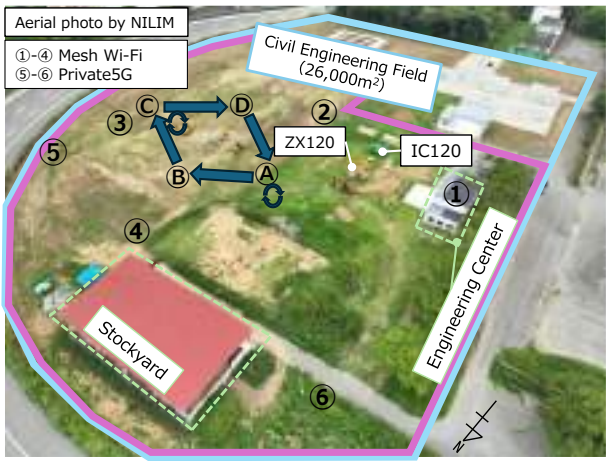


Figure 12. Experimental field

Figure 12 shows an experimental field. Four video cameras, video encoder, and network equipment were installed on the construction machine (ZX120). A 60fps video monitor (Iiyama XUB2390HS-B5), a video decoder (Hytec Inter LLC-4000), a remote controller (Hitachi S-RS-A), and network equipment are installed in the engineering center. In the experimental field, the construction machine moved around from point A to point D via points B and C, and rotated at points A and C.

Transmission experiments were conducted using an open platform called OPERA (Open Platform for Earthwork with Robotics and Autonomy) [12] facilitated for the efficient development and easy dissemination of autonomous construction technology. Table 3 shows the basic transmission performance for various wireless networks measured for 3min. For example, wireless for robots achieved 27Mbps uplink and 26Mbps downlink throughput with stable video transmission for both one channel and four channels. Similar throughputs are also obtained in public LTE and private 5G. Mesh Wi-Fi achieved 50Mbps for both uplink and downlink whereas LEO achieved 15Mbps level throughput.

In addition, the required bandwidth for the control signal was very small(<1kbps). Therefore, the control signal was transmitted in the same wireless network as that of the video data in the following experiments.

Figure 13 shows an example of one channel video transmission over the LEO satellite network at 3Mbps. As can be seen in the figure, stable video coding and transmission were realized through effective bitrate control.

Table 3. Throughputs, video coding bitrate and control signal bitrate for each network

System config.	Stay/ Move	Trhoughput Mbps		Coding bitrate per ch	Coding bitrate per 4ch	Datarate for control	Total bitrate
		To BS	From BS				
Wireless for robots	Move	27	26	3Mbps	12Mbps	Under 1kbs	Under 15Mbps
Mesh Wi-Fi	Move	52	50	3Mbps	12Mbps	Under 1kbs	Under 15Mbps
Public LTE	Move	30	26	3Mbps	12Mbps	Under 1kbs	Under 15Mbps
Private 5G	Stay	26	26	3Mbps	12Mbps	Under 1kbs	Under 15Mbps
LEO sat + fiber	Stay	15	14	1Mbps	4Mbps	Under 1kbs	Up to 6Mbps

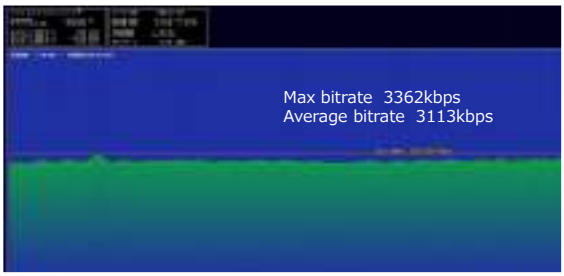


Figure 13. Transmission example of one channel video at 3Mbps over LEO satellite network

5.2 Network characteristics when the construction machine was stationary

While stationary, the networks achieved stable throughput and low packet loss, with latency varying by network type. Wireless for robots and mesh Wi-Fi

networks demonstrated minimal delay, while public LTE and private 5G showed slightly higher average delays. Table 4 shows the transmission performance when the construction machine was stationary. Due to functional limitation of measurement equipment, detailed statics such as standard deviation of jitter were not obtained.

Since less than 15Mbps throughput is obtained in LEO network, we set 1Mbps per channel for this network. Similarly, we set 3Mbps per channel for the other networks since more than 25Mbps throughputs are obtained.

As for the network latency, 4-5msec average delay was observed in the wireless for robots and mesh Wi-Fi networks, whereas there was 30-40msec average delay for the public LTE and private 5G networks. In the case of the LEO satellite network, there were 84msec average delay via the public LTE and 60msec delay via the optical fiber networks, which are much smaller latency when compared with that of the conventional GEO satellite networks.

As for the delay jitter, due to network configuration limitation for the measuring equipment, we could only have results on wireless for robots and mech Wi-Fi. Although a small jitter of +/- 7msec was observed in the former, about 50msec jitter was obtained in the latter. Although packet loss rate was only available for the wireless for robots, there was no packet losses in the experiment.

Table 4. Throughputs, packet loss, network delay, and jitter for each network when move or stay still

System config.	Stay/ Move	Throughput(Mbp		Packet loss(%)	Delay(ms)			Jitter(ms)	
		To BS	From BS		Min	Avg	Max	Min	Max
Wireless for robots	Stay	33	28	0%	2.77	3.87	8.25	-6.02	6.98
same as above	Move	27	26	0.80%	2.86	24.34	401.8	-448.5	283.7
Mesh Wi-Fi	Stay	39	44	-	2.96	5.40	30.00	-52.66	52.81
same as above	Move	52	50	-	3.47	11.82	142	-274.0	275.1
Public LTE	Stay	29	27	-	22	33	54	-	-
same as above	Move	30	26	-	22	46	175	-	-
Private 5G	Stay	26	26	-	31	44	56	-	-
LEO sat + LTE	Stay	9	10	-	55	84	203	-	-
LEO sat + fiber NW	Stay	15	14	-	33	60	290	-	-

5.3 Network characteristics when the construction machine was in motion

Table 4 also shows the transmission performance when the construction machine was moving and rotating. In this case, we compared the results of wireless for robots, mesh Wi-Fi and public LTE with stationary situations. While the throughput remained stable during movement, delays increased significantly. For instance, the wireless for robots network experienced an increase from 4msec to 24msec average delay, with jitter rising to 300-450msec due to such motion as rotation of the construction machine. On the contrary, in the case of mech Wi-Fi, although the min and max delay jitter was about 300msec level, the average delay was about 12msec due to the use of multiple routing even when the construction machine was rotating.

5.4 End-to-End delay from video camera to monitor

Table 5 and Table 6 show the end-to-end delay from video camera input to monitor output when ARQ is off and on, respectively. As can be seen in Table 3, the end-to-end delay was 115-163msec and very low delay transmission was realized. When the ARQ was on (installed in the video codec), there was an additional 100msec for this function and the total delay of less than 300msec was achieved as shown in Table 6, which satisfies the delay requirement of efficient remote construction [2].

Table 5. End-to-End delay when ARQ is off

Network	Network delay(ms)	Delay expt. NW(ms)	E-E delay(ms)	Bandwidth(Mbps)
Wireles for robot	24	103	127	27
Mesh Wi-Fi	12	103	115	52
Public LTE	46	103	149	30
Private 5G	44	103	147	26
LEO + fiber	60	103	163	15

Table 6. End-to-End delay when ARQ is on

Network	Network delay(ms)	Delay expt. NW (ms)	E-E delay(ms)	Bandwidth(Mbps)
Wireles for robot	24	201	225	27
Mesh Wi-Fi	12	201	213	52
Public LTE	46	201	247	30
Private 5G	44	201	245	26
LEO + fiber	60	201	261	15

5.5 Improvement of transmission video quality when ARQ was on

When the remote operator rotated the construction machine, there was a network disconnection between the base stations and the construction machine when wireless for robots or private 5G was used due to the limited number of base stations (one or two). In the case of ARQ=off, picture quality degradation and/or frame drop were observed due to packet loss. When the mesh Wi-Fi or the public LTE was used, the frequency of instantaneous interruption was low since the possibilities of either one of the base stations within the line of sight was high. Especially in the case of the public LTE, very stable connection was observed. However, it should be noted that the situation would vary depending on the area coverage or network congestion.



(a) ARQ is off (b) ARQ is on

Figure 14. Comparison of picture quality when ARQ is off(left) and on(right)

When ARQ was on, stable video transmission was realized at every wireless network. For example, in

Figure 14(a), the shape of the building in the right-hand side was degraded due to packet loss when ARQ was off. On the contrary, when ARQ was on, lost packet was successfully retransmitted as shown in Figure 15 and error-free decoding was realized as shown in Figure 14(b).

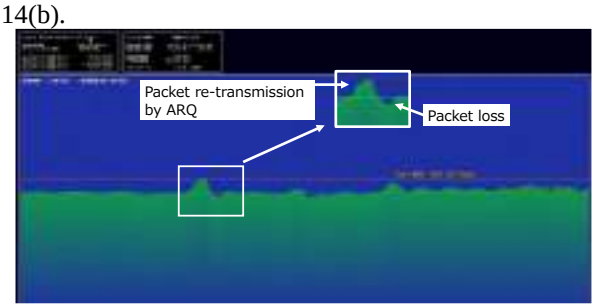


Figure 15. Packet loss and recovery via ARQ

6 Conclusions

This study focused on the development and evaluation of a multi-channel ultra-low delay video encoder to enhance remote construction efficiency. Network performance and video quality were assessed across five broadband wireless networks, including wireless for robots, mesh Wi-Fi, public LTE, private 5G, and LEO satellite. Enabling ARQ increased end-to-end delay by 100msec but significantly enhanced video quality, particularly during high-motion events such as machine rotation.

Throughout the experiments, less than 300msec of total delay was realized by the proposed system using commercially available wireless networks, meeting the delay requirements for efficient remote construction operations. While the system demonstrated high performance, simplifying the setup process remains a challenge. To support adoption in diverse construction environments, future work should focus on developing user-friendly interfaces and automated configuration tools that minimize the need for technical expertise particularly for operators with limited technical experience.

Acknowledgments

This study was carried out through the joint research work with Public Works Research Institute, Jitsuta Chugoku Co. Ltd., and Chudenko Corporation on productivity improvement of mechanical earthwork using OPERA (Open Platform for Earthwork with Robotics and Autonomy). The authors would like to thank all for discussions and support of the experiments.

The authors are also grateful to National Institute for Land and Infrastructure Management (NLIM) for the use of private 5G system.

References

[1] Japanese Infrastructure Ministry Unveils Action Plan “ i-Construction 2.0 ” to Achieve Construction Site Automation by 2040. On-line: https://www.decn.co.jp/?p=163015, Accessed: 31/03/2025.

[2] Ito S., Fujino K. and Ando H. Evaluation of work efficiency in the unmanned construction with communication delays. *The 16th Symposium on Construction Robotics in Japan*，No.06-1，2016.

[3] Miyasato T. and Kishino F. Subjective evaluation of perception of delay time between visual information and tactile information. *The Journal of the Institute of Television Engineers of Japan*, Vol.49，No.10，pp.1353-1356，1995.

[4] Proposal for Supplement 4 to the 04 series of amendments to Regulation No. 46 (Devices for indirect vision) 2016. On-line: https://unece.org/fileadmin/DAM/trans/doc/2016/wp29/ECE-TRANS-WP29-2016-089e.pdf, Accessed 31/03/2025.

[5] Sullivan G., et al. Overview of the High Efficiency Video Coding (HEVC) standard. *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649-1668, 2012.

[6] Si J., Ma S. and Gao W. Efficient bit allocation and CTU level rate control for High Efficiency Video Coding. *2013 Picture Coding Symposium (PCS)*, pp. 89-92, 2013.

[7] Starlink specifications. On-line: https://www.starlink.com/, Accessed:08/04/2025.

[8] Chayama K., Fujioka A., Kawashima K., Yamamoto H., Nitta Y., Ueki C., Yamashita A., and Asama H. Technology of Unmanned Construction System in Japan. *Journal of Robotics and Mechanics*, Vo.26 NO.4: 403-417, 2014.

[9] Xiang Y., Xu B., Su T., Brach C., Mao S.S., and Geimer M., 5G Meets Construction Machines: Towards a Smart Working Site. *2021 International Conference on Computing and Communications Applications and Technologies (I3CAT)*, 28-35, 2021.

[10] LLC-4000 -4K Video Encoder/Decoder optimized for 5G/LTE. On-line: https://hytec.co.jp/eng/prod/16271.html, Accessed: 31/03/2025.

[11] License-required Radio Equipment for Unmanned Aircraft (UA). On-line: https://www.tele.soumu.go.jp/e/sys/others/drone/, Accessed: 31/03/2025.

[12] Yamauchi G. et.al. Proposal of an Open Platform for Autonomous Construction Machinery Development. *Proc. of the 40th ISARC*, 2023.

Multi-dimensional Mapping of Confined Areas using a Hexapod Robot with Integrated Sensor Data and SLAM

Zhong Wang[1], Qipei Mei[1], Gaang Lee[1], Thomas Bock[2], and Vicente A. González[1]

[1]Infrastructure Human Tech Lab, Department of Civil and Environmental Engineering, Faculty of Engineering, Donadeo Innovation Centre for Engineering, University of Alberta, Edmonton, Alberta, Canada
[2]Chair of Building Realization and Robotics, Technical University of Munich, Munich, Germany
zhong15@ualberta.ca, qipei@ualberta.ca, gaang@ualberta.ca, bockrobotics@web.de , vagonzal@ualberta.ca

Abstract –

Mobile robots are increasingly used to explore and inspect confined environments, often in situations that are hazardous or inaccessible to humans, highlighting the need for advanced mapping and sensing capabilities. This paper presents a novel approach for constructing a multi-dimensional map by integrating sensor data from a hexapod robot with Simultaneous Localization and Mapping (SLAM). The robot, equipped with a DHT22 sensor and powered by a Raspberry Pi 4 Model B, was deployed in a mechanical room at the University of Alberta to collect humidity and temperature data while simultaneously mapping the environment. The Cartographer SLAM algorithm was used for mapping, and the sensor data was fused with the generated map by the Inertial Measurement Unit (IMU) using a bilinear interpolation algorithm. This pilot experiment serves as a demonstration of the proposed robot system, which results in a multi-dimensional map that combines 2D geographical map with sensory maps such as humidity and temperature. The map provides a visualization of the spatial distribution of environmental variables within the confined area. This approach has potential applications in various scenarios, including quick mapping of hazardous areas and routine inspections of areas with limited access. Future work will focus on incorporating 3D SLAM and exploring the use of machine learning techniques for automated anomaly detection within the multi-dimensional map, while addressing the current limitations related to real-time processing and visualization.

Keywords –

Mobile Robotics, Simultaneous Localization and Mapping (SLAM), Sensor, Inspection

1 Introduction

Mobile robots play an increasingly important role in exploring and inspecting unstructured and confined environments, particularly those deemed hazardous or inaccessible to humans [1]. A key capability for such robots is the ability to generate accurate maps of the environment while simultaneously localizing themselves within that map, even the environment is deviated from its original plan [2]. This is commonly achieved through Simultaneous Localization and Mapping (SLAM) techniques [3]. It leverages sensor data from LiDAR and cameras to enable robots to autonomously explore unknown environments and create detailed maps for navigation and task execution [3].

However, conventional SLAM algorithms predominantly focus on geometric mapping, primarily capturing the spatial layout and structural features of the environment [4]. While crucial for navigation and obstacle avoidance, these geometric maps often lack information about other critical environmental variables that can significantly impact a robot's operation or the understanding of the surrounding space [5]. Monitoring factors such as temperature, humidity, air quality, and the presence of hazardous substances is essential for robots deployed in applications such as environmental monitoring [6], industrial inspection [7], and disaster response [8]. Moreover, the map integrated with multi-dimensional layers can also enhance the situational awareness of decision makers [9]. In summary, a crucial research gap exists in multi-dimensional mapping with SLAM: the existing techniques primarily focus on geometric mapping, neglecting the integration of environmental variables crucial for understanding and operating in complex confined areas.

Researchers have explored several detection and mapping methodologies for construction environments. Borrmann, et al. [10] introduced a system utilizing a mobile robot equipped with 3D laser scanners, RGB

cameras, and thermal cameras to generate 3D thermal maps of buildings, designed to identify heat leaks and infrastructure anomalies. Amigoni and Caglioti [11] focused on efficient exploration by developing an information-based strategy for mobile robot mapping, employing laser range scanners and balancing travel distance with information gain. Although these studies effectively demonstrate the utility of mobile robots for inspection and mapping in confined spaces, they often prioritize either geometric reconstruction or specialized sensor-based detection, leading to a lack of integrated, multi-sensory map outputs. This research seeks to bridge this gap by integrating geometric and sensory mapping techniques, advancing mobile robot functionality beyond standard SLAM and detection to encompass comprehensive SLAM and sensory mapping, providing a richer environmental representation.

To fill the gap, this research investigates a method to integrate sensor data with SLAM to create a multi-dimensional map that provides a richer and more comprehensive understanding of the environment. By fusing data from environmental sensors with the spatial information obtained through SLAM, a more holistic representation of the confined area can be achieved. This multi-dimensional map not only captures the geometric features but also incorporates crucial environmental variables, enabling a more informed analysis of the surrounding space.

This research investigates the use of a hexapod robot equipped with a DHT22 sensor to create a multi-dimensional map of temperature and humidity within the mechanical room of the Li Ka Shing building at the University of Alberta. This mechanical room serves multiple key virology laboratories and requires a high level of environmental monitoring. In this pilot experiment, the goal was to investigate the robot system's capabilities to constructing a multi-dimensional map. By integrating Inertial Measurement Unit (IMU) information, the sensor data was then fused with the 2D map generated by the SLAM algorithm to create a multi-dimensional map with sensor data, enabling the visualization of the spatial distribution of temperature and humidity within the confined space.

This paper details the methods used for sensor integration, data acquisition, and map generation, and discusses the challenges encountered and potential applications of this approach. The findings contribute to the development of more comprehensive mapping techniques for robots operating in complex and confined environments, ultimately enhancing their capabilities in various real-world scenarios. The paper contributes to overcome the barriers of adopting sensors in construction industry, specifically in Multi-sensor Data Fusion and Decision Support Integration [12].

2 Robotic System Design

The high-level robotic system design employed in this research, illustrated in Figure 1, adopts a client-server architecture to facilitate both manual and semi-autonomous operation. This architecture allows for a modular and scalable approach, where the robot platform itself functions as the server, managing onboard processes such as data acquisition from the LiDAR and DHT22 sensors, as well as executing the Cartographer SLAM algorithm. The server also handles the alignment of sensor data with the SLAM-generated map using the IMU. The client-side, typically a separate computer, focuses on real-time visualization of the robot's progress and the construction of the final multi-dimensional map. This division of tasks between server and client enhances the efficiency of the system, particularly given the computational constraints of the Raspberry Pi used on the robot platform. Communication between the server and client is established wirelessly via TCP/IP over Wi-Fi, enabling remote control and monitoring of the robot's operations.

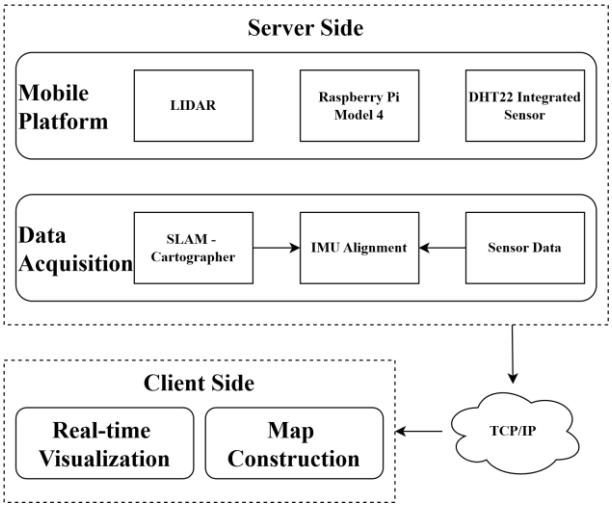


Figure 1 High-level Robotic System Design

2.1 Robot Platform

The robot prototype used in this research is a custom-assembled hexapod robot based on a Raspberry Pi 4 Model B. The Raspberry Pi serves as the central processing unit, responsible for controlling the robot's locomotion, sensor data acquisition, and communication. The robot's design is based on the RHex [13]. The choice of a hexapod robot for this research was motivated by several factors related to the challenges of navigating and collecting data within confined areas [3]. Hexapod robots offer superior mobility compared to wheeled or tracked robots, enabling them to traverse uneven terrain, overcome obstacles, and maintain stability in confined spaces. Their adaptability to different locomotion gaits

allows them to maneuver effectively in tight spaces and adjust to varying environmental conditions [14]. Additionally, the hexapod platform's inherent stability ensures that the robot can maintain its balance while carrying the necessary sensors and computing equipment, crucial for acquiring reliable data in potentially unstable environments [15]. The combination of mobility, adaptability, and stability makes the hexapod robot an ideal choice for this research.

2.2 Sensor Integration

The robot prototype was equipped with a DHT22 sensor to measure temperature and humidity. This sensor was chosen for its affordability, ease of use, and compatibility with the Raspberry Pi [16]. The DHT22 sensor was connected to the Raspberry Pi's GPIO (General Purpose Input/Output) pins, and the sensor data was read using the Adafruit_DHT library. This library provides functions for interacting with the DHT22 sensor and retrieving temperature and humidity readings [17]. Incorporating sensors with single sensing principle, such as temperature and humidity sensors, in this initial prototype is beneficial as it provides a foundational framework for integrating environmental data with SLAM, enabling the robot to assess basic environmental conditions and serving as a stepping stone towards more complex sensing capabilities for applications such as detecting thermal anomalies in industrial settings or identifying potential gas leaks [1].

2.3 SLAM Implementation

The Cartographer SLAM algorithm was employed for mapping the environment, which is visualized in real-time using Rviz, as shown in Figure 2. Cartographer is an open-source SLAM library developed by Google, known for its accuracy and robustness in various environments. It utilizes a combination of laser range finder data and inertial sensor data to create a 2D map of the surroundings [18]. In this research, the robot's movement was controlled manually, and the Cartographer SLAM algorithm generated a map of the confined area based on the robot's trajectory and sensor readings.

In order to provide the Cartographer SLAM algorithm with the necessary spatial information, a LiDAR (Light Detection and Ranging) sensor was integrated into the robot's platform. The LiDAR sensor emits laser beams and measures the time it takes for the beams to reflect back from surrounding objects, providing accurate distance measurements [19]. This data was then used by the Cartographer SLAM algorithm to construct a detailed map of the environment, including walls, obstacles, and other features. The integration of the LiDAR sensor enhances the accuracy and efficiency of the mapping process, enabling the robot to generate a comprehensive representation of the confined area [19].

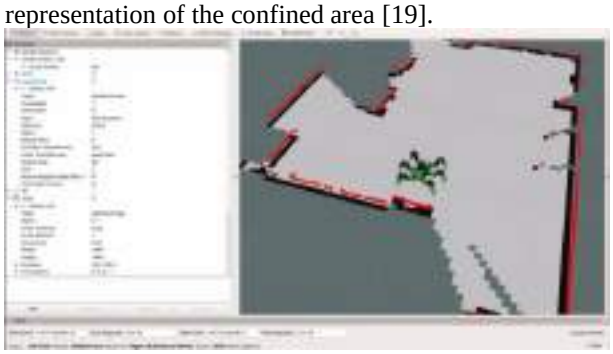


Figure 2 RViz Visualization of SLAM

2.4 Control and Communication

The robot's control and communication systems were designed to facilitate both manual and semi-autonomous operations. A client-server architecture was implemented, allowing for remote control and monitoring of the robot's activities. The server-side program, running on the Raspberry Pi, managed sensor data acquisition, SLAM processing, and communication with the client. The client-side program, running on a separate computer, provided a graphical user interface (GUI) for controlling the robot's movements, visualizing sensor data, and monitoring the mapping process.

Communication between the client and server was established over a Wi-Fi network using the TCP/IP protocol. This wireless communication enabled real-time control of the robot and allowed the operator to monitor the robot's progress and sensor readings from a safe distance. The client-server architecture provided flexibility in controlling the robot and facilitated the integration of additional sensors or functionalities in the future.

2.5 Data Acquisition

The robot prototype was deployed in a section of mechanical room of the Li Ka Shing building at the University of Alberta, as shown in Figure 3. The mechanical room served as a representative confined area, characterized by limited space, obstacles, and varying environmental conditions. In this pilot experiment, the goal was to investigate the robot system’s capabilities for constructing the multi-dimensional map. The robot was manually controlled to traverse the room, and the DHT22 sensor collected temperature and humidity data at one-second intervals. This sampling rate was chosen based on the sensors response frequency (0.5 Hz or once every two seconds) and the robot’s stabilization time (one second). The readings will be normalized to ensure the accuracy of values in Section 2.6. The Cartographer SLAM algorithm generated a 2D map of the room based on the robot's movement, and the IMU integrated with sensor

readings generated another two 2D maps for temperature and humidity.



Figure 3 Mechanical Room for the Experiment

To enhance the visualization and analysis of the acquired sensor data, a real-time bilinear interpolation algorithm was employed [20]. This technique estimates values at intermediate points within the 2D map based on the known sensor readings at surrounding locations [21]. By interpolating the temperature and humidity data, a continuous heatmap representation can be generated, providing a more comprehensive view of the environmental conditions within the confined area. Bilinear interpolation is computationally efficient and offers a reasonable approximation of the spatial distribution of sensor data, making it suitable for this application [20].

The bilinear interpolation is calculated as follows:

Let Q_{11}, Q_{12}, Q_{21}, and Q_{21} represent the temperature or humidity values at the four points of a grid cell and let $P_{(x,y)}$ be the point where the interpolated value is needed.

Interpolate in the x-direction:

$$R1 = \frac{(x_2 - x)}{(x_2 - x_1)} * Q_{11} + \frac{(x - x_1)}{(x_2 - x_1)} * Q_{21}$$

$$R2 = \frac{(x_2 - x)}{(x_2 - x_1)} * Q_{12} + \frac{(x - x_1)}{(x_2 - x_1)} * Q_{22}$$

Interpolate in the y-direction:

$$p = \frac{(y_2 - y)}{(y_2 - y_1)} * R_1 + \frac{(y - y_1)}{(y_2 - y_1)} * R_2$$

In these equations, Q represents the value of the environmental variable being interpolated using single sensing principles, whether it is temperature, humidity, or any other measured quantity. For points Q11, Q12, and Q22 that lie on the robot's trajectory, their temperature and humidity values can be directly measured by the onboard sensors. However, for the point Q21 located outside the trajectory, the value was estimated by interpolation based on the known values of its three closest neighbours (Q11, Q12, Q22) in the corresponding points in the adjacent grid cell. This approach leverages the spatial correlation between neighbouring points to provide a reasonable approximation of the environmental conditions at Q21. This generalized notation emphasizes the versatility of bilinear interpolation, as the same mathematical framework can be applied to various environmental parameters.

It is important to acknowledge that assuming a linear distribution for temperature and humidity may not always be accurate, especially in environments with complex variations. More sophisticated interpolation techniques, such as kriging [22] or Gaussian process regression [23], could be explored to better capture the spatial distribution of these variables. These methods can account for spatial correlations and provide more accurate predictions, but they also require more computational resources and may not be suitable for real-time applications [24]. The choice of interpolation method depends on the specific application and the desired balance between accuracy and computational efficiency [24].

2.6 Map Construction

The sensor data acquired by the DHT22 sensor was integrated with the 2D map generated by the SLAM algorithm to create a multi-dimensional map. This was achieved using Matplotlib, a Python plotting library [25, 26]. The temperature and humidity data were interpolated across the 2D map using bilinear interpolation, creating a heatmap representation of the environmental variables, as shown in Figure 4 and Figure 5. The resulting multi-dimensional map provides a visual representation of the spatial distribution of temperature and humidity within the mechanical room.

To ensure accurate alignment between the 2D map generated by SLAM and the heatmap created by the DHT22 sensor, the robot's MPU6050 IMU was utilized. This IMU provides orientation and movement data, enabling precise mapping of the sensor readings onto the 2D map. Specifically, the IMU's accelerometer and gyroscope readings are fused to obtain the robot's gestures and movement, which are then used to transform the sensor data from the robot's local coordinate frame to the global coordinate frame of the SLAM map. Given the temperature and humidity readings happen at one-second intervals, the IMU readings within a second are merged as one data point which aligns to the sensor readings. This alignment process ensures that the temperature and humidity readings are accurately represented in the multi-dimensional map, facilitating a more reliable analysis of the environmental conditions within the confined area. During the plotting stage, those coordinates from IMU readings are connected as a line, representing the path of navigation as shown in Figure **4** and Figure **5**.

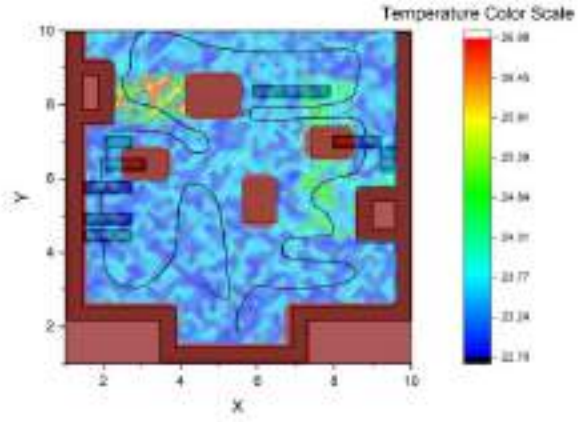


Figure 4 Temperature Heatmap

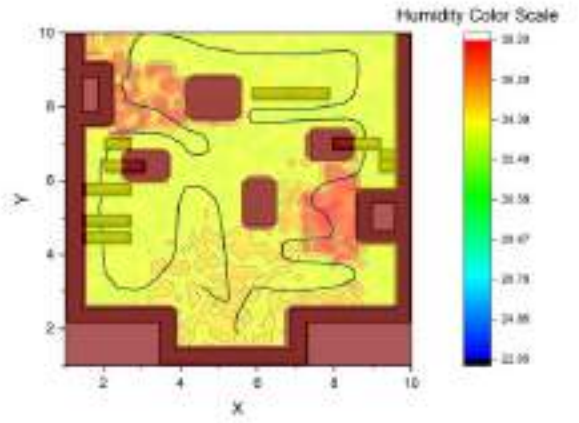


Figure 5 Humidity Heatmap

Due to the limited computational power of the Raspberry Pi, the map construction process for the sensing maps was conducted after the mapping process. The 2D map constructed by SLAM with IMU is used to create a grid for the sensing map. To perform bilinear interpolation and estimate the value at a point $P_{(x,y)}$ on the map, the four nearest grid points, denoted as Q_{11}, Q_{12}, Q_{21}, and Q_{21}, need to be identified. This involves finding the nearest point $P'_{(x\prime,y\prime)}$ on the trajectory curve to $P_{(x,y)}$.

Given the irregular shape of the curve, a piecewise cubic spline interpolation is employed to represent the trajectory. Let $S_{(t)}$ be the cubic spline function that interpolates the robot's path. The nearest point $P'_{(x\prime,y\prime)}$ on the curve to $P_{(x,y)}$ can be found by minimizing the distance:

$$P_{(x,y)} = \min\left(\sqrt{((x - S_x_{(t)})^2 + ((y - S_y_{(t)})^2}\right)$$

where $S_x_{(t)}$ and $S_y_{(t)}$ are the x and y components of the spline function $S_{(t)}$.

A local coordinate system is then established at P' using the tangent vector and its normal vector at P'. The tangent vector v at P' can be calculated as the derivative of the spline function:

$$v = \left(\frac{dS_x}{dt}, \frac{dS_y}{dt}\right) at\ t = t'$$

The normalized unit tangent vector u and unit normal vector n are then:

$$u = v/|v| = 1, n = \left(\frac{-dS_y}{dt}, \frac{dS_x}{dt}\right)/|v|$$

With a predefined grid spacing h of 0.5 meters, the four nearest grid points Q_{11}, Q_{12}, Q_{21}, and Q_{21} can be determined as:

$$Q_{11} = P' + h * u + h * n$$
$$Q_{12} = P' + h * u - h * n$$
$$Q_{21} = P' - h * u + h * n$$
$$Q_{21} = P' - h * u - h * n$$

The grid spacing h represents the resolution of the sensor data interpolation. A smaller grid spacing results in a finer-grained map with more interpolated points, potentially capturing more detailed variations in temperature and humidity [21]. However, it also increases the computational burden for interpolation. Conversely, a larger grid spacing reduces the computational load but may result in a coarser map that misses subtle environmental variations.

The choice of h = 0.5 meters was based on a balance between map resolution and computational efficiency, considering the limited processing power of the Raspberry Pi. Future work could investigate the effect of different grid spacings on map quality and optimize this parameter based on the specific environment and application requirements.

To provide a more concrete understanding of the system's performance in the bottleneck, the following metrics were collected for the map construction tasks:

Table 1 Performance Metrics

Category	Metrics
CPU Usage	95-97%
RAM Usage	68-70%
Map Generation Time	14 minutes and 37 seconds
Network Bandwidth	5 Mbps

3 Results

This pilot experiment serves as a demonstration of the prototype, which resulted in the successful generation of a multi-dimensional map of the mechanical room, integrating both geometric features and environmental variables. The Cartographer SLAM algorithm

successfully mapped the room's layout, including walls, obstacles, and open spaces. The DHT22 sensor collected temperature and humidity data throughout the robot's traversal, which was then overlaid onto the 2D map using bilinear interpolation. The IMU aligned the 2D maps constructed by SLAM and sensors together, resulting in a multi-dimensional map that effectively visualizes the spatial distribution of temperature and humidity within the confined area.

To verify and validate the accuracy of the multi-dimensional map, the robot's sensor readings were compared with manual measurements and observations of the mechanical room along the robot's path. This comparison also ensured that the readings from the DHT 22 sensor were not affected by the heat generated from the robot's motherboard or motors. The SLAM-generated map was verified to represent the room's dimensions and the location of obstacles. These verification and validation steps ensured the reliability and validity of the generated map and the associated sensor data.

The multi-dimensional map reveals that the temperature and humidity levels are not uniform throughout the mechanical room. Areas with higher concentrations of equipment tend to have slightly elevated temperatures, while humidity levels show some variation depending on the proximity to potential sources of moisture or ventilation. The map provides a comprehensive overview of the environmental conditions within the room, highlighting areas of potential concern or interest.

4 Discussion

This research demonstrates the feasibility of integrating sensor data with SLAM by IMU to create multi-dimensional maps using a hexapod robot. The approach offers several advantages, particularly for inspecting confined areas. The robot's ability to traverse complex ground conditions and collect sensor data in hard-to-reach locations makes it a valuable tool for assessing environmental conditions in challenging environments.

However, the research also highlights some challenges and limitations. The limited computational power of the Raspberry Pi poses a constraint on the complexity of the SLAM algorithms. Due to this, the mapping process took a relatively long time, and the multi-dimensional map can only be plotted after the mapping, which shows the weakness for real-time visualization. Moreover, such limited power prevented the authors to utilizing 3D SLAM. Additionally, the accuracy of the multi-dimensional map is dependent on the interpolation method used. While bilinear interpolation provides a reasonable approximation, more sophisticated techniques may be required to achieve higher accuracy, especially in environments with rapidly changing conditions. Meanwhile, the accuracy of values created from the interpolation algorithms should be systematically compare with ground truth maps to ensure the accuracy. Last but the least, the authors acknowledge the limitation that this pilot experiment only serves as a demonstration and the first field test for this robot system approach. More rigorous experimental design is needed to evaluate and improve the capabilities and applications of this system.

Despite these challenges, the potential applications of this research are significant. The ability to map confined areas and identify hazardous conditions has implications for various industries, including disaster response, environmental monitoring, and industrial inspection. Future work will focus on addressing the identified limitations and exploring the use of more advanced sensors and SLAM algorithms to further enhance the accuracy and capabilities of the multi-dimensional mapping approach.

5 Conclusion

This paper presents a method for constructing multi-dimensional maps by integrating sensor data with SLAM by IMU using a hexapod robot. The approach was successfully demonstrated in a confined area, resulting in a map that visualizes the spatial distribution of temperature and humidity. The research highlights the potential of hexapod robots for inspecting hazardous and inaccessible areas, while also identifying challenges related to computational limitations and interpolation accuracy. Future work can focus on addressing these challenges and exploring the use of more advanced sensors and 3D SLAM algorithms with better processing unit to enhance the capabilities of this approach. It is also recommended to evaluate this approach on the efficiency of decision-making.

This research sets the stage for several promising future directions. Firstly, the integration of more advanced sensors with multi-sensing principles, such as gas detectors, will enable the robot to collect a wider range of environmental data, leading to more comprehensive and informative multi-dimensional maps. Secondly, the development and utilization of more robust and computationally efficient SLAM algorithms will improve the accuracy and real-time capabilities of the mapping process, allowing for online visualization and analysis of the environment. Finally, the incorporation of machine learning techniques can automate the identification of anomalies or points of interest within the multi-dimensional map, facilitating decision-making and response in hazardous or inaccessible environments.

6 Acknowledgements

The authors would like to express their sincere acknowledgement to the Canada Research Chair Program of the Government of Canada (CRC-2021-00531).

The authors would also like to extend the acknowledgement to the following individuals for their invaluable contributions to this research:

- Darren St.Hilaire, P.Eng, CEFP. Director, Trades Infrastructure and Maintenance, University of Alberta, for providing access to the mechanical room and facilitating the data acquisition process.
- Tony Hodge, Director, Infrastructure Development, University of Alberta, for his support and guidance throughout the project.
- Neil McFarlane, Industry Liaison, Faculty of Engineering, Civil and Environmental Engineering Department, University of Alberta, for his support in establishing the collaborations.

Their expertise and assistance were instrumental in the successful completion of this research.

References

[1] S. Halder and K. Afsari, "Robots in Inspection and Monitoring of Buildings and Infrastructure: A Systematic Review," *Applied Sciences*, vol. 13, no. 4, p. 2304, 2023. [Online]. Available: https://www.mdpi.com/2076-3417/13/4/2304.

[2] A. M. Rezende *et al.*, "Indoor localization and navigation control strategies for a mobile robot designed to inspect confined environments," in *2020 IEEE 16th International Conference on Automation Science and Engineering (CASE)*, 2020: IEEE, pp. 1427-1433.

[3] H. Azpúrua *et al.*, "Towards semi-autonomous robotic inspection and mapping in confined spaces with the EspeleoRobô," *Journal of Intelligent & Robotic Systems*, vol. 101, pp. 1-27, 2021.

[4] T. Taketomi, H. Uchiyama, and S. Ikeda, "Visual SLAM algorithms: A survey from 2010 to 2016," *IPSJ transactions on computer vision and applications*, vol. 9, pp. 1-11, 2017.

[5] K. Kamarudin *et al.*, "Integrating SLAM and gas distribution mapping (SLAM-GDM) for real-time gas source localization," *Advanced Robotics*, vol. 32, no. 17, pp. 903-917, 2018/09/02 2018, doi: 10.1080/01691864.2018.1516568.

[6] M. Dunbabin and L. Marques, "Robots for environmental monitoring: Significant advancements and applications," *IEEE Robotics & Automation Magazine*, vol. 19, no. 1, pp. 24-39, 2012.

[7] S. Lu, Y. Zhang, and J. Su, "Mobile robot for power substation inspection: A survey," *IEEE/CAA Journal of Automatica Sinica*, vol. 4, no. 4, pp. 830-847, 2017.

[8] K. Nagatani *et al.*, "Emergency response to the nuclear accident at the Fukushima Daiichi Nuclear Power Plants using mobile rescue robots," *Journal of Field Robotics*, vol. 30, no. 1, pp. 44-63, 2013.

[9] H. Bavle, J. L. Sanchez-Lopez, C. Cimarelli, A. Tourani, and H. Voos, "From SLAM to Situational Awareness: Challenges and Survey," *Sensors*, vol. 23, no. 10, doi: 10.3390/s23104849.

[10] D. Borrmann *et al.*, "A mobile robot based system for fully automated thermal 3D mapping," *Advanced Engineering Informatics*, vol. 28, no. 4, pp. 425-440, 2014/10/01/ 2014, doi: https://doi.org/10.1016/j.aei.2014.06.002.

[11] F. Amigoni and V. Caglioti, "An information-based exploration strategy for environment mapping with mobile robots," *Robotics and Autonomous Systems*, vol. 58, no. 5, pp. 684-699, 2010/05/31/ 2010, doi: https://doi.org/10.1016/j.robot.2009.11.005.

[12] Z. Wang, V. A. González, Q. Mei, and G. Lee, "Sensor adoption in the construction industry: Barriers, opportunities, and strategies," *Automation in Construction*, vol. 170, p. 105937, 2025/02/01/ 2025, doi: https://doi.org/10.1016/j.autcon.2024.105937.

[13] U. Saranli, M. Buehler, and D. E. Koditschek, "RHex: A simple and highly mobile hexapod robot," *The International Journal of Robotics Research*, vol. 20, no. 7, pp. 616-631, 2001.

[14] S. Kajita and B. Espiau, "Legged robot," in *Springer handbook of robotics*: Springer Berlin/Heidelberg, Germany, 2008, pp. 361-389.

[15] X. Ding, Z. Wang, A. Rovetta, and J. Zhu, "Locomotion analysis of hexapod robot," *Climbing and walking robots*, vol. 28, pp. 291-310, 2010.

[16] Y. A. Ahmad, T. S. Gunawan, H. Mansor, B. A. Hamida, A. F. Hishamudin, and F. Arifin, "On the evaluation of DHT22 temperature sensor for IoT application," in *2021 8th international conference on computer and communication engineering (ICCCE)*, 2021: IEEE, pp. 131-134.

[17] Adafruit. "Adafruit-DHT 1.4.0." https://pypi.org/project/Adafruit-DHT/ (accessed Nov 01, 2024).

[18] W. Hess, D. Kohler, H. Rapp, and D. Andor, "Real-time loop closure in 2D LIDAR SLAM,"

in *2016 IEEE international conference on robotics and automation (ICRA)*, 2016: IEEE, pp. 1271-1278.

[19] R. Yagfarov, M. Ivanou, and I. Afanasyev, "Map comparison of lidar-based 2d slam algorithms using precise ground truth," in *2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, 2018: IEEE, pp. 1979-1983.

[20] K. T. Gribbon and D. G. Bailey, "A novel approach to real-time bilinear interpolation," in *Proceedings. DELTA 2004. Second IEEE international workshop on electronic design, test and applications*, 2004: IEEE, pp. 126-131.

[21] F. Arif and M. Akbar, "Resampling air borne sensed data using bilinear interpolation algorithm," in *IEEE International Conference on Mechatronics, 2005. ICM'05.*, 2005: IEEE, pp. 62-65.

[22] M. A. Oliver and R. Webster, "Kriging: a method of interpolation for geographical information systems," *International Journal of Geographical Information System*, vol. 4, no. 3, pp. 313-332, 1990.

[23] A. Wilson and H. Nickisch, "Kernel interpolation for scalable structured Gaussian processes (KISS-GP)," in *International conference on machine learning*, 2015: PMLR, pp. 1775-1784.

[24] J. Li and A. D. Heap, "A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors," *Ecological Informatics*, vol. 6, no. 3-4, pp. 228-241, 2011.

[25] Matplotlib. "Matplotlib: Visualization with Python." https://matplotlib.org/ (accessed Nov 01, 2024).

[26] S. Tosi, *Matplotlib for Python developers*. Packt Publishing Ltd, 2009.

Exploring Cognitive Load and Task Complexity in Dynamic Tracking Tasks: Insights for Construction Workflows

Amira Eltahan[1], Gaang Lee[1] and Farook Hamzeh[1]

[1]Department of Civil and Environmental Engineering, University of Alberta, Canada
eltahan@ualberta.ca, gaang@ualberta.ca, hamzeh@ualberta.ca

Abstract -
Construction tasks involve dynamic and visually demanding activities that require continuous monitoring, tracking, and decision-making. These demands can overwhelm workers' cognitive capacity and increase mental strain and reduce efficiency. This study combines two cognitive load assessment methods: objective cognitive load assessment using wearable eye tracking and task-based performance analysis by calculating performance error. It examines the impact of complexity in dynamic tracking tasks on performance and cognitive load. The metrics include blink rate, fixation rate, saccadic amplitude, saccadic peak velocity, saccadic mean velocity, and tracking error. Saccadic amplitude showed a strong positive correlation (r = 0.891) which reflect the need to scan broader areas for higher visual complexity. In contrast, saccadic peak velocity (r = -0.96), blink rate (r = -0.967), and fixation rate (r = -0.671) demonstrated negative correlations with task complexity, which suggests increased cognitive demands and a prioritization of accuracy over speed. Saccadic mean velocity showed minimal correlation (r = -0.07), which suggests it might not be a sensitive metric for evaluating the impact of task complexity. Performance error was measured as the Euclidean distance between gaze and target, and it revealed a strong positive correlation with task complexity (r = 0.967). It indicates reduced performance as complexity increased. These findings highlight the significant impact of complexity on cognitive and visual performance. This is particularly relevant in construction tasks that require continuous monitoring, tracking, and visual processing. Understanding these relationships helps optimize construction workflows, reducing cognitive strain, improving efficiency, and enhancing safety in visually demanding tasks.

Keywords –
Task complexity, Cognitive load, Visual processing, Eye tracking metrics, Dynamic tracking

1 Introduction

Construction tasks are inherently dynamic and visually demanding. This requires workers to continuously monitor, track, and make decisions under challenging conditions. These tasks often involve processing complex visual information while maintaining high levels of accuracy and speed. The cognitive and visual demands of such tasks can overwhelm workers, leading to mental strain, decreased performance, and an increased likelihood of errors. Cognitive load refers to the mental effort required to process information, solve problems, and make decisions [1, 2] . It is influenced by task complexity, time constraints, and environmental factors [3]. Understanding how cognitive load and task complexity impact performance is critical to optimize construction workflows and improve workers' safety and productivity.

To optimize cognitive load, we first have to measure it. Three methods can be used to measure cognitive load which are subjective, objective, and task- and performance-based techniques[4]. Subjective measurement methods rely on self-reporting tools like the NASA Task Load Index (NASA-TLX) [5], the Subjective Workload Assessment Technique (SWAT), and the Cognitive Task Load Analysis (CTLA) [5]. These approaches involve participants rating their mental workload on predefined scales. However, these methods can be limited by individual biases and interruptions to workflow which reduce their reliability and practicality. Objective measurement techniques utilize physiological indicators to assess cognitive load. It includes monitoring heart rate variability (HRV), brain activity via electroencephalogram (EEG), eye activity, cortisol levels, and skin conductance [4]. This method provides real-time, less-intrusive approach to quantify cognitive load. The task and performance-based technique is usually implemented in experimental studies to supplement wearable sensors and support the results. It assesses cognitive load by measuring task performance through metrics like reaction time, accuracy and error rate [6].

EEG is a widely used tool for assessing cognitive load in construction [7]. However, its application in real-world scenarios is constrained by challenges such as the requirement for controlled environments [8]. Eye-tracking technology is gaining traction as a practical alternative for evaluating cognitive load in construction settings [9, 10] because of its adaptability in the dynamic construction setting. Unlike EEG, its less invasive, portable, and less sensitive to artifacts from the external environment. Previous studies have established that cognitive load influences visual behavior with metrics such as fixation rate, blink rate, saccadic peak velocity, pupil dilation, and saccadic amplitude[11, 2, 12]. However, there is limited research on how these metrics interact in the context of dynamic tracking tasks coupled with varying complexities commonly encountered in construction. This study aims to fill this gap by evaluating the relationship between task complexity and cognitive load in dynamic tracking tasks using eye-tracking data. These findings will contribute to the development of strategies to reduce cognitive strain, enhance tracking performance, and optimize workflows in visually intensive construction tasks.

2 Literature Review

2.1 Cognitive Load and Its Impact on Construction

In the construction industry, cognitive load plays a significant role in determining workers' efficiency, decision-making [12] and safety [13]. A study by Kwon et al (2024) highlight the risks posed by cognitive overload in safety-critical situations as a result of multiple safety signs. The study concluded that excessive cognitive demands reduce workers' ability to process visual warnings effectively, particularly when more than nine items are displayed [14]. Mitropolus and Memartian (2013) [15] examined the dimensions of task demands in masonry work. They divided it to mental, physical, and temporal demands and studied their implications on performance and management strategies construction. The study provided evidence that high task complexity and temporal pressures are critical factors contributing to mental overload and reduced attentiveness, which can lead to errors, rework and safety incidents on construction sites. They proposed practical management strategies, such as reducing the task design complexity that creates high physical or mental demands. Memarian et al., (2012)[16] identified work practices that balance high productivity and safety while errors in construction tasks through a case study that analyzed the work practices of a high-reliability field supervisor in a concrete construction project. The findings emphasized the importance of reducing task complexity and aligning high-demand tasks with highly skilled workers. However, the correlation between task complexity in visual tracking tasks and the corresponding impact on cognitive load remains unexplored in literature. Even though several studies have highlighted the impact of task complexity on performance through subjective methods or through wearable sensors, the studies quantifying the impact of task complexity on cognitive load and performance is still lacking.

2.2 Cognitive Load Measurement in Construction

Subjective self-reporting methods have been deployed to assess mental fatigue and cognitive load in construction by Zhang et.al.(2013) [13] using Fatigue Assessment Scale for Construction Workers (FASCW). Others have used NASA Task Load Index (NASA-TLX) to assess cognitive load [14]. However, these subjective methods were proven to bias due to individual variability and recall bias due to participants' reliance on their memory to report their experiences [18]. To overcome this, researchers have shifted towards using objective methods using advanced wearable sensors to measure cognitive load. These sensors measure physiological responses due to cognitive load such as Electroencephalography (EEG), wearable eye trackers, and physiological metrics recorders to objectively measure cognitive states in construction environments [20, 21, 22, 23].

A systematic literature review was conducted by Jamil Uddin et al., (2024) [7] on using technologies to measure mental fatigue of construction workers. The authors reported that EEG was one of the commonly used tools to assess cognitive load in construction. However, EEG has a limited applicability in construction setting due to its sensitivity to the external environmental factors, such as head motions and electromagnetic interferences. Eye tracking has been emerging as an alternative practical measurement to cognitive load [20, 21]. However, research quantifying the relationship between task complexity, performance and cognitive load for dynamic tracking tasks using site-friendly wearable sensors is still lacking. Therefore, this study aims to explore and quantify the impact of task complexity on cognitive load and performance using wearable eye trackers in dynamic tracking tasks.

3 Methods

3.1 Task Design

The experiment will employ a within-subject design where participants perform three identical tasks with one controlled variable "task complexity". This controlled setup ensures that differences in the eye metrics are

highly attributable to the varying task complexity. A dynamic task was designed to evaluate participants' ability to track a moving gaze point by displaying a dot that follows a predefined path at three varying speeds. This is particularly important in construction, where workers need to track moving objects, process dynamic visual information, and maintain situational awareness. Deficiencies in situational awareness are a major contributor to workplace accidents, as shown by 28,512 lost-time injuries reported in 2023 [23]with majority of incidents resulting from workers being struck by or caught in objects and equipment**.** These accidents often stem from difficulties in effectively tracking moving objects. For instance, tasks such as crane operations, vehicle maneuvering, and heavy equipment handling require precise visual tracking of moving targets to ensure safety and efficiency. This study aims to capture fundamental visual-cognitive processes that underlie real-world construction tasks by isolating tracking ability in a controlled experiment. The results can inform strategies for reducing cognitive overload and improving workers' ability to maintain focus in complex environments. Level 1 represents the baseline condition and Levels 2 and 3 indicate increasing levels of complexity. The tasks were presented in a randomized order to minimize bias from learning effects. A black background was used as it provides higher contrast with a brightly coloured moving dot (bright green). Green is considered less harsh on the eyes compared to red or blue, making it a good choice for tasks that require sustained tracking.

3.2 Data Collection

The study employs objective and task and performance-based measurement methods of cognitive load. Five metrics were collected which include blink rate (blinks/sec), fixations rate (fixations/second), saccadic amplitude (px), saccadic peak velocity (px/sec), and saccadic mean velocity (px/sec). Also, gaze coordinates were collected to calculate the average distance between the participants' gaze points and the actual target coordinates by calculating Euclidean distance. All metrics were recorded continuously during each trial using Pupil Invisible Eye Tracking Glasses by Pupil Labs. The eye tracking glasses collects gaze data at 200 HZ, fixations, saccades, blinks, RGB scene video at 30 Hz using a detachable scene camera and gaze mapping. Pearson correlation coefficient was calculated to assess the correlation between task complexity and the metrics.

A total of seven graduate students voluntarily agreed to participate after signing informed consent forms with the option to withdraw two weeks after data collection without consequences. Prior to the experiment, each participant was briefed on the objectives and procedures of the study and their questions or concerns were addressed to ensure clarity and comfort. The participants were asked to as accurately as possible follow the moving object. The sample size aligns with similar studies using physiological metrics [13]. In eye-tracking studies, similar sample sizes have been commonly used due to the exhaustive nature of data collection and analysis, as well as the high accuracy of the observed trends [10].

3.3 Data Analysis

To measure the impact of cognitive load on eye movement, five different metrics were calculated. These metrics include saccadic peak velocity (px/s), saccadic amplitude (px), saccadic mean velocity (px/s), blink rate (blink/second), and fixation rate (fixation/second). The metrics were measured for the three different task complexities for each participant. This enabled the comparison of each participant's metrics that reflect his corresponding cognitive load given the varying task complexity. Also, the performance of each participant was calculated by the average Euclidean distance between the gaze and target points for the three complexities. Furthermore, to understand the correlation between the metrics and task complexity, Pearson correlation coefficient was calculated.

4 Results and Discussion

4.1 Eye tracking metrics analysis

4.1.1 Saccadic Peak Velocity (px/s)

Saccadic peak velocity is an essential metric in eye-tracking research as it represents the speed and effectiveness of eye movements during saccades. It is the highest speed achieved by the eye during a saccadic movement. Figure 1 shows the results across all subjects in three tasks of varying complexity. The figure shows that saccadic peak velocity decreases systematically as task complexity increases for all participants. The baseline condition (Level 1) exhibits the highest saccadic velocities which indicates minimal cognitive strain and higher efficiency in executing saccades. Our results reveal a strong negative correlation between task complexity and saccadic peak velocity with an average Pearson correlation coefficient of -0.96. This indicates an inverse relationship between task complexity saccadic peak velocities. It could possibly be because participants prioritize accuracy over speed when visual demand increase.

These findings aligns with existing research which highlighted that saccadic peak velocity decrease as mental load and task complexity increase[11, 24]. And have been reported as an effective measurement for mental load variations in complex tasks [26, 27].

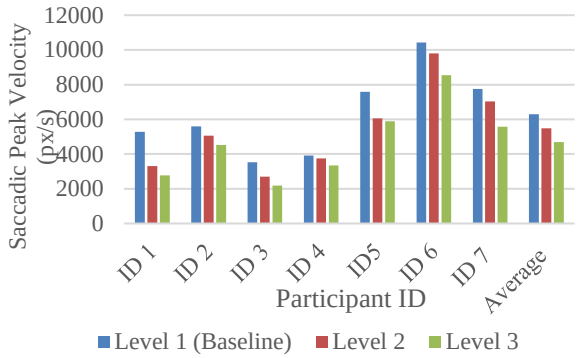


Figure 1. Saccadic Peak Velocity Across Task Complexity Levels

It should be noted that there is variability in saccadic peak velocity across participants. For example, ID 6 consistently exhibits the highest velocities across all levels, while IDs 3 and 4 have relatively lower values. This variability could be attributed to differences in individual cognitive capacity, visual acuity, or familiarity with the task.

4.1.2 Saccadic Amplitude (px)

Saccadic amplitude is the distance the eye travels during a single saccadic movement. Figure 2 presents saccadic amplitude (in pixels) across three levels of task complexity for seven participants (ID 1 to ID 7). The average amplitude is calculated for each level to provide insights into general trends. The average saccadic amplitude increased as task complexity increased. This indicates that participants made larger eye movements as task complexity increased this could be due the need to scan broader areas or manage more challenging visual demands. There is significant variability among participants. ID 6 consistently exhibits the largest amplitudes across all levels, with a sharp increase from 97.73 px (Level 1) to 169.45 px (Level 3). ID 1 and ID 7 show smaller amplitudes compared to others, with more modest increases between levels. The variability suggests individual differences in visual search strategies, task adaptation, or cognitive effort.

It can be observed that saccadic peak velocity decreased with task complexity while saccadic amplitude increased. However, saccadic peak velocity has been sought to typically increase with the amplitude of eye movements, a relationship known as the main sequence[27]. This relationship is generally considered fixed but can be modulated by external factors such as incentives and motivation[28]. If a task involves scanning a visually complex scene under cognitive strain, participants might make longer saccades to cover larger areas while slowing their saccadic velocity to ensure accuracy[29].

Also, our results show a strong positive correlation between task complexity and saccadic amplitude with an average Pearson correlation coefficient of 0.896. This indicates that saccadic amplitude increases significantly as task complexity increases. This results in longer saccades which lead to an increase in saccadic amplitude. Under cognitive strain, the brain often prioritizes accuracy over speed. This trade-off can cause a reduction in saccadic peak velocity, even as the amplitude increases as reported in our findings. This behaviour aligns with findings that saccadic metrics, such as peak velocity, can vary under cognitive load, while amplitude adapts to task demands[23].

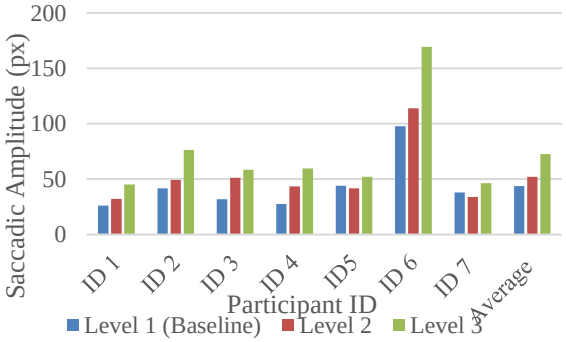


Figure 2. Saccadic Amplitude Across Task Complexity Levels

4.1.3 Saccadic Mean Velocity (px/s)

Saccadic mean velocity is the average speed of the eye during a saccadic movement. Figure 3 shows the saccadic mean velocity across all participants. The average mean velocity decreases as task complexity increases with an overall reduction of 7.26% from Level 1 to Level 3 which indicate that participants generally slow their saccades as task complexity rises indicating higher cognitive demand.

ID 3 shows a consistent increase in velocity across levels, moving from 1432.04 px/s in Level 1 to 1946.94 px/s in Level 3. ID 2 followed a similar trend. This suggests that ID 2 and 3 adapts to increasing task complexity by speeding up their saccadic movements, contrary to the general trend. ID 6 mean velocity increased from Level 1 to Level 2 suggests that ID 6 may approach intermediate complexity with higher efficiency or speed. ID 4 maintains relatively steady velocities which indicate a consistent visual and cognitive strategy across all task levels.

Our findings show a weak negative correlation between saccadic mean velocity and task complexity with an average Pearson correlation coefficient of -0.077. While the overall trend shows a decrease in mean velocity with task complexity, the substantial variability among participants weakens its reliability as a robust indicator of cognitive load. This is in line with existing research that suggest [26, 31] that task complexity had no

effect on saccadic mean velocity.

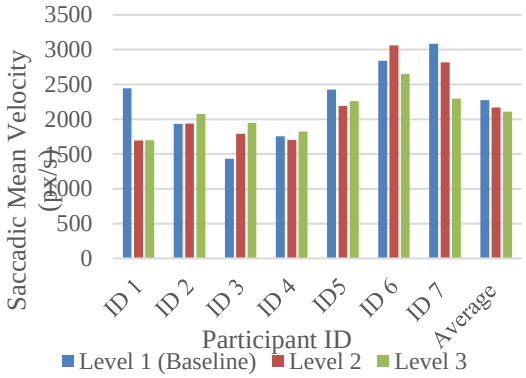


Figure 3. Saccadic Mean Velocity Across Task Complexity Levels

4.1.4 Blink Rate (Blinks/second (Hz))

Another indicator for cognitive load is blinking rate. Figure 4 shows the blinking rate for the seven participants across three levels of task complexity as well as their average. The average blink rate decreases consistently from 1.23 blinks (Level 1) to 0.96 (Level 2) and 0.7 (Level 3) as task complexity increases. Also, our results showed a strong negative correlation between blink rate and task complexity with an average Pearson correlation coefficient of -0.967. This indicates that blink rate decreases significantly as task complexity increases.

This reflects increasing cognitive demand as reduced blinking is commonly linked to increased focus and engagement. Blink rate varies among participants due to individual differences which is why a baseline was established. Despite these variations all participants showed a downward trend and a strong negative correlation coefficient. This aligns with previous research [33, 34] and reinforcing blink rate's reliability as a measure of cognitive load and its strong linkage to task complexity.

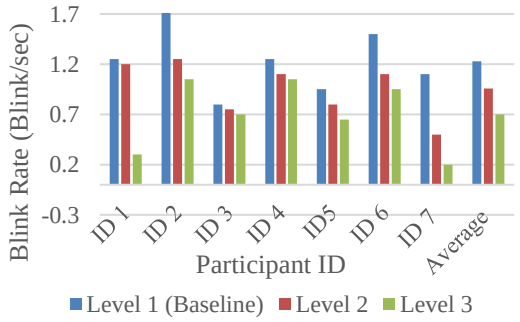


Figure 4. Blink Rate Across Task Complexity Levels

4.1.5 Fixation Rate (Fixations/Second (Hz))

Fixation rate refers to the frequency of fixations made by the eye per unit of time during a task. A fixation occurs when the eye stops moving and focuses on a single point to gather visual information. Fixation rate is typically expressed as fixations per second (Hz) and is a key metric in eye-tracking research. Figure 5 presents data for seven participants across three levels of task complexity with their individual values and overall averages provided for each level. The average fixation rate values decreased as task complexity increased. Also, our results indicate a moderate negative correlation between fixation rate and task complexity with an average Pearson correlation coefficient of -0.671.This aligns with trends seen in cognitive load research, where participants tend to adapt their behaviour as complexity rises and prioritizing precision over speed [2, 35]. All participants showed the same trend except for ID 5 who showed a slight upward trend which might indicate individual variability as adaptation or improved efficiency to handle complexity

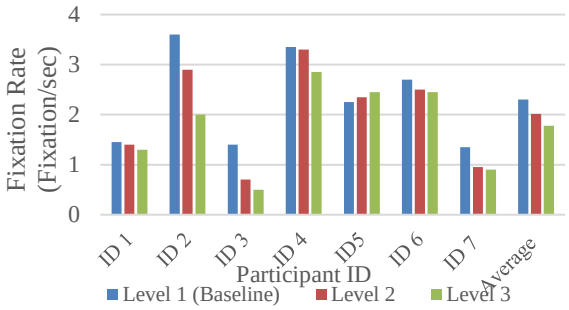


Figure 5. Fixation Rate Across Task Complexity Levels

4.2 Participants' Performance Evaluation

The impact of task complexity and cognitive load on performance was assessed by measuring the average distance between the participants' gaze points and the actual target coordinates. This measurement shows how accurately participants tracked the stimuli as the complexity of the task increased. A higher distance indicates higher error in tracking performance. It indicates that participants found it harder to keep their gaze aligned with the target as tasks became more challenging. The results align with this trend, with higher distances seen at higher task complexity levels.

Figure 6 represents the average Euclidean distance between gaze points and the actual coordinates of the stimuli for seven participants across three levels of task complexity. The Euclidean distance increased consistently as complexity increased. This indicate that as task complexity rises participants experience more difficulty tracking stimuli accurately. While the average trend supports this relationship, individual differences highlight variability in participants' ability to adapt to cognitive and visual demands. This suggests that tracking performance under complex tasks is influenced by both task characteristics and individual capabilities.

Our results show a very strong positive correlation

between Euclidean distance and task complexity with an average Pearson correlation coefficient of 0.967. This supports our findings that tracking error has a direct relationship with task complexity.

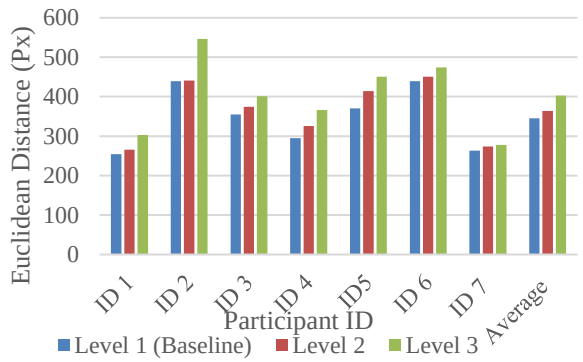


Figure 6. Euclidean Distance Across Task Complexity Levels

5 Insights for Construction Workflows

This study highlights the impact of visually demanding tasks on cognitive load and consequently performance. This could be particularly relevant to tasks of complex demand that require continuous monitoring, tracking, precise focus, and rapid decision-making based on visual information. This could include operating heavy machinery, rigging and lifting operations, welding and cutting, working at heights, and installation tasks.

Table 1 outlines specific recommendations for optimizing construction workflows based on findings from eye-tracking metrics and cognitive load analysis. Each recommendation is directly supported by observed data trends. This table provides actionable recommendations based on insights from cognitive load and eye-tracking metrics. By addressing specific challenges like task complexity, visual distractions, and worker fatigue, these strategies can enhance safety, productivity, and efficiency in construction workflows. Each recommendation is supported by findings that directly link eye-tracking data to cognitive performance under varying task complexities.

Table 1. Insights for Construction Workflows

Recommendation	Support by Findings
Focus on Reducing Task Complexity	Higher Euclidean distances indicate that increased task complexity reduces tracking accuracy. Construction tasks should be simplified where possible by breaking them into smaller, manageable steps to reduce cognitive load and improve worker precision.
Enhance Training for High-Complexity Task	Participants who showed increased errors at higher complexity levels highlight the need for specialized training.
Design Visual Environment	The findings suggest that visual demands influence performance. Construction workflows should minimize unnecessary visual distractions and use clear, standardized markings to guide workers' attention, ensuring accurate tracking of critical elements.
Leverage wearable sensors to monitor workers cognitive load	Metrics like blink rate, fixation rate, peak velocity reflects how task complexity impact cognitive load. Real-time monitoring could identify high-stress periods and offer feedback to enhance accuracy.
Monitor Worker Performance and Adjust Workloads	Significant differences in metrics like mean velocity and tracking performance highlight the need for personalized task assignments.
Incorporate Rest Breaks in Workflows	Reduced blink rates at higher complexity could reflect visual fatigue which can be mitigated with rest breaks.

6 Limitations and Future Work

The study revealed significant individual differences in metrics like saccadic mean velocity, fixation rate, and tracking performance. These variations limit the ability to generalize findings across a broader population. Also, experiments were conducted in a controlled setting which may not fully replicate the dynamic and unpredictable nature of real-world construction sites. The participant pool was relatively small, which may not provide sufficient statistical power to generalize results to all workers in construction or other fields. Furthermore, the study tasks were simplified representations of real-world construction workflows. More complex or nuanced tasks could yield different cognitive load and performance patterns. While metrics like blink rate, fixation rate, and saccadic velocities were useful, they may not capture all dimensions of cognitive load. Other physiological indicators such as heart rate variability or pupil dilation, could provide additional insights.

7 Conclusion

This study examined the impact of task complexity on cognitive load and performance in construction

workflows using eye-tracking metrics. These metrics include blink rate, fixation rate, saccadic amplitude, saccadic peak velocity, saccadic mean velocity, and tracking accuracy (Euclidean distance). The findings provide insights into how workers respond to visually and cognitively demanding tasks.

The findings indicate that blink rate decreased with increased task complexity which could reflect higher cognitive load. Fixation rate and saccadic amplitude increased which suggests higher visual processing demands and the need to scan broader areas as task complexity increased. Saccadic peak velocity decreased which align with previous research that indicates a compensatory mechanism to prioritize accuracy over speed in complex scenarios. Saccadic mean velocity showed significant individual variability and a weak correlation coefficient; therefore, it's a less reliable metric for measuring cognitive load across participants. The participants' tracking performance was measured by Euclidean distance between gaze points and target coordinates. All participants showed higher error at higher complexity levels. This highlight reduced accuracy in visually challenging tasks. These results emphasize the strong relationship between task complexity, cognitive load, and performance.

Recommendations to enhance worker focus, accuracy, and safety include simplifying task designs, providing targeted training, leveraging wearable eye-tracking technology, and incorporating regular rest breaks. This research provides a foundational understanding of how cognitive and visual demands affect worker performance and offers actionable recommendations for optimizing construction tasks and improving workplace safety and efficiency.

References

[1] J. Sweller, "Cognitive Load During Problem Solving: Effects on Learning," *Cognitive Science*, vol. 12, no. 2, pp. 257–285, 1988, doi: 10.1207/s15516709cog1202_4.

[2] J.-C. Liu, K.-A. Li, S.-L. Yeh, and S.-Y. Chien, "Assessing Perceptual Load and Cognitive Load by Fixation-Related Information of Eye Movements," *Sensors*, vol. 22, no. 3, Art. no. 3, Jan. 2022, doi: 10.3390/s22031187.

[3] J. Sweller, P. Ayres, and S. Kalyuga, *Cognitive Load Theory*. New York, NY: Springer New York, 2011. doi: 10.1007/978-1-4419-8126-4.

[4] F. Paas, J. E. Tuovinen, H. Tabbers, and P. W. M. Van Gerven, "Cognitive Load Measurement as a Means to Advance Cognitive Load Theory," *Educational Psychologist*, vol. 38, no. 1, pp. 63–71, Jan. 2003, doi: 10.1207/S15326985EP3801_8.

[5] S. Rubio, E. Díaz, J. Martín, and J. M. Puente, "Evaluation of Subjective Mental Workload: A Comparison of SWAT, NASA-TLX, and Workload Profile Methods," *Applied Psychology*, vol. 53, no. 1, pp. 61–86, 2004, doi: 10.1111/j.1464-0597.2004.00161.x.

[6] B. Mehler, B. Reimer, J. F. Coughlin, and J. A. Dusek, "Impact of Incremental Increases in Cognitive Workload on Physiological Arousal and Performance in Young Adult Drivers," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2138, no. 1, pp. 6–12, Jan. 2009, doi: 10.3141/2138-02.

[7] S. M. Jamil Uddin, N. Tabassum, A. Ovid, A. Alsharef, and A. Albert, "Measuring Mental Fatigue in Construction: State of the Science and Future Opportunities," in *Construction Research Congress 2024*, Des Moines, Iowa: American Society of Civil Engineers, Mar. 2024, pp. 688–698. doi: 10.1061/9780784485293.069.

[8] H. Li, D. Wang, J. Chen, X. Luo, J. Li, and X. Xing, "Pre-service fatigue screening for construction workers through wearable EEG-based signal spectral analysis," *Autom. Constr.*, vol. 106, p. 102851, Oct. 2019, doi: 10.1016/j.autcon.2019.102851.

[9] M. Noghabaei, K. Han, and A. Albert, "Feasibility Study to Identify Brain Activity and Eye-Tracking Features for Assessing Hazard Recognition Using Consumer-Grade Wearables in an Immersive Virtual Environment," *Journal of Construction Engineering and Management*, vol. 147, no. 9, p. 04021104, Sep. 2021, doi: 10.1061/(ASCE)CO.1943-7862.0002130.

[10] J. Li *et al.*, "Identification and classification of construction equipment operators' mental fatigue using wearable eye-tracking technology," *Autom. Constr.*, vol. 109, p. 103000, Jan. 2020, doi: 10.1016/j.autcon.2019.103000.

[11] L. L. Di Stasi, M. Marchitto, A. Antolí, T. Baccino, and J. J. Cañas, "Approximation of on-line mental workload index in ATC simulated multitasks," *Journal of Air Transport Management*, vol. 16, no. 6, pp. 330–333, Nov. 2010, doi: 10.1016/j.jairtraman.2010.02.004.

[12] L. L. Di Stasi, A. Antolí, M. Gea, and J. J. Cañas, "A neuroergonomic approach to evaluating mental workload in hypermedia interactions," *International Journal of Industrial Ergonomics*, vol. 41, no. 3, pp. 298–304, May 2011, doi: 10.1016/j.ergon.2011.02.008.

[13] G. Lee, B. Choi, H. Jebelli, and S. Lee, "Assessment of construction workers' perceived risk using physiological data from wearable sensors: A machine learning approach," *Journal of Building Engineering*, vol. 42, p. 102824, Oct. 2021, doi: 10.1016/j.jobe.2021.102824.

[14] Y. H. Kwon, Y. B. Kwon, D. C. Nwagbala, and J. Y. Park, "The Cognitive Load Limits of Multiple Safety Signs," *Buildings*, vol. 14, no. 8, p. 2391, Aug. 2024, doi: 10.3390/buildings14082391.

[15] P. Mitropoulos and B. Memarian, "Task Demands in Masonry Work: Sources, Performance Implications, and Management Strategies," *J. Constr. Eng. Manage.*, vol. 139, no. 5, pp. 581–590, May 2013, doi: 10.1061/(ASCE)CO.1943-7862.0000586.

[16] B. Memarian, B. Memarian, and P. Mitropoulos, "Organizing production for speed and safety: Lessons from a multi-story concrete building," 2012.

[17] Z. Zhang, T. Xiang, H. Guo, L. Ma, Z. Guan, and Y. Fang, "Impact of physical and mental fatigue on construction workers' unsafe behavior based on physiological measurement," *Journal of Safety Research*, vol. 85, pp. 457–468, Jun. 2023, doi: 10.1016/j.jsr.2023.04.014.

[18] Y. Han, R. Jin, H. Wood, and T. Yang, "Investigation of Demographic Factors in Construction Employees' Safety Perceptions," *KSCE J Civ Eng*, vol. 23, no. 7, pp. 2815–2828, Jul. 2019, doi: 10.1007/s12205-019-2044-4.

[19] J. Li, H. Li, H. Wang, W. Umer, H. Fu, and X. Xing, "Evaluating the impact of mental fatigue on construction equipment operators' ability to detect hazards using wearable eye-tracking technology," *Autom. Constr.*, vol. 105, p. 102835, Sep. 2019, doi: 10.1016/j.autcon.2019.102835.

[20] I. Mehmood, H. Li, W. Umer, A. Arsalan, M. S. Shakeel, and S. Anwer, "Validity of facial features? geometric measurements for real-time assessment of mental fatigue in construction equipment operators," *Adv. Eng. Inform.*, vol. 54, p. 101777, Oct. 2022, doi: 10.1016/j.aei.2022.101777.

[21] D. Wang, J. Chen, D. Zhao, F. Dai, C. Zheng, and X. Wu, "Monitoring workers' attention and vigilance in construction activities through a wireless and wearable electroencephalography system," *Automation in Construction*, vol. 82, pp. 122–137, Oct. 2017, doi: 10.1016/j.autcon.2017.02.001.

[22] X. Xing, H. Li, J. Li, B. Zhong, H. Luo, and M. Skitmore, "A multicomponent and neurophysiological intervention for the emotional and mental states of high-altitude construction workers," *Autom. Constr.*, vol. 105, p. 102836, Sep. 2019, doi: 10.1016/j.autcon.2019.102836.

[23] AWCBC, "National Work Injury, Disease and Fatality Statistics-2020-2022," *Association of Workers' Compensation Boards of Canada*, 2023, [Online]. Available: https://awcbc.org/files/publications/National-Work-Injury-Disease-and-Fatality-Statistics-2020-2022.pdf

[24] L. L. Di Stasi *et al.*, "Saccadic Peak Velocity Sensitivity to Variations in Mental Workload," *aviat space environ med*, vol. 81, no. 4, pp. 413–417, Apr. 2010, doi: 10.3357/ASEM.2579.2010.

[25] L. L. Di Stasi, A. Antolí, and J. J. Cañas, "Main sequence: An index for detecting mental workload variation in complex tasks," *Applied Ergonomics*, vol. 42, no. 6, pp. 807–813, Nov. 2011, doi: 10.1016/j.apergo.2011.01.003.

[26] S. Aatrai, S. K. Jha, and R. Guha, "Visual Perception and Performance: An Eye Tracking Study," in *Proceedings of the 2023 Symposium on Eye Tracking Research and Applications*, in ETRA '23. New York, NY, USA: Association for Computing Machinery, May 2023, pp. 1–5. doi: 10.1145/3588015.3588424.

[27] B. L. Zuber, L. Stark, and G. Cook, "Microsaccades and the Velocity-Amplitude Relationship for Saccadic Eye Movements," *Science*, vol. 150, no. 3702, pp. 1459–1460, Dec. 1965, doi: 10.1126/science.150.3702.1459.

[28] K. Muhammed, E. Dalmaijer, S. Manohar, and M. Husain, "Voluntary modulation of saccadic peak velocity associated with individual differences in motivation," *Cortex*, vol. 122, pp. 198–212, Jan. 2020, doi: 10.1016/j.cortex.2018.12.001.

[29] C. M. Harris and D. M. Wolpert, "The Main Sequence of Saccades Optimizes Speed-accuracy Trade-off," *Biol Cybern*, vol. 95, no. 1, pp. 21–29, Jul. 2006, doi: 10.1007/s00422-006-0064-x.

[31] D. P. Munoz and S. Everling, "Look away: the anti-saccade task and the voluntary control of eye movement," *Nat Rev Neurosci*, vol. 5, no. 3, pp. 218–228, Mar. 2004, doi: 10.1038/nrn1345.

[32] C. Page, C. C. Liu, J. Meltzer, and S. Ghosh Hajra, "Blink-Related Oscillations Provide Naturalistic Assessments of Brain Function and Cognitive Workload within Complex Real-World Multitasking Environments," *Sensors*, vol. 24, no. 4, Art. no. 4, Jan. 2024, doi: 10.3390/s24041082.

[33] V. Bachurina and M. Arsalidou, "Multiple levels of mental attentional demand modulate peak saccade velocity and blink rate," *Heliyon*, vol. 8, no. 1, p. e08826, Jan. 2022, doi: 10.1016/j.heliyon.2022.e08826.

[34] K. Walter and P. Bex, "Cognitive load influences oculomotor behavior in natural scenes," *Sci Rep*, vol. 11, no. 1, p. 12405, Jun. 2021, doi: 10.1038/s41598-021-91845-5.

Adjustments and Assembly of a Novel Robotic End-effector for Construction Automation

Megan Rudo[1], Jiansong Zhang, Ph.D.[2]

[1] Elmore Family School of Electrical and Computer Engineering, Purdue University, United States
[2] School of Construction Management Technology, Purdue University, United States
mrudo@purdue.edu, zhan3062@purdue.edu

Abstract -

The construction industry is currently behind other non-farming industries on the trends of updating technologies, especially in implementing robotic automation in its processes. Because of this, there are numerous opportunities for development of new technologies in the construction domain for automating tasks, lowering risks, and decreasing costs. The focus of this paper was to take a newly patented design for a robotic end effector to build wooden frames and start to build a physical prototype. More specifically, unfinished or nonfunctional parts are updated, and portions of the end effector were assembled.

Keywords

Robotic Automation, End Effector, Timber Construction, Construction Robotics

1 Introduction

It is widely known that construction industry has lagged behind other non-farming industries in its productivity improvement, yet not enough has been done with advancements in technology to improve that in recent years. This research investigates one of the possible solutions through robotic automation, specifically the development of a prototype to demonstrate how a new end effector can help automate the construction of timber frames.

Robotic automation is set out to address different challenges within the construction community such as in improving safety. According to Pinto et al [1], some of the causes of construction safety hazards include large company sizes and therefore a large number of workers to train, constantly changing worksites, worker specialization to one machine or area, insufficient protective equipment, use of drugs on the job, and long-term health problems from the stress of lacking job security [1]. Robotic automation provides a promising solution for mitigating such risks. However, for the last 50 years, limited advancement in robotic automation has come to the construction industry [2], which is notably delayed compared to other industries [2].

In addition to safety risk exposure, the construction industry is also costly, partially due to the number of employees, materials, equipment, and accidents related. It is estimated in a study by Everett and Frank Jr that the cost of accidents in the construction industry has risen from 6.5% in 1979 of total cost of nonresidential construction to between 7.9% to 15.0% in 1996 [3]. That increase, which is at least 1.4%, is millions of dollars that could have been spent on updated technologies. In 2002, it's estimated that $11.5 billion was spent on injuries in construction [4]. According to the federal reserve bank of St Luis, the average annual cost of construction in the US in 1994 was 530,980 million dollars. In 2023, that average was raised to 2,021,695 million dollars, a 280.8% increase [5]. As projects become more complicated and more costly, the need to mitigate these costs increases. Robotic automation is considered a promising solution. One way to pursue this is to use robots to construct elements that are used multiple times within a project, such as timber frames. Zhang et al. [6] patented a technology in that regard and this paper focuses on implementing the patented technology to an existing robotic arm.

2 Methodology

The goal is to build a working prototype of the patented end effector design [6] (Figure 1) as a proof of concept for companies to adopt this technology in future construction operations. This integrates: (1) a mount for a robotic arm, (2) two rotating shafts enabling it to encircle timber pieces, (3) suction plates for secure wood handling, and (4) a nail gun for efficient timber assembly. This configuration ensures

precise and robust manipulation of timber, enhancing the overall automation process in construction tasks.

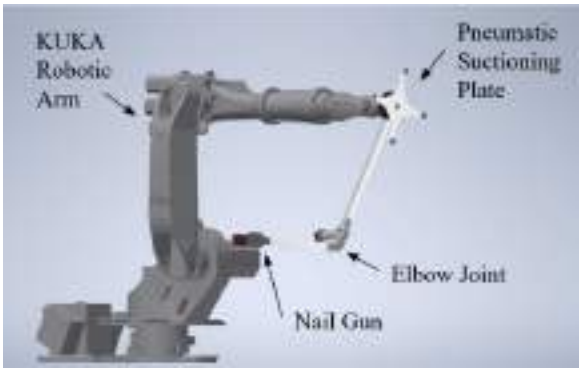


Figure 1: Patented End-effector Design [6]

Figure 2 shows the work breakdown structure of this research, representing all tasks to be completed. These tasks are categorized into design, manufacturing, assembly, electrical, programming, and documentation. The prioritization of tasks was based on assemble sequence (onto the arm), with documentation being produced and used throughout the design iterations. CAD designs and sample prints were used for validation of design.



Figure 2: Work Breakdown Structure

3 Previous Research

The researchers chose to use the KUKA KR 16 L6 robotic arm with a KR C2 controller for implementation, following a lineage research [7]. Before powering the robotic arm, an inspection was done to ensure all connections were secure and accurately placed. Although a design for the end effector was almost fully implemented in CAD, the nail gun mount and trigger were not yet designed. All parts of the CAD design had been ordered, printed, or manufactured [2].

4 End Effector Updates

4.1 Design

The first piece updated was the wrist mount, which is a 3D printed attachment that directly connects the end effector to the KUKA's mounting plate. When initially attempting assembly, the screws rested about ¾" above the surface of the wrist mount, making the connection to another bracket impossible. As a solution, the wrist mount was extended, and

countersinks were placed (Figure 3). This edit ensures the screws are underneath the top surface of the bracket, so the connecting bracket lies flush on the surface and connecting screws can be placed through the holes between the two.



Figure 3: Countersunk Wrist Mount

Since the thin slots in the wrist mount form a weak point, which could potentially snap and cause problems later, they will be connected instead by a solid block. As shown in Figure 4 below, the cutout is now one rectangle in the counterpart to the mount, and on the mount it is a rectangular prism extrusion with holes for the connection screws. For the bearing case, the existing CAD was duplicated and split into two halves, as shown in Figure 5. There are already built-in holes for screws to attach the two pieces.



Figure 4: Wrist Mount Updated CAD

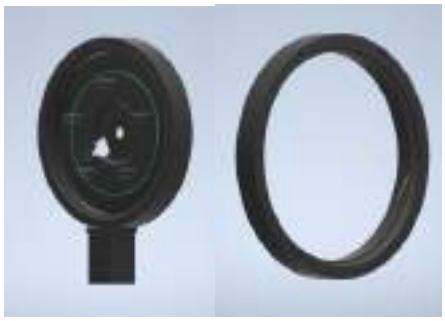


Figure 5: Bearing Case Updated CAD

Redesign of the motor mount was also conducted which is challenging due to small tolerances, yet it still needs to be able to provide ample support for this motor and the attaching encoder. Figure 6 shows the redesign results which extended the previous mount design on both ends by ¼", and then split in half.



(1)



(2)

Figure 6: (1) Motor Mount CAD and (2) Halved Versions

In addition, two larger holes were added on the opposite sides of the motor mount for the shaft collar adjustment to be possible once the motor is mounted (Figure 7).

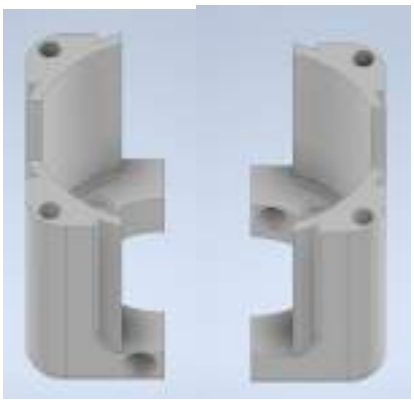


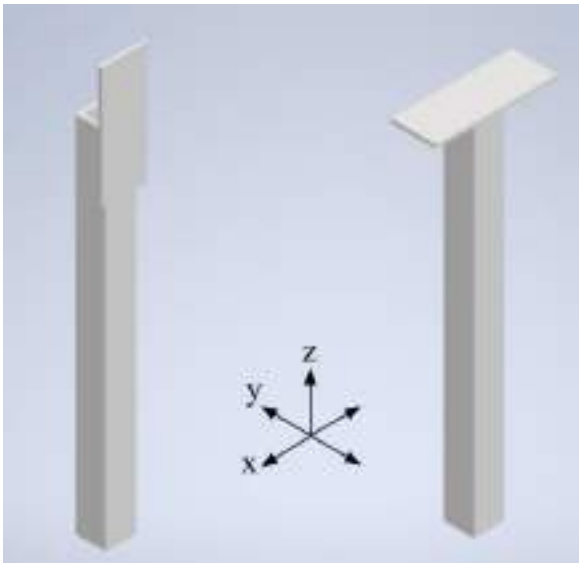
Figure 7: New Motor Mount CAD

The top innermost portion of the bearing case, known as eccentric bearing 2 in previous design [7], was too far off center to fit properly flush with the case due to the off-centered nature of the dovetail joint portion. The outer ring is cut down significantly to ensure its fit (Figure 8).



Figure 8: Updated Bearing

For the nail gun mount, the main concern was the amount of torque applied on the connected arm due to the kickback force when the nail gun is fired. We therefore designed the nail gun to be attached to a metal bar. This needs to be secure, because the kickback force when a nail gun is fired will apply a lot of torque on the rest of the system. Figure 9 below shows how the different mounting positions will affect the direction the kickback force is applied. For the option on the left, the force is dispersed throughout the metal bar, applying torque on the joint the bar is attached to. For the option on the right, the torque is applied directly on the mount for the nail gun and the bar itself. With a sturdy bar and mount, this would be the best option. Our prototype is currently being built with a hollow aluminum bar, which would most likely twist over time with continued use. For long-term solutions, a solid aluminum bar or a hollow steel bar would be good alternatives.



Orientation 1 Orientation 2

Figure 9: Mounting Positions

To address this kickback force issue, the forces were analyzed to determine which direction the torque should be applied to minimize the impact of the kickback force. In orientation 1, the rotational kickback force is applied along the xz plane, possibly bending the metal bar in that plane. In orientation 2, that force is applied along the xy plane, potentially twisting the metal bar in that xy plane. Because the metal bar is hollow, it needs additional support (not shown in figure) in both orientations to prevent damage with use. In orientation 1, this bracket would need to extend along a significant portion of the bar. However, with orientation 2, the bracket only needs to span the top portion (either covering a small portion of the sides or within the hollow bar) to prevent any twisting. Correspondingly, a design was created to mount in orientation 2, but it should be noted that the kickback force could shatter the bearing and might need to be addressed in another design update. This is currently partially addressed by the choices in manufacturing. Since the latest print is made in PLA+ filament instead of regular PLA, it can handle a relatively stronger torque being applied onto it, and with 60% infill and gyroid pattern infill, this has the

potential to be able to handle the forces applied in orientation 2.

For mounting the nail gun, the initial conceptual design is to put the nail gun in a fitted 3D printed box, and zip tie it tight.

The very first nail gun mount design is shown in Figure 10. It is a box slightly larger than the nail gun itself, (⅛" walls, inner box 2" x 9.5") for the gun to slide into with minimal wiggle room. In addition, there is a slot at the bottom of the box for the tip of the nail gun to sit in, and for nails to shoot out from.

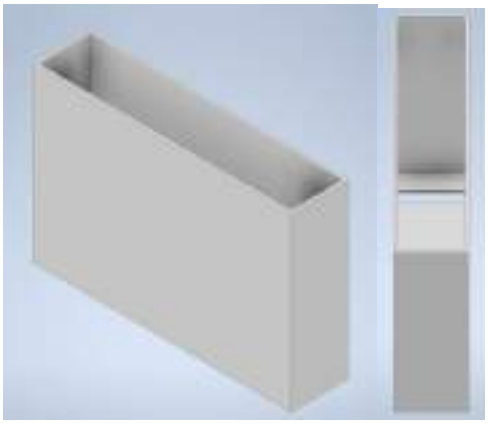


Figure 10: Nail Gun Initial CAD

The updates to the nail gun CAD are shown in Figure 11. For the update to the nail gun mount, 4 ¼" holes were added around the main handlebar for the nail gun. In addition, the slot previously in the design was changed to be a smaller centered hole, in the hope of minimizing any jitter by the gun affecting the aim of the nails, by having that extension of the nail gun more tightly slotted into place. Also, the back corner of the box was cut out just enough to be able to access the nail refill easily, without taking away much structural integrity from the mount.

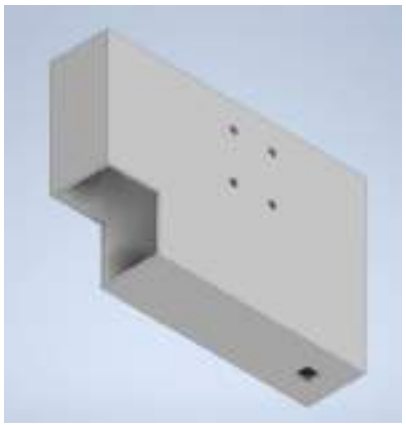


Figure 11: Nail Gun Updated CAD

By making the mount a flat sheet instead of a full box, it will be significantly easier, faster, and cheaper to build. After a few iterations, Figure 12 shows our final design. The updated nail gun CAD from Figure 11 was copied, and then cut to be the back plate only. More holes were then added for security of the nail gun.



Figure 12: Buildable Nail Gun Mount, Overlay

The nail gun selected has two triggers, as shown in Figure 13. The bottom trigger is a safety, and the trigger on top is for shooting the nail.



Figure 13: Nail Gun

4.2 Manufacturing

3D printing was used for manufacturing. When selecting the filament type, the goal was min/maxing cost/strength. PLA+ presented a good balance, so eSUN PLA+ 1.75mm in a 2.2lb spool was selected, because of its high tensile strength, comparable to that of POM filament but with a fraction of the cost. It is necessary to have high tensile strength on these prints given the amount of weight they will be carrying and the resultant gravitational force being placed, so regular PLA or ABS was passed.

With the PLA+ filament and 60% gyroid infill and supports where needed, the parts were printed (Figure 14).



Figure 14: Finished 3D Prints

In addition, during the 3D printing process, the distance between supports and the print was decreased. This does make the supports more difficult to remove and uses more filament, but it made the printing results significantly better than before so is worth the added effort. The difference between the two prints (white vs red, as shown in Figure 15) is the direction the lines were drawn (linear vs concentric, respectively). The white one came out with smoother finish and less cracking/holes, so it will be the one attached to the build.



Figure 15: Eccentric Bearing Reprints

4.3 Assembly

Once the prints were completed, an assembly of the end effector was attempted.



Figure 16: Elbow Arm Assembly



Figure 17: Motor Mount Assembly

Parts were sanded down in order to fit in their respected places due to a bit of filament scraps stuck on. For example, both wrist mount pieces have some sharp spots on the corners that were sanded down before sliding into place. The bearing case assembled is shown in Figure 18.



Figure 18: Partially Assembled Bearing Case



Figure 19: Motor Mount Too Thick

Currently attached to the KUKA, as shown in Figure 20, is the wrist mount, bearing, and the encoder and motor.



Figure 20: Current Assembly

5 Conclusions and Future Work

A partially functional prototype of a patented end effector for robotic automation in timber construction is assembled, new and updated parts are added onto the previous design, and a possible solution to the previously reported issue on the robotic arm is identified. For the future, the nail gun mount needs to be built and attached, and aluminum sheet can be used as its material. The electronic parts of the end effector then need to be connected. Once the end effector is fully assembled and connected, the KUKA arm is ready to be programmed with the teach pendant and tested for full functionality. For evaluation at that stage performance metrics include for example a 95% success rate of assembly of the wooden frames and no damage to the end effector after 20 heavy uses.

6 Acknowledgements

The authors would like to thank the National Science Foundation (NSF). This material is based on work supported by the NSF under Grant No. 1827733. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

References

[1] A. Pinto, I. Nunes, and R. Ribeiro,
Occupational risk assessment in construction industry – Overview and reflection. *Safety Science*, 49(5):616-624, 2011.

[2] W. Alaloul, M. Liew, N. Zawawi, and B. Mohammed. Industry Revolution IR 4.0: Future opportunities and challenges in construction Industry. In *MATEC Web Conference*, 2018.

[3] J. Everett and P. Frank. Costs of Accidents and Injuries to the construction industry. *Journal of Construction Engineering and Management*, 122(2):158, 1996.

[4] G. Waehrer, X. Dong, T. Miller, E. Haile, and Y. Men. Costs of occupational injuries in construction in the United States. *Accident; analysis and prevention*, 39(6):1258-1266, 2007.

[5] Total Construction Spending. *Federal Reserve Bank of St. Louis*, U.S. Federal Reserve Economic Data, On-line: https://fred.stlouisfed.org/series/TTLCONS?form=MG0AV3#0. Accessed: 12/30/2024

[6] J. Zhang, M Lacny C., and Reardon N. Autonomous robotic system for placing and fastening paneling material for building construction operations, U.S. Patent Number: 12090628, September 17, 2024.

[7] A. Borgo, A. Ruiz, J. Zhang, and L. Felix-Herran. Implementation of a robotic manipulator end effector for construction automation, In *Proceedings of the 2024 International Symposium on Automation and Robotics in Construction (ISARC 2024)*, Pages 177-184, Lille, France, 2024.

Tailored Vision-Language Solutions for Comprehensive Hazard Identification on Construction Sites

Qihua Chen[1] and Xianfei Yin[1,*]

[1]Department of Architecture and Civil Engineering, City University of Hong Kong, Hong Kong SAR
*Corresponding Author
qihua.chen@cityu.edu.hk, xianfyin@cityu.edu.hk

Abstract –

The identification of hazards is critical for mitigating accident risks on construction sites. Although current automatic recognition methods based on computer vision have demonstrated significant success, their performance can be limited by inadequate generalization capabilities and challenges in handling dynamic, complex scenes, leading to incomplete detection of various types of construction hazards. This study introduces a novel approach that leverages a tailored vision-language model (VLM) to enhance comprehensive hazard identification on construction sites. By capitalizing on the cross-modal understanding and robust generalization abilities of VLMs, this method aims to improve both the accuracy and comprehensiveness of construction hazard identification. To validate this approach, we constructed a specialized dataset comprising 1,139 images across 31 fine-grained categories of construction hazards for fine-tuning and evaluation. The experimental results revealed that, when fine-tuned on the Qwen2-VL-7B model, the VLM achieved a precision of 0.856, representing a substantial improvement from 0.495 precision in the non-fine-tuned state. Moreover, fine-tuning significantly reduced prediction errors and enhanced the model's capability to understand and identify construction hazards. In addition, advanced VLM models demonstrate greater application potential compared to traditional models such as CLIP. This study advances the accuracy and comprehensiveness of hazard identification on construction sites, offering a new technological perspective for integrated automatic hazard detection in construction environments.

Keywords –

Construction management; Hazard identification; Vision-language model; Fine-tuning

1 Introduction

There is a wide variety of hazards on construction sites, such as workers not wearing safety helmets and electrical wires submerged in water, which constitute major factors leading to construction safety accidents [1,2]. To mitigate these risks, traditional construction safety management has relied on the Behavior-Based Safety (BBS) approach, centered around human inspections and timely corrections [3]. However, with the rapid advancement of artificial intelligence technology, especially in computer vision, and the development of advanced construction site monitoring systems utilizing cameras [4,5], automated methods for detecting safety hazards have begun to emerge. These methods, for example, checking compliance with personal protective equipment (PPE) [6], monitoring workers' access to hazardous areas [7], and ensuring the orderly stacking of construction materials [8], have demonstrated efficacy. Despite this progress, most existing methods are limited to identifying a small number of hazards and fail to consider multiple hazard factors in an integrated manner. Moreover, because model training depends on labeled data, it limits the models' generalization capabilities and hinders their adaptation to unseen, complex dynamic scenes [6]. Consequently, there remains a critical need for a method that can accurately and comprehensively recognize various types of construction hazards through advanced scene understanding and perception, particularly within the complex and dynamic construction environments.

With the development of pre-trained vision-language model (VLMs), such as CLIP [9]，BLIP [10]，LLaVA [11], significant progress has been made in understanding and generating textual descriptions associated with images. These models provide richer and more detailed insights by accumulating robust image and text feature representation capabilities through pre-training on large-scale datasets. Compared to traditional computer vision methods, VLMs exhibit higher robustness and generalization capabilities when handling complex scenes [12]. Additionally, the self-supervised learning mechanism of VLMs allows them to acquire knowledge from unlabeled data, thereby reducing reliance on manually labeled data [13].

Therefore, the application of VLMs in construction

management holds great promise. In past studies, Yong et al. [14] utilized BLIP to identify ergonomics issues and their solutions with the aid of data augmentation. The study by Liang et al. [15] demonstrated the potential of using CLIP for few-shot learning to classify construction site objects, achieving an increase in classification accuracy from 58.33% to 83.33% with only one sample for training. For construction safety management, Tsai et al. [12] automatically generated construction safety observations by fine-tuning CLIP, attaining an average accuracy of 73.7% across nine types of violations. However, the proposed hazard identification methods generate caption content that requires professionals to judge the construction hazard categories, rather than directly outputting violation categories from the input images. Moreover, construction sites present a variety of complex hazards, such as unguarded rotating parts of machinery, necessitating more specific and fine-grained hazard identification capabilities. Traditional VLMs, like CLIP , show limited advanced language comprehension and understanding of complex scenarios [16]. It needs to be combined with other language models to generate the corresponding textual content [12]. Therefore, research gaps still exist in the end-to-end construction hazard identification and reporting process, which requires sophisticated visual reasoning capabilities. This highlights the need to explore and apply more advanced VLMs, such as the pre-trained Qwen2-VL [17], to address these challenges.

Therefore, to address the limitations and achieve more specific, fine-grained identification of construction hazard categories, this study proposes an end-to-end construction safety hazard recognition framework based on a fine-tuned VLM. This framework leverages the advanced image comprehension and reasoning capabilities of pre-trained VLMs to identify construction site safety hazards. The main contributions of this paper are as follows: (1) constructing a dataset with 31 fine-grained categories of construction hazards for fine-tuning and evaluating the hazard identification model; (2) demonstrating the effectiveness of advanced VLMs and fine-tuning methods in construction hazard identification; and (3) implementing an automated end-to-end process for construction site hazard inspections.

2 Methodology

Figure 1 illustrates the proposed framework, which comprises three principal components. Firstly, dataset construction: this stage involves collecting and annotating a series of image data related to construction hazards, serving as the foundation for fine-tuning the VLM. Secondly, VLM fine-tuning process: we utilize an advanced open-source pre-trained VLM and apply advanced fine-tuning techniques to optimize the model using the constructed construction hazard dataset. The performance of the fine-tuned model is subsequently evaluated using a test set. Lastly, deployment for construction hazard identification: the validated and optimized model is deployed at construction sites, leveraging real-time image data streams from CCTV systems to enable immediate monitoring and identification of construction hazards. Detailed descriptions of these three critical stages are provided in the following sections.

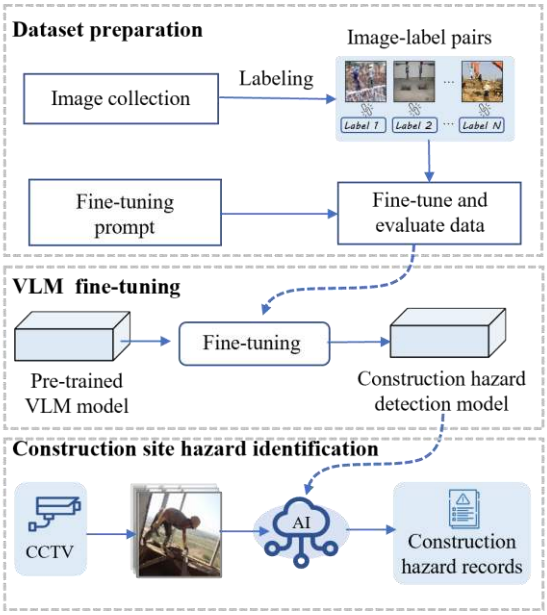


Figure 1. The framework of the proposed method

2.1 Dataset preparation

In BBS-based management, construction hazards encompass not only the use of workers' PPE but also other aspects, such as the utilization of construction tools and equipment. Based on the previous research by Chen et al. [3], Lee et al. [18] and practical experience from construction sites, this study has systematically identified common construction hazards, categorizing them into 31 fine-grained categories as presented in Table 1. Overall, Table 1 includes violations related to PPE wearing, the status of construction equipment, and the construction environment. Subsequently, based on the categories of construction safety hazards outlined in Table 1, we gathered images representing various types of construction hazards. The majority of these images were sourced from historical inspection records of construction projects. These hazard inspection logs contain images and detailed descriptions of the hazard categories. However, given the limited availability of historical inspection data, we supplemented images with publicly available datasets containing construction site images, such as CPPED [6] and image data collected

from the Internet. Subsequently, we conducted a filtering process for all images, which involved verifying their relevance to construction site scenarios and removing low-quality or blurry images. As a result, we assembled a dataset comprising 1,139 high-quality construction hazard images.

Table 1. 31 categories of construction hazards

No.	Construction hazards
H01	Not wearing safety helmets
H02	Wearing a safety helmet but not fastening the chin strap
H03	Not wearing safety vests
H04	Not wearing protective shoes
H05	Workers not wearing gloves during operations such as handling materials.
H06	Workers performing high-altitude work without wearing safety harnesses or improperly securing the harness hooks
H07	Illegally climbing over safety railings or barriers
H08	Debris piled on scaffolding
H09	Construction machinery illegally transporting or hoisting construction workers
H10	Workers smoking at the construction site
H11	Entering the operating range of an excavator
H12	Unauthorized use of open flames
H13	Rotating parts of machinery lack guards
H14	Crane operations without setting up safety warning lines around the crane
H15	Gas cylinders lacking protection
H16	Hooks missing safety pins
H17	Edge protection barriers not installed or having gaps
H18	Openings and shafts lack protective measures
H19	Guardrails not installed at open edges of staircases
H20	Distribution boxes left with doors open
H21	Distribution boxes not elevated (in contact with the ground)
H22	Electrical wires not elevated (laid directly on the ground) or poorly organized
H23	Unauthorized use of plug-in boards
H24	Electrical wires submerged in water
H25	Standing water present at the construction site
H26	Lack of clearance for construction waste
H27	Construction materials piled disorderly
H28	Temporary passageways at the construction site insufficiently stabilized or lacking safety measures
H29	Safety nets on exterior scaffolding damaged
H30	Dust prevention nets damaged
H31	Gaps existing in the perimeter fencing

According to the predefined categories, we categorized these images, ensuring that each image represented at least one category during the annotation process. The data annotation process is subject to standardized quality control measures, including cross-check inspections by annotators and final approval by experienced construction safety managers, to ensure the reliability of the dataset. Samples of hazard images and corresponding labels are illustrated in Figure 2.

In order to verify the performance of fine-tuned VLMs, a portion of the data from the dataset is extracted for testing, ensuring that the number of labels in each category within the test set is not less than 8. The stratified sampling method is used to divide the dataset into a test set and a training set in a 3:7 ratio. The final proportion of each label in both the training set and the test set is shown in Fig. 3. Subsequently, we generate the final datasets for both the training and test sets using designed VLM fine-tuning prompt as illustrated in Fig. 4



Figure 2. Examples of construction hazards and corresponding labels



Figure 3. Number of labels in the test and train sets

User: You are an expert in identifying construction site hazards. Please list all categories of hazards present in the image.

ChatCH: The construction site hazard categories identified in the image are as follows: {*Hazard Categories Included*}

Figure 4. Prompt for VLM fine-tuning

2.2 VLM fine-tuning

Fine-tuning a VLM is a process that adapts a pre-trained VLM to a specific task or domain. Initially, the pre-trained model undergoes extensive training on a large-scale dataset, enabling it to learn general feature representations. Subsequently, during fine-tuning, the model leverages task-specific data to enhance its performance on specialized tasks and improve its generalization capabilities in targeted applications. Through this adaptation, the fine-tuned VLM can better address the unique challenges of specific domains. To enhance the model's performance in construction hazard identification, further fine-tuning is necessary to improve the VLM model's understanding of construction-specific scenarios. Low-rank adaptation (LoRA) [19] stands out as an effective fine-tuning strategy that addresses challenges posed by traditional computer vision methods, such as overfitting. Figure 5 compares the direct use of pre-trained weights with the LoRA-based fine-tuning process. As illustrated in Figure 5, LoRA's core principle involves updating only a select subset of parameters deemed most crucial for the new task. Specifically, LoRA achieves parameter updates by incorporating low-rank matrices into the pre-trained model, rather than directly altering a vast number of original parameters. This method not only reduces the number of parameters requiring training, thereby lowering computational expense, but also aids in preserving the pre-trained model's knowledge and mitigating catastrophic forgetting.

In LoRA, the original weight matrix W is updated to $W = W_0 + AB$, where W_0 denotes the pre-trained VLM weights, and AB represents the product of matrices A and B. Here, A and B are designed to be low-rank matrices with rank r, which allows for minimal changes to W_0 while enabling efficient fine-tuning. The result of this multiplication is a low-rank matrix that is added to the original weights to form the new weight matrix. During the fine-tuning process, the objective is to adjust these parameters by minimizing the loss function $L(\theta, A, B)$ on the target task, which is formulated as follows:

$$L(\theta, A, B) = \sum_i \ell(f(x_i; W_0 + AB), y_i), \quad (1)$$

where θ represents the other parameters, x_i denotes the i-th sample of the input data, and y_i denotes the true label of the i-th sample. By precisely tuning the model parameters, LoRA enables the VLM to more effectively understand and predict various safety hazards in construction sites, thereby providing robust technical support for safety management and risk prevention.

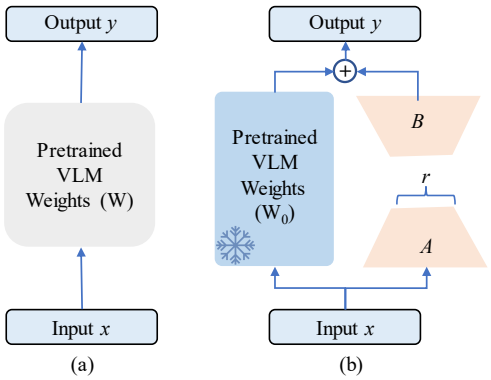


Figure 5. Comparison of two data processing workflows: (a) direct use of pre-trained weights, (b) LoRA-based fine-tuning process

2.3 Construction site hazard identification

The fine-tuned VLM is deployed on the server side, as illustrated in Fig. 1. The system interfaces with a CCTV monitoring system at the construction site to acquire live or stored video streams. To optimize computational efficiency and reduce processing load, key frames are selected at fixed time intervals. For each extracted frame, the VLM extracts rich visual features and semantic information relevant to the construction scene.

Based on these extracted features, the VLM classifies images to identify whether they contain categories of construction hazards that require inspection, as outlined in Table 1. Identified hazards, along with associated metadata such as location, type, and timestamp, are aggregated into a hazard identification dataset. Leveraging an LLM, the system can automatically generate a structured report on construction hazards based on this aggregated data. The report includes detailed textual descriptions, and relevant image evidence may be embedded to facilitate a more intuitive understanding for site managers of the conditions. Through this process, the fine-tuned VLM significantly contributes to real-time monitoring and alerting of potential hazards on the construction site, thereby enhancing the smart level of safety management.

3 Experiment and Result

3.1 Experiment setup

Based on the constructed construction hazard dataset, this study employs the LoRA technique to fine-tune the

VLM. Specifically, two different parameter-size versions of Qwen2-VL are selected for this purpose: a 2-billion-parameter model and a 7-billion-parameter model. Specialized fine-tuning and validation experiments are conducted for both versions. During the testing phase, Qwen2.5-32B LLM is utilized to assist in evaluating the match between the fine-tuned VLM predictions and the actual labels. This evaluation ensures that the VLM's performance is accurately assessed and optimized for real-world applications.

To establish an effective benchmark for comparison, the CLIP model with ViT-B/32 as the image encoder is used as a baseline for classifying construction hazard images. The process involves fine-tuning the CLIP model using image-label pairs from the training dataset. During the validation phase, similarity scores are calculated between the images in the test set and all possible hazard labels. If the similarity score exceeds a predefined threshold d, the corresponding label is considered a prediction, thereby evaluating model performance. This approach aligns with the methodology proposed by Chen et al. [20].

For evaluating the models' performance, a series of widely accepted evaluation metrics are applied. Precision (the proportion of correctly identified positive samples among those predicted as positive), recall (the proportion of actual positives which are correctly identified as such), and F1-score (the harmonic mean of precision and recall) are calculated for each category using the following formulas:

$$Precision = \frac{TP}{TP + FP}, \tag{2}$$

$$Recall = \frac{TP}{TP + FN}, \tag{3}$$

$$F1\ score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}, \tag{4}$$

where TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives. Additionally, the F2 score (β = 2) is introduced to emphasize recall, focusing more on minimizing missed identifications. It is calculated using the following formula:

$$F2\ score = (1+\beta^2) \cdot \frac{Precision \cdot Recall}{(\beta^2 \cdot Precision) + Recall}. \tag{5}$$

This is critical given the serious consequences of safety accidents at construction sites. The F2 score places greater weight on recall compared to precision, ensuring that fewer potential hazards are overlooked. Moreover, Cohen's Kappa coefficient is used to measure the agreement between observed and expected consistency, applying the formula:

$$\kappa = \frac{p_o - p_e}{1 - p_e}, \tag{6}$$

where p_o is the observed proportion of agreement and p_e is the expected probability of random agreement [21]. This coefficient verifies whether the model predictions significantly outperform random guessing, thereby increasing confidence in the model's reliability. A Cohen's Kappa coefficient value close to 1 indicates almost perfect agreement, suggesting that the model significantly outperforms random chance. In contrast, values near or below 0 indicate agreement no better than, or even worse than, what would be expected by random chance, highlighting potential limitations in the model's ability.

3.2 Result

As illustrated in Figure 6, the performance of the fine-tuned Qwen2-VL-2B and Qwen2-VL-7B models in identifying construction hazards as a function of Epoch is depicted. The figure demonstrates that, at the outset (Epoch = 0), prior to fine-tuning, both models exhibit lower precision, recall, and F1 scores. As the number of Epochs increases, there is a marked improvement in model performance. Specifically, the precision for both the Qwen2-VL-7B and Qwen2-VL-2B models rises from approximately 0.4 to around 0.8 following fine-tuning. The comparison with their no fine-tuning performance can be seen in Table 2. Notably, the fine-tuned Qwen2-VL-7B model achieves a precision of 0.856 by the 9th Epoch, while the fine-tuned Qwen2-VL-2B model reaches a precision of 0.818 by the 27th Epoch. Similarly, recall values for both models increase from about 0.2 to approximately 0.6, and F1 scores rise from roughly 0.2 to nearly 0.7. These observations indicate that fine-tuning significantly enhances the models' ability to identify construction hazards. Furthermore, it is evident from the figure that, in their initial state, the Qwen2-VL-7B model in recall and F1 score outperform the Qwen2-VL-2B model. However, as the number of fine-tuning Epochs increases, the performance gap between the two models narrows, suggesting that with sufficient fine-tuning, the smaller model can achieve performance comparable to that of the larger model. This finding underscores the importance of fine-tuning strategies in optimizing model performance, highlighting that even smaller-scale models can attain significant improvements when subjected to effective fine-tuning procedures.

The loss curve depicted in Figure 7 illustrates the trend of the loss value as a function of epochs during the training process of the CLIP model. Initially, the loss value is relatively high but gradually decreases over time before eventually plateauing. Meanwhile, the recall curve remains close to 1 for most of the training period, indicating near-perfect identification of positive samples

by the model. However, it exhibits a decline in later stages, suggesting that the model's ability to accurately identify positive samples diminishes over extended training. Despite undergoing fine-tuning, the CLIP model does not demonstrate significant improvements in precision or F1 score, as shown in Table 2. The consistently low precision indicates that the model struggles to effectively differentiate construction hazards, while the low F1 score reflects an overall deficiency in balancing precision and recall. This suggests that the model fails to fully comprehend and accurately distinguish the various features associated with construction hazards. Therefore, when applied to the task of recognizing construction hazards, the pre-trained CLIP model, despite achieving high recall, exhibits limitations in enhancing precision and the F1 score.

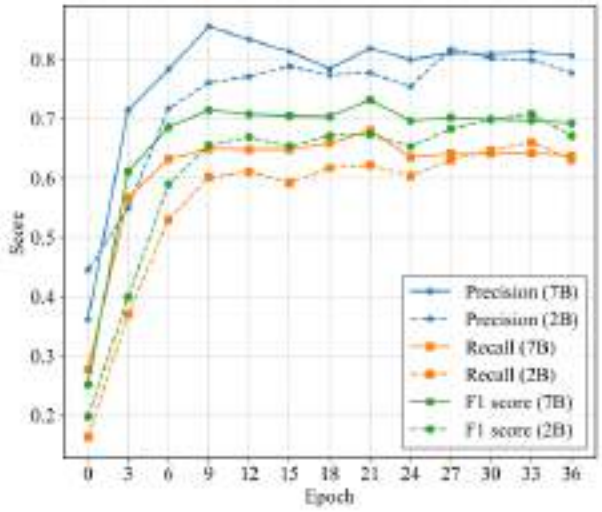


Figure 6. Comparison of Qwen2-VL-2B and 7B's construction hazard identification performance as a function of Epoch

Table 2 also presents the F2 scores and Cohen's Kappa coefficients for various models. The table compares two primary models: CLIP (ViT-B/32), and Qwen2-VL in both 2B and 7B, each evaluated in terms of their pre-trained and fine-tuned performance on the test dataset. For the CLIP model using the ViT-B/32 as the image encoder, both the F2 score and Cohen's Kappa coefficient remain nearly unchanged, even after fine-tuning, indicating that the fine-tuning process did not yield significant improvements for this model. In contrast, the Qwen2-VL models exhibit substantial gains following fine-tuning. Initially, the Qwen2-VL-2B has an F2 score of 0.186 and a Cohen's Kappa of 0.118, which significantly improves to an F2 score of 0.661 and a Cohen's Kappa of 0.613 after fine-tuning. Similarly, the Qwen2-VL-7B starts with an F2 score of 0.290 and a Cohen's Kappa coefficient of 0.194, which then increases to an F2 score of 0.683 and a Cohen's Kappa of 0.600 after fine-tuning, with respective increases of 0.393 and 0.406. These results underscore the effectiveness of fine-tuning in enhancing the performance of the Qwen2-VL models, particularly in terms of improving both precision-recall balance.

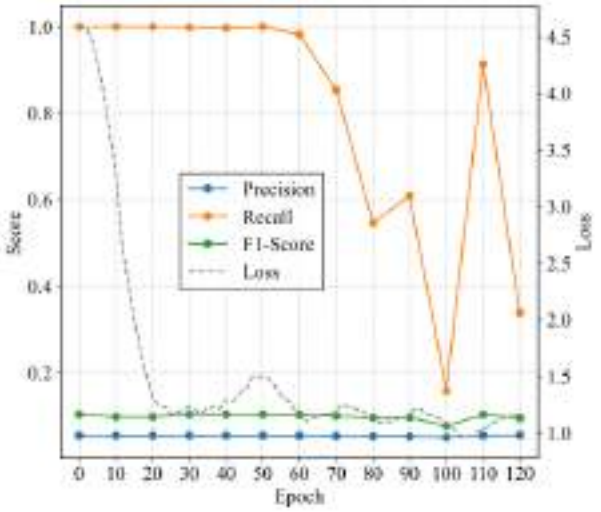


Figure 7. Performance and loss curves for construction hazard identification based on fine-tuned CLIP as a function of Epoch

Table 2. Performance comparison of different models before and after fine-tuning

Models		Precision	Recall	F1 score	F2 score	Cohen's Kappa coefficient
CLIP	ViT-B/32	0.055	**1.0**	0.104	0.225	0
	ViT-B/32 (Fine-tuned)	0.055 (+0.0)	0.999(-0.001)	0.104(+0.0)	0.225(+0.0)	0(+0.0)
Qwen 2-VL	2B	0.445	0.163	0.198	0.186	0.118
	2B (Fine-tuned)	0.818 (+0.373)	0.631 (+0.468)	0.683 (+0.485)	0.661 (+0.475)	**0.613** (0.495)
	7B	0.361	0.277	0.252	0.290	0.194
	7B (Fine-tuned)	**0.856** (+0.495)	0.650 (+0.373)	**0.715** (+0.463)	**0.683** (+0.393)	0.600 (+0.406)

Note: Values in parentheses indicate the change after model fine-tuning compared to before.

The effectiveness of the proposed method in identifying construction safety hazards is illustrated using real-world images extracted from an actual construction site video, with one of the extracted images shown in Figure 8. The selected construction site is a stockpile area, where our approach demonstrates its capability to accurately detect potential construction hazards present in the images and generate corresponding hazard reports for further documentation. This application highlights the practical utility of our method in a real-world setting, showcasing its ability to provide reliable hazard identification. Such improvements are crucial for preventing accidents and ensuring worker safety, thereby underscoring the significant impact and value of this method in advancing construction site management.

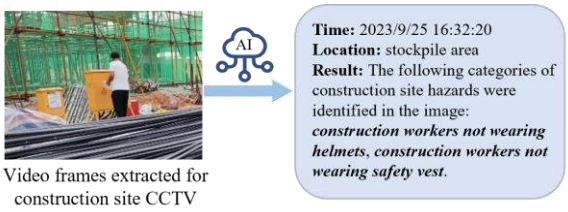


Figure 8. A case of identifying construction hazards

4 Discussion and Conclusion

This study introduces a fine-tuned VLM designed to comprehensively identify safety hazards on construction sites. Specifically, our work encompasses the following aspects: (1) we developed a specialized dataset for construction hazards, classifying 31 types of common hazards at a fine-grained level. (2) we demonstrated the effectiveness of using pre-trained VLMs and fine-tuning techniques for identifying construction hazards. Additionally, we compared the performance differences among various pre-trained VLMs in the context of construction hazard identification. (3) we implemented an automated end-to-end process for construction site hazard inspections, ranging from image acquisition to report generation, thereby enhancing the automation level of construction management.

The experimental results indicate that the fine-tuning process significantly enhances the VLM model's ability to handle complex and dynamic scenarios, improving its generalization capability and robustness. Through enhanced cross-modal understanding, the precision of the fine-tuned Qwen2-VL-7B reaches 0.856, marking a substantial improvement from 0.495 precision in the non-fine-tuned state. The use case of real-world site videos further demonstrates its future application potential in construction management, such as being deployed on robotic dogs or drones to facilitate the development of automated construction hazard inspections.

Moreover, the limitations of traditional VLMs in comprehending construction scenarios underscore the necessity for advanced VLM modeling applications. Our study also revealed that even VLM models with fewer parameters can achieve recognition results comparable to those of larger VLM models when appropriately fine-tuned. This insight suggests that future research should focus on developing specialized, compact models to achieve high-precision construction hazard recognition while using minimal computational resources.

Although the proposed method has shown promising performance, there are two limitations. First, we conducted a preliminary application on a real construction site to assess its usability. However, this case study was limited in scale and environmental control, warranting future systematic validation. Moreover, the study not analyse the model's performance for each category, further investigation, particularly in exploring the model's robustness, few-shot learning capabilities, and speed comparisons will conduct in future work.

Acknowledgements

The authors gratefully acknowledge the financial support provided by the National Natural Science Foundation of China (Grant No. 72404233), the Guangdong Basic and Applied Basic Research Foundation (Grant No. 2025A1515010190) and the New Faculty Start-up Grant from the City University of Hong Kong (Project No. 9610701).

The authors also thank Kaisheng Liu, a construction safety manager, for his assistance in processing and categorizing the raw construction hazard images.

References

[1] B. Choi, S. Lee. An empirically based agent-based model of the sociocognitive process of construction workers' safety behavior. *Journal of Construction Engineering and Management*, 144 (2): 04017102, 2018. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001421

[2] B. Yuan, S. Xu, M. Niu, K. Guo. Will improved safety attitudes necessarily curb unsafe behavior? Hybrid method based on NCA and SEM. *Journal of Environmental and Public Health*, 2022 (1): 9271690, 2022. https://doi.org/10.1155/2022/9271690

[3] Q. Chen, D. Long, C. Yang, H. Xu. Knowledge graph improved dynamic risk analysis method for behavior-based safety management on a construction site. *Journal of Management in Engineering*, 39 (4): 04023023, 2023. https://doi.org/10.1061/JMENEA.MEENG-5306

[4] B. Xiao, Y. Zhang, Y. Chen, X. Yin. A semi-supervised learning detection method for vision-based monitoring of construction sites by integrating teacher-student networks and data augmentation, *Advanced Engineering Informatics*. 50 101372, 2021. https://doi.org/10.1016/j.aei.2021.101372.

[5] B. Xiao, X. Yin, S.-C. Kang. Vision-based method of automatically detecting construction video highlights by integrating machine tracking and CNN feature extraction. *Automation in Construction*, 129 103817, 2021. https://doi.org/10.1016/j.autcon.2021.103817.

[6] Q. Chen, D. Long, S. Wang, Q. Chen, B. Yuan. Real-Time Detection of Personal Protective Equipment Violations for Construction Workers Using Semi-Supervised Learning and Video Clips. *Journal of Construction Engineering and Management*, 2025. https://doi.org/10.1061/JCEMD4/COENG-15310.

[7] X. Mei, X. Zhou, F. Xu, Z. Zhang. Human intrusion detection in static hazardous areas at construction sites: Deep learning–based method. *Journal of Construction Engineering and Management*, 149 (1): 04022142, 2023. https://doi.org/10.1061/(ASCE)CO.1943-7862.0002409

[8] Y.G. Lim, J. Wu, Y.M. Goh, J. Tian, V. Gan. Automated classification of "cluttered" construction housekeeping images through supervised and self-supervised feature representation learning. *Automation in Construction*, 156 105095, 2023. https://doi.org/10.1016/j.autcon.2023.105095.

[9] A. Radford, J.W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark. Learning transferable visual models from natural language supervision. International conference on machine learning, PMLR, 2021, pp. 8748-8763. https://proceedings.mlr.press/v139/radford21a.

[10] J. Li, D. Li, C. Xiong, S. Hoi. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. International conference on machine learning, PMLR, 2022, pp. 12888-12900. https://proceedings.mlr.press/v162/li22n.html.

[11] H. Liu, C. Li, Q. Wu, Y.J. Lee. Visual instruction tuning, *Advances in neural information processing systems*. 36, 2024.

[12] W.-L. Tsai, P.-L. Le, W.-F. Ho, N.-W. Chi, J.J. Lin, S. Tang, S.-H.J.A.i.C. Hsieh. Construction safety inspection with contrastive language-image pre-training (CLIP) image captioning and attention. *Automation in Construction*, 169 105863, 2025. https://doi.org/10.1016/j.autcon.2024.105863

[13] D. Gil, G. Lee. Zero-shot monitoring of construction workers' personal protective equipment based on image captioning. *Automation in Construction*, 164 105470, 2024. https://doi.org/10.1016/j.autcon.2024.105470.

[14] G. Yong, M. Liu, S. Lee. Explainable Image Captioning to Identify Ergonomic Problems and Solutions for Construction Workers. *Journal of Computing in Civil Engineering*, 38 (4): 04024022, 2024. https://doi.org/10.1061/JCCEE5.CPENG-5744

[15] Y. Liang, P. Vadakkepat, D.K.H. Chua, S. Wang, Z. Li, S. Zhang. Recognizing temporary construction site objects using CLIP-based few-shot learning and multi-modal prototypes. *Automation in Construction*, 165 105542, 2024. https://doi.org/10.1016/j.autcon.2024.105542.

[16] Y. Ouali, A. Bulat, A. Xenos, A. Zaganidis, I.M. Metaxas, G. Tzimiropoulos, B.J.a.p.a. Martinez. Discriminative Fine-tuning of LVLMs. *arXiv preprint arXiv:2412.04378*, 2024. https://doi.org/10.48550/arXiv.2412.04378

[17] P. Wang, S. Bai, S. Tan, S. Wang, Z. Fan, J. Bai, K. Chen, X. Liu, J. Wang, W. Ge. Qwen2-vl: Enhancing vision-language model's perception of the world at any resolution. *arXiv preprint arXiv:.12191*, 2024. https://doi.org/10.48550/arXiv.2409.12191

[18] P.-C. Lee, J. Wei, H.-I. Ting, T.-P. Lo, D. Long, L.-M. Chang. Dynamic analysis of construction safety risk and visual tracking of key factors based on behavior-based safety and building information modeling. *KSCE Journal of Civil Engineering*. 23 (10): 4155-4167, 2019. https://doi.org/10.1007/s12205-019-0283-z.

[19] E.J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, W. Chen. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:.09685*, 2021. https://arxiv.org/pdf/2106.09685v1/1000

[20] Z. Chen, H. Chen, M. Imani, R. Chen, F. Imani. Vision language model for interpretable and fine-grained detection of safety compliance in diverse workplaces. *Expert Systems with Applications*. 265 125769, 2025. https://doi.org/10.1016/j.eswa.2024.125769.

[21] M. Ferdous, S.M.M. Ahsan. PPE detector: a YOLO-based architecture to detect personal protective equipment (PPE) for construction sites. *PeerJ Computer Science*, 8 e999, 2022. https://doi.org/10.7717/peerj-cs.999.

Performance Evaluation Model to Enhance Life Cycle Productivity in Modular and Off-Site Construction: A Process-Oriented Approach

Babak Manouchehri [a], Sang Hyeok Han [a], and Fuzhan Nasiri [ap]

[a] Department of Building, Civil and Environmental Engineering, Concordia University, Montréal, Canada
Email: Babak.manouchehri@mail.concordia.ca, sanghyeok.han@concordia.ca, fuzhan.nasiri@concordia.ca

Abstract

The construction sector has long faced low productivity relative to other industries. Modular and off-site construction (MOC) offers a promising solution by assembling modules in a controlled facility environment and benefiting from repetitive nature of operations and the ability to conduct processes simultaneously. These features increase process flow continuity and enhance productivity across the life cycle. However, MOC has not yet realized its full potential, mainly due to the influence of conventional construction mindsets and the ineffective integration of lean practices, leading to process disruptions and persistent waste throughout the life cycle. To address this, a process-oriented performance evaluation (PE) model is introduced, using system dynamics to model life cycle processes and the existing interactions among different life cycle stages. This approach systematically evaluates life cycle productivity by accounting for the state of utilized resources (workers and workspace) and employing proposed MOC key performance indicators (MOC-KPIs) to identify the root causes of process waste with respect to resources utilization.

Keywords

Off-site construction, Key performance indicators, System dynamics, Productivity evaluation, Process waste.

1 Introduction

Over the past two decades, global construction productivity has remained stagnant, increasing by only 10% (0.4% annually) between 2000 and 2022, while the global economy and manufacturing sectors grew by 50% and 90%, respectively. Productivity further declined by 8% between 2020 and 2022, causing significant value loss, project delays, and cost overruns [1]. Despite rising demand for housing, the industry struggles with workforce shortages, escalating costs, and affordability challenges. Rising inflation and interest rates have exacerbated the disparity between housing costs and incomes, with homeownership now consuming up to 54% of household income in Canada [2]. Improving life cycle productivity is critical to addressing these challenges and enhancing housing availability. MOC, which shifts traditional on-site activities to controlled off-site environments, offers potential solutions by adopting lean manufacturing principles. It aims to boost productivity, quality, safety, and cost efficiency [3, 4]. However, MOC adoption remains limited, holding only 6.03% market share in 2022, and faces challenges such as high installation and transportation costs and inefficient lifecycle performance [3]. Many stakeholders report higher-than-expected costs compared to conventional methods, limiting MOC's ability to achieve widespread cost reductions and productivity improvements [5, 6].

MOC leverages lean manufacturing principles to reduce costs, minimize waste, and maintain quality. While research has demonstrated its potential in applications like high-rise buildings and manufactured housing [8], MOC often fails to fully benefit from lean practices. Conventional methodologies and fluctuating market demands requiring complex customization have hindered its adoption, leaving processes dominated by traditional construction mindsets [9]. These issues result in unbalanced lifecycle workflows and inefficiencies unique to MOC, such as repetitive operations and simultaneous activities across workstations. Current PE tools like the Balanced Scorecard, conventional construction KPIs, and Earned Value Management fail to effectively assess MOC projects [10]. Their limitations include poor tracking of concurrent activities, inadequate resource utilization assessment, inability to identify waste causes, and failure to address logistics challenges critical to MOC's success. These gaps impede precise PE and hinder the industry's shift toward more efficient, lean-driven practices. This paper introduces a process-oriented PE approach, using System Dynamics Modeling (SDM) to assess MOC lifecycle productivity through proposed MOC-KPIs in order to identify the root causes

of process waste. SDM is a simulation-based approach used to understand and analyze complex systems over time. It employs feedback loops, stock-and-flow structures, and time delays to model real-world processes dynamically. The approach involves three key steps: (1) proposing KPIs to analyze resource utilization and identify process waste; (2) developing an SDM simulation to map MOC lifecycle dynamics and evaluate current-state life cycle productivity; and (3) testing intervention scenarios to validate the KPIs' effectiveness in identifying waste and improving life cycle process efficiency.

2 Literature Review

2.1 Measuring progress in conventional construction

Progress measurement involves quantifying the effectiveness and efficiency of actions to assess a project's current state [11]. As projects grow in size and complexity, they require more resources, generating vast amounts of data [12]. To obtain accurate performance outcomes, it is essential to analyze this data using specific indicators to identify root causes of problems promptly [14]. KPIs play a crucial role in improving project efficiency by supporting decision-making, identifying areas for improvement [15]. Literature reveals that KPIs in construction and related fields (e.g., manufacturing, road transportation) often focus on sustainability and productivity [16, 17]. These efforts primarily introduce new indicators for PE in areas such as time and cost. Mellado et al. [18] ranked commonly used KPIs, while Orgut et al. [11] identified factors to enhance project control reliability at different phases of the construction lifecycle. Habibi et al. [19] also explored phase-based KPIs and their impact on cost and schedule. However, due to the project-oriented nature of conventional construction, KPIs vary significantly between projects, making performance comparisons and future forecasts challenging. Efforts to reduce inefficiencies on construction sites have led to the development of various performance metrics to improve progress measurement [20, 21].

Furthermore, conventional KPIs related to lean manufacturing, road transportation, and on-site construction, focusing on productivity and process flow, are investigated. For example, KPIs include defect per unit which tracks errors in products, aiming to reduce rework and scrap costs [22]; overall equipment effectiveness that measures manufacturing equipment efficiency to enhance productivity [22]; loading/unloading time, focusing on minimizing delays in transportation operations [7, 13]; and on-site rework rate which measures defects and reworks during construction to minimize delays and the related costs [23, 24]. However, these KPIs may not fully address the unique needs of MOC projects, such as work sequence and resource flow, and may overlook key factors like laydown area availability. They may also struggle to accommodate MOC's dynamic nature, where disruptions in manufacturing (e.g., material shortages) affect on-site installation. Additionally, they may not pinpoint root causes of waste (e.g., rework due to module misfit or damage during shipment) or provide a complete view of resource allocation, especially for tasks like module repairs. Furthermore, traditional road transportation KPIs often fail to capture transportation processes across the MOC life cycle, such as the duration modules spend in staging areas before shipment.

2.2 Evaluating productivity in MOC

Effective resource planning is essential in construction management to meet project constraints and maintain process flow, as its absence can lead to productivity losses [25]. Concepts such as optimal-size capacity buffers and multi-skilled resources have been proposed to improve flexibility and productivity [25, 26]. Using multi-skilled resources enhances process integration and production efficiency [26]. Arashpour et al. [27] examined various cross-training strategies in MOC, finding they address capacity imbalances and time variability, measured by indicators like throughput rate and cycle time. However, these strategies are costly due to workforce turnover and resource availability [28].

Arashpour et al. [28] proposed a customized methodology for tracking the MOC production workflow, using discrete event simulation (DES) to reveal that production output is nonlinear due to variability in the MOC environment. To address this, they recommended setting a capacity buffer rather than aligning production targets with capacity. Manouchehri [29] applied a dynamic productivity model using SDM to MOC, estimating productivity in labor cost per ton of module installed on-site. By modeling the impact of performance factors (e.g., waiting times for transit) through causal loop diagrams and Stock and Flow diagrams, the model predicted future productivity and provided more accurate estimates for integrated project phases. Wang et al. [30] proposed an optimization model for the precast concrete component production line, incorporating lean production theory and Value Stream Mapping (VSM). They developed current- and future-state value maps and introduced KPIs such as delivery cycle, value-added time ratio, number of personnel, and production line balance rate. The model demonstrated improvements in production line balance, reduced delivery cycle, decreased value-added time ratio, and fewer workers. Similarly, Garza-Reyes et al. [17] developed a methodology combining time-based study, DES, and

trial and error to optimize line balancing in manufacturing. The study identified workload distribution and process flow performance, with DES models assessing different scenarios to determine the optimal balance. They introduced the "balancing loss" indicator to measure idle time, achieving reductions in balancing losses ranging from 1.82 to 36.32 percent. The KPIs developed for MOC are listed in Table 1.

Table 1. Overview of existing KPIs for assessing productivity and resource allocation in MOC

MOC Phase	MOC KPI
Manufacturing	Workforce density
	Workload density
	Production line balance rate
	Value-added time ratio
	Balancing loss

In reviewing previous studies, several key gaps were identified in the evaluation of MOC performance: (1) Many studies relied on conventional construction KPIs that fail to account for the unique workflow and sequencing of MOC, leading to ineffective assessments of productivity and resource utilization; (2) Most studies focused solely on the manufacturing phase, neglecting the integrated nature of MOC across all life cycle stages; (3) They often overlooked on-site installation dynamics, such as site-specific constraints; (4) The interaction between life cycle processes, like delays in module quality inspections affecting on-site installation, was not adequately considered; (5) There is a lack of methodologies that compare actual performance with predefined benchmarks, hindering performance comparisons across projects; and (6) Existing methods, like VSM, do not effectively address concurrent activities or identify root causes of inefficiency in MOC life cycle performance. Given the differences between off-site and on-site construction methods, MOC should be treated as a process-oriented approach, leveraging repetitive operations throughout its life cycle. As a result, the literature supports the need for a dedicated PE framework for MOC, which should: (1) adopt a process-oriented approach that supports repetitive operations and the distinct flow of MOC processes; (2) incorporate MOC life cycle and resource leveling to identify waste and improve efficiency; and (3) include KPIs to assess MOC life cycle performance.

3 Proposed Method

This paper introduces a process-oriented approach to systematically evaluate the MOC life cycle productivity (MOC-LCP), using proposed MOC-KPIs. The approach aims to help production managers efficiently identify root causes of process waste related to resource utilization (e.g., workforce and workspace). As illustrated in Figure 1, the methodology involves three key steps: (1) developing MOC-KPIs tailored to various life cycle phases to measure resource utilization (e.g., workforce); (2) creating a simulation model integrated with these KPIs to assess life cycle productivity while considering resource allocation; and (3) run intervention scenarios using the simulation model to analyze the effectiveness of the proposed MOC-KPIs in identifying root causes of waste.

3.1 MOC-KPIs proposal

The proposed MOC-KPIs enable a detailed analysis of resource utilization to identify inefficiencies and the root causes of process waste. Grounded in lean principles, the approach aims to enhance life cycle efficiency through continuous improvement and waste minimization, transforming MOC into a more efficient and cost-effective construction method. While tools like VSM offer useful metrics for performance measurement [9], they have notable limitations: (1) an incomplete set of metrics for systematically assessing life cycle performance; (2) an inability to capture complex interdependencies among life cycle processes, such as the ripple effects of delays across stages; (3) insufficient capacity to model and analyse parallel workflows that require precise coordination (e.g., interactions between manufacturing and site operations); and (4) limited capability to guide investigations into root causes of waste (e.g., rework) to improve the efficiency and effectiveness of operations, particularly regarding resource utilization (e.g., workforce and workspace).

To propose KPIs aligned with the MOC life cycle processes, the delivery flow of a single module at the MOC system level was analyzed. A detailed process flow was constructed using insights from MOC literature, published MBI reports, and interviews with industry professionals (e.g., Stack Modular, MODLOGIQ, Smart Modular Canada). However, the MOC life cycle process is prone to disruptions that introduce significant waste, hinder operational efficiency, and increase resource utilization, thereby reducing MOC-LCP and escalating project costs. To address these challenges, MOC-KPIs are essential for identifying, measuring, and diagnosing the root causes of process waste. They enable recognizing issues, quantifying their impact, and analyzing contributing factors, facilitating the timely detection of inefficiencies. This improves resource utilization across life cycle stages, streamlines process flow, and supports continuous improvement. Ultimately, these efforts enhance MOC-LCP and lower overall project costs.

The proposed KPIs include workflow interruption (WFI), module defect modification (MDM), module quality inspection release (MQIR), staging area

occupancy (SAO), at-factory module loading (AFML), module travelling time (MTT), shipment damage inspection (SDI), module lift handling (MLH), module in-place alignment (MIPA), module connection work (MCW), and site-fit rework (SFR).

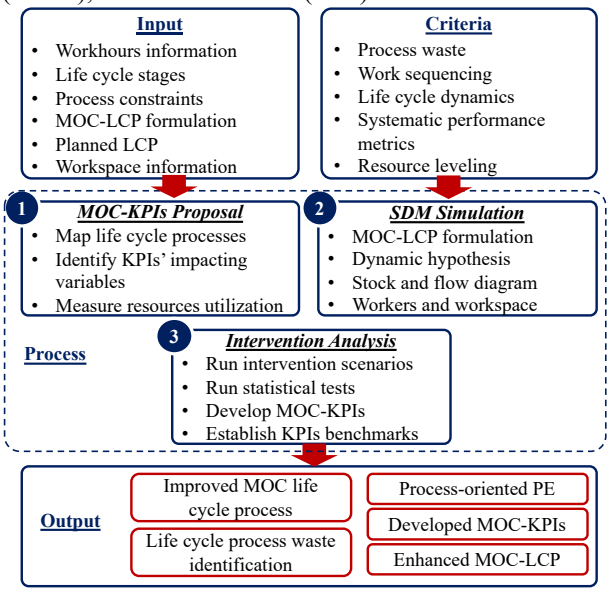


Figure 1. MOC-LCP evaluation process

3.2 SDM Simulation

Simulation is recognized as an effective technique for tasks such as productivity assessment and resource planning. However, before developing a life cycle simulation model, it is essential to formulate an equation capable of efficiently estimating current-state productivity in MOC. The MOC-LCP is represented by Equation 1.

$$MOC - LCP = \frac{S_{inst}}{T_{MOC}} \quad (1)$$

Where S_{inst} = surface area of all modules installed on-site per square feet; T_{MOC} = total MOC project spent workhours.

Additionally, further literature review was conducted to identify factors influencing the proposed KPIs and constraints affecting module delivery within the modular project life cycle. Further, SDM is used which is an efficient tool that enables evaluating a system's ability in coping with the incurred changes (i.e., most-contributing KPIs functionality), while examines new scenarios that could be implemented within the system's environment (i.e., life cycle process) for the sake of effective decision-making process (i.e., identification of root cause of process waste with respect to resources utilization). To better understand the interrelated processes within the MOC life cycle and visually depict the dependencies among various KPIs and their parameters, a causal loop diagram of the module delivery process flow model is developed [20]. (Figure 2).

Moreover, building on the established CLD, a stock and flow diagram (SFD) is subsequently formulated. It is divided into five interconnected sub-SFDs to model process of module (*i*) manufacturing, including module assembly and quality inspection; (*ii*) transportation; (*iii*) on-site installation; and (*iv*) MOC-LCP assessment. To effectively represent the dynamic relationships among various life cycle processes, additional key components and their interactions are integrated into the SFD diagram. These relationships are articulated through mathematical equations and a set of defined rules.

3.3 Intervention Analysis

The objective of scenario analysis is to evaluate the impact of various interventions on a system's KPIs [25]. This involves several key steps: (*i*) identifying interventions and their boundaries, with the proposed MOC-KPIs serving as response variables, and employing the developed and validated SDM to examine the behavior of MOC-KPIs under each scenario; (*ii*) performing statistical tests to assess whether the interventions significantly affect overall life cycle performance, explore relationships among the proposed MOC-KPIs, evaluate the combined effects of interventions, and test hypotheses regarding the impact of each intervention on individual MOC-KPIs; and (*iii*) analyzing MOC-KPIs' behavior across different interventions to verify which one efficiently identifies root causes of process waste within each phase of the MOC. To conduct intervention analysis, various scenarios, as shown in Table 2, were identified through literature review and interviews with MOC experts.

Table 2. Scenarios studied for intervention analysis

Intervention type	Intervention scenario
Manufacturing (assembly and quality inspection)	S1: Ramp up module production S2: Apply overtime policy S3: Reduce rework during module assembly and quality inspection S4: Reduce quality inspection space unavailability
Transportation (truck-trailer and crane)	S5: Increase transportation fleet availability S6: Timely crane preparation at factory S7: Reduce lack of space on laydown area
On-site installation	S8: Timely crane preparation S9: Speed up lift preparation S10: Boost module alignment

	process
	S11: Accelerate module connection
	S12: Reduce site-fit rework

4 Case Example

A residential modular project located in Kenora, Ontario, was selected as the case study. The project involved the construction of a single building using six wooden-framed modules, manufactured by Smart Modular Canada. These modules, each weighing 14 tons and measuring 14 feet by 50 feet, were fabricated at a factory in Thunder Bay, Ontario, and fully equipped with mechanical, electrical, and plumbing systems, as well as internal finishes. Although the installation of the final module was initially planned to be completed within 242 workhours, the process ultimately required 268 workhours due to various challenges encountered throughout the project life cycle, including detected defects during assembly, delays in truck availability, and damages sustained during shipment.

Moreover, the scenarios were simulated using the SDM, and the resulting MOC-KPI data were analysed to assess normality. This analysis employed Quantile-Quantile (Q-Q) plots, kurtosis, skewness, and the Shapiro-Wilk test. Q-Q plots were constructed for each MOC-KPI to visualize their distribution (11 plots). Figure 3 presents the Q-Q plots for the manufacturing proposed KPIs (module assembly and inspection), namely, WFI, MQIP, and MDM. The Q-Q plots reveal that the datasets for each MOC-KPI deviate from normality, as the data points cluster away from and do not align closely with the 45-degree reference line. To provide a more comprehensive assessment of normality, additional statistical measures, including kurtosis, skewness, and the Shapiro-Wilk test, were employed, as summarized in Table 3. The skewness and kurtosis values exceeded the acceptable thresholds for some MOC-KPIs, thereby violating the normality assumption (thresholds: −2 to +2 for skewness and −7 to +7 for kurtosis). These values are highlighted in Table 3. Furthermore, the Shapiro-Wilk test yielded *p*-values below the significance level of 0.05, leading to the rejection of the null hypothesis (normal distribution). This result corroborates that the dataset deviates significantly from normality. Therefore, non-parametric statistical tests were subsequently employed at this stage of the analysis.

Table 3. Statistical measures and test for data normality assessment

Independent variables (MOC-KPIs)	Skewness	Kurtosis	Shapiro-Wilk *p*-value
WFI	**-0.700**	7.470	2.008×10^{-9}
MQIR	2.478	12.133	1.591×10^{-9}
MDM	4.885	27.870	4.132×10^{-11}
WFD	4.401	25.909	1.693×10^{-11}
SAO	**-0.211**	**4.552**	1.430×10^{-8}
AFML	5.375	31.888	3.577×10^{-12}
MTT	4.115	21.327	1.677×10^{-10}
SDI	3.861	16.086	3.437×10^{-9}
MLH	2.787	**6.999**	2.014×10^{-10}
MIPA	3.337	21.307	3.154×10^{-11}
MCW	5.388	32.000	2.794×10^{-12}

Spearman's Rank Correlation Coefficient analysis was conducted to examine the strength and direction of relationships among the proposed MOC-KPIs, as summarized in Table 4. The results reveal an acceptable level of linear relationship between certain MOC-KPIs ($0.2<r<0.9$). However, the correlation between some variables, such as MDM and MTT is less than 0.2, indicating a very weak and negligible relationship. The strongest correlation is observed between AFML and MIPA (r=0.64). Interestingly, while MIPA shows no correlation with other KPIs (r = 0.0), it exhibits the second-highest correlation with AFML (r = 0.642). These correlations, though weak in some cases, underscore the interconnected nature of various processes across the MOC life cycle, highlighting the factors and MOC-KPIs that are influenced by these operations. Furthermore, since all coefficients fall within the acceptable range, the proposed MOC-KPIs are deemed suitable for inclusion in the developed model, supporting their potential relevance in capturing the dynamics of the life cycle processes.

At this stage, quantile regression analysis was utilized to assess the robustness of the proposed MOC-KPIs and to define a reliable framework for identifying those that significantly contribute to MOC-LCP. The analysis involved multiple iterative rounds, during which the *p*-value for each MOC-KPI was computed. KPIs with *p*-values exceeding 0.05, signifying a lack of statistical significance, were systematically excluded from subsequent iterations. This process continued until all remaining MOC-KPIs demonstrated statistical significance, confirming their importance to MOC-LCP and their effectiveness in identifying and evaluating root

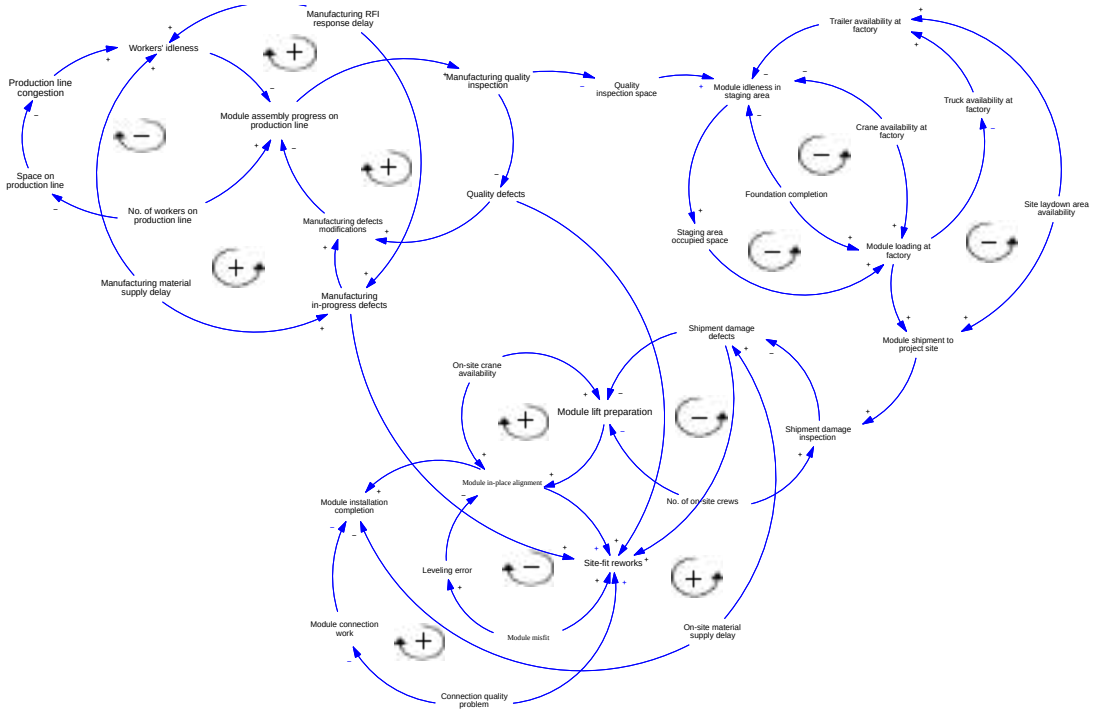


Figure 2. The causal loop diagram of MOC process flow to evaluate life cycle productivity

Table 4. Correlation between MOC-KPIs using Spearman Rank Correlation Coefficient

	WFI	MDM	MQIR	SAO	AFML	MTT	SDI	MLH	MIPA	MCW	SFR
WFI	1										
MDM	0.0	1									
MQIR	0.0	**0.49**	1								
SAO	0.0	-0.03	-0.07	1							
AFML	0.0	**0.6**	**0.5**	-0.03	1						
MTT	0.0	-0.01	-0.02	0.23	-0.01	1					
SDI	0.0	-0.02	-0.04	-0.44	-0.02	-0.02	1				
MLH	0.0	-0.03	-0.07	-0.10	-0.03	-0.03	-0.06	1			
MIPA	0.0	0.00	0.00	0.00	**0.64**	0.00	0.00	0.0	1		
MCW	0.0	-0.02	-0.04	-0.06	-0.02	-0.02	-0.03	-0.06	0.0	1	
SFR	0.0	-0.01	-0.02	-0.03	-0.01	-0.01	-0.02	-0.03	0.0	**-0.45**	1

causes of process waste with respect to resource utilization. The outcomes of the quantile regression analysis are detailed in Table 5.

Table 5. Results from multiple iterations of quantile regression analysis (pp-value)

MOC-KPI	*p*-value				
	R1	R2	R3	R4	R5
WFI	0	0	0	0	0
MQIR	0	**0.586**	-	-	-
MDM	0	0	0	0	0
WFD	0	0	0	0	0
SAO	0	0	0	0	0
AFML	0	0	0	0	0
MTT	0	0	0	0	0
SDI	**0.596**	-	-	-	-
MLH	0	0	0	**0.999**	-
MIPA	0	0	**0.991**	-	-
MCW	0	0	0	0	0

As shown in Table 5, five rounds of quantile regression analysis identified WFI and MDM in manufacturing, SAO, AFML, and MTT in transportation, as well as SDI and SFR in the on-site installation stages, as statistically significant at the 95% confidence level. These findings indicate that these factors play a critical role in influencing MOC life cycle performance. The statistical significance at the 95% confidence level suggests a strong correlation between these variables and overall project efficiency. Further analysis is required to explore the extent of their impact and potential strategies for optimizing these key indicators across different project phases. These MOC-KPIs are deemed robust, as they contribute meaningfully to identifying the root causes of waste across the MOC life cycle by accounting for resource utilization and enabling an effective and efficient evaluation of the MOC-LCP.

5 Conclusion

Certain studies have proposed mathematical and simulation models to evaluate productivity and process flow in MOC, focusing on minimizing variability. Other research works have developed KPIs in manufacturing to enhance MOC operational efficiency. However, few studies have addressed the root causes of process waste related to resource utilization (e.g., workers and workspace), nor have they proposed MOC-KPIs and considered the dynamic interactions among life cycle stages. This study fills this gap by proposing a process-oriented PE model that incorporates the proposed MOC-KPIs. Using SDM, the model simulates the interactions among life cycle stages and their processes, systematically assessing the current-state life cycle performance in terms of productivity (i.e., MOC-LCP) by comparing it to a baseline while considering process constraints and resource utilization (workers and workspace). The study also introduces new MOC-KPIs to identify, quantify, and diagnose the root causes of waste and enhance coordination across life cycle stages in MOC. The results show that MOC-KPIs such as WFI and MDM in manufacturing, SAO, AFML, and MTT in transportation, and SDI and SFR in on-site installation phases can effectively assess the state of resource utilization and identify process inefficiencies. Future research should explore the impact of applying the model to projects with a larger number of modules, as its calibration was based on a 6-module project, which may affect result reliability. Additionally, further studies should investigate the relationship between MOC-LCP improvement and project cost savings. Also, social indicator (e.g., employee turnover) are excluded due to their subjective and qualitative nature, as well as the lack of standardized assessment methods. This research improves the understanding of the key factors affecting MOC life cycle performance, helping general contractors and modular developers manage constraints and optimize project outcomes.

References

[1] McKinsey Global Institute. *Delivering on construction productivity is no longer optional*, New York, USA, 2024.

[2] R. Hogue. *Housing trends and affordability*. Royal Bank of Canada. Canada. 2023.

[3] M. Kamali, and K. Hewage. Development of performance criteria for sustainability evaluation of modular versus conventional construction methods. *Journal of Cleaner Production,* 142(4): 3592–3606, 2017.

[4] Modular Building Institute (MBI). *2023 Permanent modular construction report*. Charlottesville, USA, 2023.

[5] National House Building Council (NHBC). *Modern Methods of Construction: who's doing what?*. UK. 2018.

[6] H. Xue, S. Zhang, Y. Su, Z. Wu, and R.J. Yang. Effect of stakeholder collaborative management on off-site construction cost performance. *Journal of Cleaner Production*, 184: 490-502, 2018.

[7] L.d.C. Ng Corrales, M.P. Lambán, P. Morella, J. Royo, J.C. Sánchez Catalán, and M.E. Hernandez Korner. Developing and implementing a lean performance indicator: Overall process effectiveness to measure the effectiveness in an operation process. *Machines*, 10(133), 2022.

[8] H. Yu, T. Tweed, M. Al-Hussein, and R. Nasseri. Development of lean model for house construction using value stream mapping. *Journal of Construction Engineering and Management,*

135(8): 782–790, 2009.

[9] Y. Zhang, Z. Lei, S.H. Han, A. Bouferguene, and M. Al-Hussein. Process-oriented framework to improve modular and offsite construction manufacturing performance. *Journal of Construction Engineering and Management*, 146(9), 2020.

[10] T. Kim, Y.W. Kim, and H. Cho. Customer earned value: Performance indicator from flow and value generation view. *Journal of Management in Engineering*, 32(1), 2016.

[11] R.E. Orgut, J. Zhu, M. Batouli, A. Mostafavi, and E.J. Jaselskis. Critical factors for improving reliability of project control metrics throughout project life cycle. *Journal of Management in Engineering,* 36:(1), 2020.

[12] A.S. Hanna, W. Lotfallah, D.G. Aoun, and M. El-Asmar. Mathematical formulation of the project quarterback rating: new framework to assess construction project performance. *Journal of Construction Engineering and Management*, 140(8), 2014.

[13] J. Rezaei, A. Hemmes, and L. Tavasszy. Multi-criteria decision-making for complex bundling configurations in surface transportation of air freight. *Journal of Air Transportation Management,* 61:95-105, 2017.

[14] R.E. Orgut, J. Zhu, M. Batouli, A. Mostafavi, and E.J. Jaselskis. Metrics that matter: Core predictive and diagnostic metrics for improved project controls and analytics. *Journal of Construction Engineering and Management*, 144(11), 2018.

[15] A. Ibrahim, Jing, W. and Wenge, D. Key performance indicators supporting decision-making affecting Malaysian enterprise' project performance in China. *American Journal of Applied Sciences*, 7(2):241-247, 2010.

[16] S. Dabirian, M. Moussazadeh, M. Khanzadi, and S. Abbaspour. Predicting the effects of congestion on labour productivity in construction projects using agent-based modelling. *International Journal of Construction Management*, 1-24, 2021.

[17] J.A. Garza-Reyes, I. Oraifige, H. Soriano-Meier, P.L. Forrester, and D. Harmanto. The development of a lean park homes production process using process flow and simulation methods. *Journal of Manufacturing Technology Management*, 23(2):178–197, 2012.

[18] F. Mellado, E.C.W. Lou, and C.L.C. Becerra. Synthesising performance in the construction industry: An analysis of performance indicators to promote project improvement. *Engineering, Construction and Architectural Management,* 27(2): 579-608, 2019.

[19] M., Habibi, S. Kermanshachi, and B. Rouhanizadeh. Identifying and measuring Engineering, Procurement, and Construction (EPC) key performance indicators and management strategies. *Infrastructures*, 4(14):1-19, 2019.

[20] J.D. Sterman, Business dynamics: Systems thinking and modelling for a complex world. Irwin Professional, McGraw-Hill, New York, 2000.

[21] H.S. Cha, and C.K. Kim. Quantitative approach for project performance measurement on building construction in South Korea. K*SCE Journal of Civil Engineering*,15(8):1319-1328, 2011.

[22] R.D. Rodriguez, K. Medini, and T. Wuest. A DMAIC framework to improve quality and sustainability in additive manufacturing - A Case Study. *Sustainability*, 14(581), 2022.

[23] R. Arashpour, Wakefield, N. Blismas, and E.W.M. Lee. Analysis of disruptions caused by construction field rework on productivity in residential projects. *Journal of Construction Engineering and Management*, 140(2), 2014.

[24] H. Leon, H. Osman, M. Georgy, and M. Elsaid. System dynamics approach for forecasting performance of construction projects. *Journal of Management in Engineering,* 34(1), 2018.

[25] M. Arashpour, R. Wakefield, N. Blismas, and E.W.M. Lee. Framework for improving workflow stability: deployment of optimized capacity buffers in a synchronized construction production. *Canadian Journal of Civil Engineering*, 41(12): 995–1004, 2014.

[26] M. Arashpour, R. Wakefield, B. Abbasi, M. Arashpour, and R. Hosseini. Optimal process integration architectures in off-site construction: Theorizing the use of multi-skilled resources. *Architectural Engineering and Design Management*, 1-14, 2018.

[27] M. Arashpour, R. Wakefield, N. Blismas, and J. Minas. Optimization of process integration and multi-skilled resource utilization in off-site construction. *Automation in Construction*, 50: 72–80, 2015.

[28] M. Arashpour, R. Wakefield, N. Blismas, and T. Maqsood. Autonomous production tracking for augmenting output in off-site construction. *Automation in Construction,* 53:13–21, 2015.

[29] B. Manouchehri. Dynamic Modelling of Productivity in Modular Construction. In *Proceedings of the 2019 Canadian Society for Civil Engineering Annual Conference (CSCE 2019)*, Canadian Society for Civil Engineering. 2019.

[30] S. Wang, J. Tang, Y. Zou, and Q. Zhou. Research on production process optimization of precast concrete component factory based on value stream mapping. *Engineering, Construction and Architectural Management*, 27(4): 850-871, 2020.

Enhancing Construction Inspection with Augmented Reality: An Analysis of Accuracy and Observation of its Application in Training

Jeffrey Kim[1] and Darren Olsen[1]

[1]McWhorter School of Building Science, Auburn University, US
Jeff.Kim@auburn.edu, DAO0002@auburn.edu

Abstract –

The construction industry faces significant challenges, including declining quality performance, skilled labor shortages, and resistance to technological innovation. These issues have hampered the sector's operational efficiency and reputation. A critical component of quality control—accurate material placement—remains dependent on manual interpretation of 2D drawings, a task often prone to errors due to workforce inexperience. To address these concerns, this study investigates augmented reality (AR) as a tool to modernize construction inspection processes and enhance workforce training under the framework of Construction 4.0. This research evaluates the accuracy and practical utility of AR devices, including the Apple iPhone 13 Pro Max, Apple iPad Pro 12.9-inch, Trimble GPS Enabled SiteVision, and Trimble XR10 - Microsoft HoloLens 2, for inspecting the placement of underground utilities in a simulated construction environment. Thirty-two construction management students participated, using AR tools to position survey prisms at perceived utility locations. Results revealed the HoloLens achieved the highest accuracy and precision, followed by the SiteVision and iPad, with the iPhone exhibiting the largest deviations. Findings highlight AR's potential as an educational tool, enabling students to bridge theoretical knowledge with practical applications while engaging in immersive, hands-on learning. Despite limitations, including environmental variability and current technological tolerances, AR demonstrates significant promise in training future professionals and improving quality control processes. This study underscores the need for continued innovation in AR technology to advance the construction industry's efficiency, reputation, and workforce readiness.

Keywords –

Augmented Reality, Quality Control, Experiential Learning, Construction 4.0

1 Background and Rationale

The construction industry, a cornerstone of infrastructure development, is at a critical transition concerning standards in quality control. Extensive research underscores a significant decline in quality performance over recent decades. Key factors contributing to this deterioration include a persistent shortage of skilled labor [1], insufficient investment in quality management systems [2][3], and inadequate leadership in adopting technological innovations [4]. These systemic challenges have impaired the industry's competitiveness and operational efficiency. Moreover, entrenched resistance to process change and a tarnished public reputation exacerbate the sector's difficulties in adoption [5]. Concurrently, the industry is grappling with the loss of experienced professionals while relying on a less-skilled incoming workforce that requires substantial training time to achieve proficiency [6]. These intertwined issues further hinder efforts to improve productivity and public perception.

Precision in material placement is a critical aspect of construction processes. Currently, this task heavily depends on manual interpretation of two-dimensional (2D) drawings to construct three-dimensional (3D) structures. Inexperienced installers often face challenges in accurately interpreting these representations, leading to errors and delays in construction. Effective layout and quality control require practiced visualization skills typically developed through extensive field experience. However, the departure of skilled professionals has widened the expertise gap, compounding the industry’s quality control challenges and adversely affecting its reputation for reliability and customer satisfaction.

To address these issues, innovative technologies are being explored under the banner of Construction 4.0, the industry’s response to Industry 4.0 advancements. Among these, augmented reality (AR) has emerged as a transformative tool, with applications ranging from enhancing safety training [7] and improving design

visualization [8] to owner training and building maintenance support [9]. AR offers a potential solution for bridging the skills gap by facilitating training and comprehension. Its potential for quality control is particularly appealing. Application of AR in quality control has been demonstrated with the verification of proper rebar installation, albeit with the assistance of laser scanners [10]. More recently, large timber installation inspections using AR has been shown to improve fastener placement for improved accuracy and safety of those installed structures [11]. Despite these past studies and others like it, none are focused on the tool for its comparable training purpose, identifying a gap in this narrowly focused viewpoint. Nevertheless, leveraging AR effectively requires a nuanced understanding of construction-specific quality control procedures.

One illustrative example is the measurement of placement tolerances for utilities embedded in concrete structures. Ensuring accurate placement is essential to prevent costly rework. Early identification of deficiencies by inspectors is crucial to maintaining quality standards. For AR to succeed in augmenting this inspection process, it must deliver consistent accuracy and precision in identifying defects. Additionally, AR's capacity to overlay virtual objects onto real-world environments must be rigorously assessed for practicality and reliability.

Research highlights AR's efficacy in enhancing both classroom and on-site training [12][13]. With the proliferation of mobile devices, AR tools can now be deployed with minimal effort, enabling construction professionals and students to integrate these technologies seamlessly into their workflows. However, while the potential of AR for improving quality control processes is widely acknowledged, the current body of research remains speculative, emphasizing the need for empirical studies to validate its applications [8][14][15].

This study addresses these gaps through a dual focus. First, it examines students' perceptions of AR tools as a means of teaching essential inspection and analysis skills. Second, it evaluates the accuracy of AR inspection tools in the classroom setting, hypothesizing that their effectiveness in educational contexts could translate to practical applications on construction sites. The research assumes that AR can aid students in developing foundational skills essential for future industry practitioners. Furthermore, data on the accuracy and precision of AR tools could inform their integration into real-world quality control workflows, offering practical solutions for outdated inspection practices.

By establishing a foundation for incorporating AR tools into construction education and professional practices, this research contributes to modernizing the industry. It aims to address the knowledge gap left by departing experts while enhancing productivity and reputation. These efforts signify a critical step toward an industry that struggles to embrace innovation, retain expertise, and deliver higher-quality products.

1.1 Research Rationale

How can the construction industry modernize its outdated quality control processes, improve its reputation for quality, and address the challenges posed by a declining skilled workforce? The industry's traditional mindset—relying on methods that "worked in the past"—is no longer sufficient. A paradigm shift [16] is necessary, beginning with equipping the incoming workforce with advanced training and tools to meet modern demands effectively.

This research focuses on evaluating the accuracy and potential of AR technology as a tool for construction inspections. By integrating AR into training programs, the study aims to address two critical issues: enhancing quality control and bridging the skills gap in the workforce. The research involves students from a construction management program, providing them with hands-on experience and exposure to innovative tools. Through this approach, the study seeks to demonstrate AR's capacity to improve construction processes while preparing the next generation of professionals to meet industry challenges head-on.

2 Methodology

This study employs an action research protocol, integrating insights from various disciplines to assess AR tools' effectiveness in evaluating the placement accuracy of underground utilities. By superimposing utilities over a real-world view, the study aims to prepare construction students as future professionals while connecting classroom concepts to practical applications [17].

Previous research on AR accuracy and precision informs this study. Scherl et al. demonstrated error tolerances within 1.33 mm using fiducials in surgical procedures [18], while Khan et al. found object placement accuracies within 1.1 mm in virtual environments under stationary conditions and controlled lighting [19]. These findings influenced the authors' device selection and methodology, leveraging best practices to evaluate AR-assisted placement accuracy in construction.

2.1 General Participant Demographics

The study involved students from a four-year construction management program in the Southeastern United States. Participation was voluntary, and the experiment took place during a regularly scheduled class session. These students were chosen because of their

foundational knowledge aligning with the study's goals and their aspirations to become project managers or superintendents in commercial construction. Most participants, aged 20 to 23, had gained hands-on experience through internships, giving them a practical understanding of construction practices and the critical role of accuracy in utility placement. This background enabled them to effectively distinguish between correctly and incorrectly placed objects, contributing to the study's credibility. Thirty-two students participated in this study.

2.2 Devices Used in this Study

Devices were selected to align the study's emphasis on accuracy and precision, as informed by the findings of Khan et al. [19] and Scherl et al. [18]. The devices chosen for the study (shown in Figures 1 and 2 below) included the Apple iPhone 13 Pro Max (iPhone), the Apple iPad Pro 12.9-inch 5th generation (iPad), and the Trimble GPS Enabled SiteVision, representing a *passive*, non-fiducial AR device category. Additionally, the Trimble XR10 - Microsoft HoloLens 2 (HoloLens) was selected as an *active*, non-fiducial AR device type. The terms "passive" and "active" refer to the level of user interactivity: the HoloLens offers a more immersive experience, placing the user into a virtual environment compared to the other devices. Additionally, the iPhone and iPad were chosen for their familiarity among students, simplifying data collection and enhancing its reliability. The HoloLens and the Trimble SiteVision were included due to their potential as cutting-edge tools for professionals, providing students with an opportunity to engage with technology that could modernize traditional inspection workflows.



Figure 1. (a.) iPhone & (b.) iPad



Figure 2. (a.) HoloLens & (b.) SiteVision

2.3 The Experimental Space and AR Authoring Tools

The experimental case study was conducted within a simulated construction environment designed to mimic pre-concrete placement inspections commonly performed on construction sites. This phase is critical as it occurs before concrete is poured, allowing for the identification and resolution of deficiencies to avoid costly rework or delays.

The simulated site (Figure 3.a.) spanned 20 feet by 15 feet (6.1 m x 4.6 m) and was enclosed by wooden forms, creating a flat, stable surface for students to conduct observations using the AR devices described earlier (see Figures 1 and 2). Students used the devices to position a survey prism mounted on a PVC pipe supported by a tripod (Figure 3.b.), simulating the top of an underground utility pipe. The accuracy and precision of these placements were later evaluated against control points established within the simulation.



Figure 3. (a.) Experimental Workspace & (b.) Locating

Prisim

The site was outdoors and partially shaded, reflecting typical lighting conditions encountered on construction sites. Conducting the study in natural sunlight introduced practical challenges associated with real-world inspections using AR, enhancing the study's relevance to industry applications. While outdoor lighting and weather conditions were expected to influence data collection, they were not specifically measured during this study.

The AR experience was facilitated using Trimble FieldLink MR, a platform built on Building Information Modeling (BIM) principles widely adopted in the industry [18]. A BIM model of the simulated site was initially created in Revit, then converted to a SketchUp file for seamless integration with the Trimble FieldLink MR platform (see Figure 4). This cloud-based solution ensured consistent access to the AR experience across all devices used in this study, promoting uniformity in participant interactions and data collection.

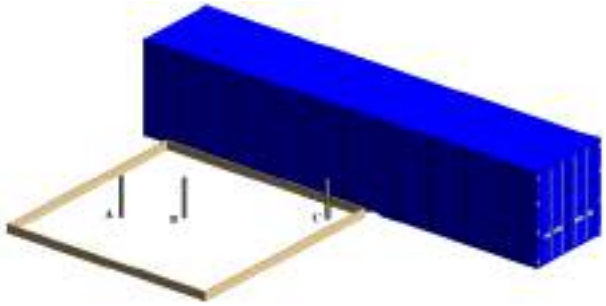


Figure 4. Rendering of the AR Display

Within the model, three virtual PVC pipes (A, B, and C) were positioned to represent utility penetrations typically encountered on construction sites. These digital proxies allowed participants to observe and assess placement accuracy before concrete placement, aligning the experiment with real-world inspection practices and ensuring practical relevance to the construction industry.

2.4 The Experimental Workflow

At the start of the study, control points were established within the simulated construction site and validated using a robotic total station. During the experiment, students navigated the simulated site (Figure 3.a.) with their assigned AR device (Figure 1 and 2), which displayed virtual pipes and their spatial positions (Figure 4). Upon identifying a virtual pipe, students positioned a survey prism stand (Figure 3.b.) at its perceived center. The robotic total station then recorded the student's identified pipe location, capturing both northing and easting coordinates for subsequent analysis. To evaluate the accuracy of the observations, the recorded pipe coordinates from the students were compared against the control coordinates of the pipes.

3 Results and Discussion

Each pipe was individually located by students utilizing their designated AR devices, allowing every participant to make up to three positional observations specific to their assigned device.

The positional data (northing and easting coordinates) for all students were compared against the control positions of the simulated pipes using the formula outlined in Figure 5.

$$\sqrt{(N_n - N_y)^2 + (E_n - E_x)^2}$$

Figure 5. Formula for Calculating Placement Difference (Northing and Easting coordinates).

Figure 5. (N_n and E_n = pipe control coordinates and N_y and E_x = student observed coordinates)

3.1 Accuracy and Precision

A summary of the calculated error averages and standard deviations is presented in Table 1.

Table 1. Error Calculations

PIPE		DEVICE TYPES			
		HoloLens	SiteVision	iPhone	iPad
A	Average	2.47	3.05	7.51	3.30
	SD	1.23	1.85	4.22	1.95
	n	8	8	8	8
B	Average	2.73	3.60	5.97	3.66
	SD	1.69	2.40	2.80	1.17
	n	8	8	8	8
C	Average	4.65	3.91	5.30	3.79
	SD	1.34	2.53	1.43	1.39
	n	8	8	8	8

Note: All error measurements are represented in (SI) inches.

Analyzing the results for each AR device—the HoloLens, SiteVision, iPhone, and iPad—reveals notable variations in performance. The HoloLens consistently demonstrated the highest placement accuracy, with an overall average error of 3.28 inches across all three pipes (A, B, and C). In comparison, the SiteVision recorded an average error of 3.52 and the iPad recorded a near similar average error of 3.59 inches. The iPhone exhibited the largest error, averaging 6.26 inches. Precision, assessed through standard deviation, further underscored the HoloLens' reliability, as it had the lowest standard deviation of 1.42, followed closely by the iPad at 1.50. The SiteVision and iPhone trailed with a standard

deviation of 2.26 and 2.82 respectively.

These findings align with the expectations set by Trimble's whitepaper [21], affirming the superior performance of the HoloLens' WLT technology compared to devices lacking such features. Similar to Trimble's conclusions, the HoloLens excelled in anchoring virtual elements to the real-world environment. It should be noted, however, that this study differed methodologically by excluding fiducials and stationary observers, leading to accuracy results that diverged from prior research [19][18]. These deviations were anticipated, given the experimental conditions.

The iPad and the SiteVision emerged as a practical alternatives to the HoloLens, with accuracy and precision closely mirroring those of the HoloLens. Despite this, the iPad offers a less immersive experience, as its AR interface relies on a handheld screen. In contrast, the HoloLens provides a more engaging, hands-free experience by overlaying virtual elements directly onto the user's field of view via a headset. This hands-free operation enhances mobility and safety, whereas the need to hold the iPhone or iPad can restrict movement and increase safety risks [22]. Consequently, the decision to use these devices should balance factors such as cost, immersion, and safety. The SiteVision accommodates for this ergonomic consideration by providing a handle for operation, but it too is a device that requires the use of at least one of the operator's hands.

While the observed tolerances in this study indicate that the current AR tools may not meet the precise standards required in many construction industry applications—where tolerances often range between 1/8" and 1/4"—they nonetheless highlight the potential of AR as a valuable educational tool. The technology's ability to bridge theoretical learning with practical applications makes it a promising resource for advancing industry training and education.

3.2 Practicality for Academic and Industry Use

The AR tool proves to be a highly valuable educational resource, especially as a training platform for aspiring inspectors. While its accuracy may not yet align with industry standards, its immersive nature offers a unique opportunity for hands-on learning. Through interactive simulations and practical exercises, trainees can refine their ability to detect and address construction deficiencies, gaining a deeper understanding of inspection processes. Participant feedback highlights the AR tool's potential as a transformative educational aid. As one student expressed in their open-ended comments:

"The experience was enjoyable and gave me a better grasp of how to inspect for construction errors."

Anecdotally, this feedback reflects a broader trend of positive student engagement and impactful learning experiences when action research is implemented [23]. Research supports the effectiveness of active, experiential learning strategies in improving educational outcomes, particularly when cognitive load is managed effectively [12][24].

These findings underscore the AR tool's promise as a teaching aid within the construction field. By leveraging its immersive capabilities to foster active participation, the tool not only enhances practical competencies but also encourages a culture of ongoing learning and innovation. The study's results suggest that the AR tool positively transforms the educational experience, signalling a shift toward more dynamic and interactive training methods in the industry.

3.3 Considerations for the Future of this Research

While this study sought to manage key variables influencing outcomes, certain factors arose that merit consideration in future investigations.

Firstly, environmental conditions, such as sunlight and weather, were not controlled. This approach aimed to mimic real-world variability on construction sites, ensuring practical relevance for industry applications. However, during the experiment, delays occurred as researchers waited for optimal sunlight to continue, given that poor lighting could compromise the HoloLens experience and overheating issues were noted with the iPhone and iPad. These challenges led to the loss of several data points, requiring rework to maintain study integrity.

Secondly, the limited size of the student cohort may have constrained the breadth of the findings. A larger participant pool could provide a more comprehensive dataset, yielding deeper insights into the effectiveness of the AR tool.

Lastly, the requirement for students to work in pairs to record virtual pipe locations introduced an additional variable. Those using the iPhone and iPad faced limitations due to the need to hold their devices, leaving only one free hand for marking tasks. Conversely, HoloLens users had the advantage of hands-free operation, potentially influencing the overall results. Those using the SiteVision were afforded more flexibility since the device was equipped with a handle and wrist strap to precent dropping the device. While theses discrepancy likely impacted the study, no specific data was collected to confirm its effect.

These considerations point to areas where future research can refine methodologies, address limitations, and further explore the AR tool's potential to enhance learning and professional development in the construction industry.

4 Conclusion

The findings from this research reveal the transformative potential of AR as both an educational tool and a technology capable of modernizing quality control practices in the construction industry. With construction increasingly defined by its challenges—such as a declining skilled workforce, a need for higher-quality outcomes, and resistance to technological change—AR offers a pathway to innovation and improvement.

Through this study, AR tools demonstrated their ability to bridge the gap between theoretical learning and practical application, particularly in training students on critical inspection and analysis skills. The immersive and interactive nature of AR facilitates experiential learning, enabling students to practice identifying construction errors within a controlled, simulated environment. Feedback from participants underscores this value, with many noting increased engagement and comprehension. Active, hands-on learning methodologies, supported by the AR platform, align with established educational research, which emphasizes the importance of reducing cognitive load to optimize learning outcomes.

From a technical perspective, the study highlighted the relative strengths and limitations of AR devices, such as the HoloLens, SiteVision, iPhone, and iPad, in achieving placement accuracy for simulated utilities. While the HoloLens emerged as the most precise tool, with accuracy levels surpassing other devices, the findings suggest that current AR technology may not yet meet the stringent tolerances required in some industry-specific applications. However, these tools show significant promise as cost-effective and accessible alternatives for training purposes. The hands-free operation of the HoloLens was shown to enhance user mobility and safety, offering an additional layer of practicality for both educational and field applications.

This research contributes to the growing body of evidence supporting AR's integration into construction practices. It highlights AR's dual role in addressing pressing industry challenges: by equipping the next generation of professionals with advanced skills and by improving workflows in quality control. These outcomes position AR as an instrumental technology for advancing the goals of Construction 4.0, fostering innovation, and enhancing the reputation and productivity of the construction sector [25].

In conclusion, this study demonstrates that augmented reality represents a significant opportunity for the construction industry to modernize its approaches to education and quality control. By fostering a culture of innovation and continuous learning, AR has the potential to mitigate the challenges associated with a declining skilled workforce while enhancing the quality and efficiency of construction processes. As the industry continues to evolve, embracing technologies like AR will be critical to building a future characterized by greater resilience, adaptability, and excellence in construction practices.

References

[1] Allen, S., 1985. Why Construction Industry Productivity is Declining. Macroeconomics eJournal. https://doi.org/10.2307/1924811.

[2] Rosenfeld, Y., 2009. Cost of quality versus cost of non‐quality in construction: the crucial balance. Construction Management and Economics, 27, pp. 107 - 117. https://doi.org/10.1080/01446190802651744.

[3] Arditi, D., & Gunaydin, H., 1997. Total quality management in the construction process. International Journal of Project Management, 15, pp. 235-243. https://doi.org/10.1016/S0263-7863(96)00076-2.

[4] Wright, R., 1989. Technology for competitiveness of the U.S. construction industry. . https://doi.org/10.6028/nist.ir.89-4099.

[5] Moraru, G., & Popa, D., 2021. Potential Resistance of Employees to Change in the Transition to Industry 5.0. MATEC Web of Conferences. https://doi.org/10.1051/matecconf/202134307005.

[6] Welfare, K., Sherratt, F., & Hallowell, M., 2021. Perceptions of Construction Work: Views to Consider to Improve Employee Recruitment and Retention. Journal of Construction Engineering and Management. https://doi.org/10.1061/(ASCE)CO.1943-7862.0002057.

[7] Li, X., Yi, W., Chi, H., Wang, X., & Chan, A., 2018. A critical review of virtual and augmented reality (VR/AR) applications in construction safety. Automation in Construction, 86, pp. 150-162. https://doi.org/10.1016/J.AUTCON.2017.11.003.

[8] Hajirasouli, A., Banihashemi, S., Drogemuller, R., Fazeli, A., & Mohandes, S., 2022. Augmented reality in design and construction: thematic analysis and conceptual frameworks. Construction Innovation. https://doi.org/10.1108/ci-01-2022-0007.

[9] Chi, H. L., Kim, M. K., Liu, K. Z., Thedja, J. P. P., Seo, J., & Lee, D. E. (2022). Rebar inspection integrating augmented reality and laser scanning. Automation in Construction, 136, 104183. https://doi.org/10.1016/j.autcon.2022.104183

[10] Fazel, A., & Adel, A. (2024). Enhancing construction accuracy, productivity, and safety with augmented reality for timber fastening. Automation in Construction, 166, 105596. https://doi.org/10.1016/j.autcon.2024.105596

[11] Harikrishnan, A., Abdallah, A., Ayer, S., Asmar, M., & Tang, P., 2021. Feasibility of augmented reality technology for communication in the construction industry. Adv. Eng. Informatics, 50, pp. 101363. https://doi.org/10.1016/J.AEI.2021.101363.

[12] Kim, J. and Irizarry, J., 2021. Evaluating the use of augmented reality technology to improve construction management student's spatial skills. International Journal of Construction Education and Research, 17(2), pp.99-116.

[13] Olbina, S. and Glick, S., 2023. Using integrated hands-on and virtual reality (VR) or augmented reality (AR) approaches in construction management education. International Journal of Construction Education and Research, 19(3), pp.341-360.

[14] Adebowale, O., & Agumba, J., 2022. Applications of augmented reality for construction productivity improvement: a systematic review. Smart and Sustainable Built Environment. https://doi.org/10.1108/sasbe-06-2022-0128.

[15] Delgado, J., Oyedele, L., Beach, T., & Demian, P., 2020. Augmented and Virtual Reality in Construction: Drivers and Limitations for Industry Adoption. Journal of Construction Engineering and Management-asce, 146, pp. 04020079. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001844.

[16] Kuhn, Thomas S. The structure of scientific revolutions. Vol. 962. Chicago: University of Chicago press, 1997.

[17] Zuber‐Skenitt, O., 1993. Improving Learning and Teaching Through Action Learning and Action Research. Higher Education Research & Development, 12, pp. 45-58. https://doi.org/10.1080/0729436930120105.

[18] Scherl, C., Stratemeier, J., Karle, C., Rotter, N., Hesser, J., Huber, L., Dias, A., Hoffmann, O., Riffel, P., Schönberg, S., Schell, A., Lammert, A., Affolter, A., & Männle, D., 2020. Augmented reality with HoloLens in parotid surgery: how to assess and to improve accuracy. European Archives of Oto-Rhino-Laryngology, 278, pp. 2473 - 2483. https://doi.org/10.1007/s00405-020-06351-7.

[19] Khan, F., Muvva, V., Wu, D., Arefin, M., Phillips, N., & Swan, J., 2021. Measuring the Perceived Three-Dimensional Location of Virtual Objects in Optical See-Through Augmented Reality. 2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 109-117. https://doi.org/10.1109/ismar52148.2021.00025.

[20] Sacks, R., Eastman, C., Lee, G. and Teicholz, P., 2018. BIM handbook: A guide to building information modeling for owners, designers, engineers, contractors, and facility managers. John Wiley & Sons.

[21] Trimble White Paper on Accuracy of the Trimble XR-10, October 2020.

[22] Alzhrani, A., Cook, M., Johnstone, K., & Cooper, J., 2018. Handheld Mobile Devices—How Do We Use Them at Work? A University Case Study. Advances in Intelligent Systems and Computing. https://doi.org/10.1007/978-3-319-96083-8_16.

[23] Zuber‐Skenitt, O. (1993). Improving learning and teaching through action learning and action research. Higher education research and development, 12(1), 45-58. https://doi.org/10.1080/0729436930120105

[24] Sweller, J., 2020. Cognitive load theory and educational technology. Educational Technology Research and Development, 68(1), pp.1-16.

[25] Sawhney, A., Riley, M. and Irizarry, J., 2020. Construction 4.0: Introduction and overview. In Construction 4.0 (pp. 3-22). Routledge.

Fuzzy-Chaos Framework for Analyzing and Quantifying Uncertainty in Labour Productivity in Complex Construction Environments

Elyar Pourrahimian[1], Diana Salhab[1], Aminah Robinson Fayek[1] and Farook Hamzeh[1]

[1]Department of Civil and Environmental Engineering, University of Alberta, Canada
Elyar@ualberta.ca, Salhab@ualberta.ca, Aminah@ualberta.ca, Hamzeh@ualberta.ca

Abstract –

Effective management of labour productivity in construction projects is important for ensuring project success. This paper introduces a novel approach to quantifying uncertainty in labour productivity through entropy calculations based on outputs from a fuzzy expert system. Two distinct scenarios are analyzed to demonstrate the practical application of the proposed methodology, one representing a week with high variability and the other a stable and predictable week. These scenarios highlight how entropy can be a powerful tool for identifying periods of high uncertainty and assisting project managers in making informed decisions about resource allocation, risk management, and strategic planning. The findings suggest that high entropy values indicate weeks requiring increased management attention and resources, whereas lower entropy values correspond to more stable conditions, allowing for standard operational procedures. This approach enhances understanding of dynamic project conditions and supports proactive project management practices. This study contributes to the body of knowledge by integrating fuzzy logic with entropy calculations, which offers a robust framework for managing the complexities of labour productivity in construction projects. Through real-world application and comparative analysis, this study validates the effectiveness of entropy analysis as a diagnostic and decision-making tool in the construction industry.

Keywords –

Construction management; Fuzzy expert systems; Chaos theory; Entropy

1 Introduction

Traditional linear and deterministic approaches for analyzing labour productivity in construction environments often fail to fully address the dynamic nature of construction projects [1]. Management strategies must evolve as projects become complex and are influenced by numerous interacting variables and unpredictable environmental conditions.

In this context, fuzzy logic offers ways to develop a methodological framework capable of handling the ambiguity and imprecision inherent in construction project management. Fuzzy logic can be used to more effectively deal with uncertain parameters, such as estimations of time and resources, by classifying them into degrees of truth rather than binary true-or-false categories [2]. This allows for better decision-making that aligns with the complexity of construction projects and expert knowledge.

Chaos theory also provides a valuable lens through which the unpredictable nature of construction projects can be understood and managed. It suggests that small variations in project conditions can lead to disproportionately large effects, a phenomenon often referred to as the “butterfly effect” [3]. Chaos theory has helped highlight the non-linear behaviours of project dynamics, thus emphasizing the need for project managers to anticipate and prepare for significant fluctuations in project trajectories that traditional methods might not predict [3,4].

Combining chaos theory and fuzzy logic addresses the dual challenges of unpredictability and imprecision by providing tools that predict potential chaos in project management while offering pragmatic approaches to handling complexities through adaptable and flexible decision-making frameworks [5]. Such an integrated approach is particularly effective in managing the interdependencies and dynamic interactions within construction projects, which traditional project management techniques often struggle to handle effectively.

In the context of this study, labour productivity is defined specifically as the measure of output produced per labour hour. This definition focuses on the efficiency of labour in delivering tangible results within a given time frame. In construction research, productivity can

also encompass broader measures such as cost, duration, and overall project success.

This paper proposes a novel approach integrating chaos theory and fuzzy logic to enhance the understanding and management of labour productivity in construction environments. The objectives of this framework are to generate better understanding of the factors impacting labour productivity in complex construction environments and, ultimately, to facilitate more-informed management decisions. By understanding and applying the frameworks using these complex theories, project managers can enhance their ability to navigate the uncertainties of the construction field, leading to better project outcomes in terms of efficiency, cost, and time management. Thus, this approach advocates for a shift from conventional to more dynamic and responsive project management strategies that can better address the complexities of modern construction projects.

2 Background

Labour productivity in the construction industry is a critical measure of efficiency [6] and reflects the output per labour hour expended on construction tasks. It directly influences a project's cost, duration, and overall success. However, productivity in construction is uniquely challenged by the industry's inherent variability and complexity as well as the unpredictable nature of work environments and workforce dynamics [7]. It exhibits complex variability that is not normally distributed, which suggests the need to apply chaos theory to understand it [8].

Traditionally, construction productivity has been evaluated using time–motion studies, manual data collection, and direct observations of activities on-site [9]. Although useful, these methods often fail to capture the distinctions of complex environments where multiple interdependent variables affect productivity [10, 11]. Such methods are limited in handling construction projects' dynamic adaptations and stochastic nature, so more robust, adaptive analytical tools are required.

Fuzzy logic offers a means to handle uncertainty and imprecision, characteristics that are common in construction data [12]. It offers the ability to develop flexible mathematical frameworks for modeling the vagueness associated with subjective human judgments and qualitative data. For instance, studies have successfully implemented fuzzy systems to enhance decision-making under uncertainty in project management, particularly in areas such as risk assessment, cost forecasting, and the mitigation of potential delays and cost overruns [13, 14, 15, 16, 17].

Chaos theory is pertinent to understanding construction projects' complex and dynamic nature [18]. It demonstrates how even minor variations in initial conditions can cascade into disproportionately large impacts, a principle especially relevant to the non-linear dynamics of construction management [19, 20]. Through use of sensitivity analysis and fractal geometry, chaos theory provides methods for identifying and addressing sensitive points in the project lifecycle and repeating patterns that may significantly influence outcomes [19, 20]. Although less commonly applied in project management than fuzzy logic, chaos theory has the potential to provide insights into labour productivity by modeling how small variations in work conditions or team compositions can lead to disproportionate impacts on productivity. This approach can be particularly valuable in planning and managing workflows in real time to mitigate the effects of emergent issues [21].

Entropy is a measure of disorder or randomness in a system. Entropy analysis has been extensively explored as a tool for quantifying uncertainty in construction projects, thus offering insights into the stability and predictability of project systems by measuring deviations in cost, schedule, and performance [22, 23]. Employing entropy calculations alongside traditional project management metrics, such as cost performance index and schedule performance index, allows managers to monitor project progress and proactively address emergent issues [1].

Recent studies have demonstrated the utility of entropy-based approaches in risk assessment, resource allocation, and schedule stability, showing the ability of such approaches to predict project disruptions and enable dynamic adjustments. For example, entropy analysis has been used in scheduling resource-constrained projects to measure disorder and forecast potential delays [22], and it has been integrated with Bayesian inference to refine predictions based on new data [23].

Integrating fuzzy logic and chaos theory in construction management is a novel approach that has seen limited exploration, especially in terms of enhancing productivity. This integrative approach offers a framework capable of managing the inherent uncertainties of construction projects and predicting and adapting to the non-linear dynamics that affect labour productivity. The proposed framework thus enhances traditional project management approaches by integrating advanced analytics specifically designed to navigate the complexities of modern construction environments, thereby significantly improving labour productivity.

3 Methodology

The methodology, presented in Figure 1, explains a systematic process that begins with factor shortlisting, where key labour productivity factors are identified and

prioritized based on their relevance and impact on construction projects. Subsequently, these factors are evaluated through a fuzzy expert system (FES) that classifies each factor into qualitative states (*Poor*, *Fair*, *Good*) based on expert evaluation. This classification facilitates pattern analysis, enabling us to identify and interpret trends and patterns in the data, which provides insights into the factors' behaviors and interactions. The final step in this methodology is entropy analysis, where uncertainty and variability of the factors' states is quantified based on the FES results. This entropy analysis serves as a measure of disorder or randomness in the system, offering a deeper understanding of the project's complexity and better guiding strategic decision-making in managing construction project productivity.



Figure 1. Proposed methodology

3.1 Labour Productivity Factors

A comprehensive list of factors comprising eight main categories and 137 sub-factors that can positively or negatively influence labour productivity is adapted from the framework proposed by Raoufi and Fayek [24]. The methodology proposed by Pourrahimian et al. (2024) is employed to shortlist this extensive list [25] and focus on the most relevant factors for the case study. The detailed selection process and criteria, however, are beyond the scope of this study. This framework was selected due to its proven applicability in similar contexts, where handling fuzzy, uncertain, and non-linear data is crucial. Notably, the framework's robustness in dealing with complex variability in labour productivity makes it particularly suitable for construction projects affected by a wide range of fluctuating conditions. Figure 2 illustrates the distribution of sub-factors across each main factor category.

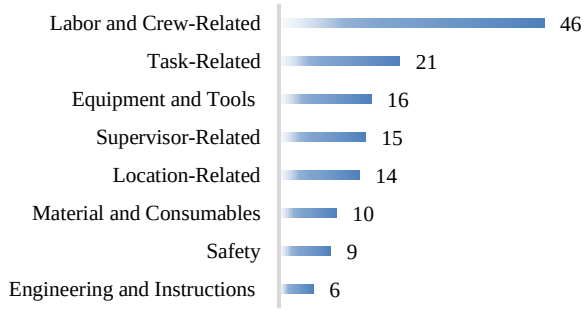


Figure 2. The number of sub-factors under each main factor

3.2 Fuzzy Expert System

FES is applied to manage the ambiguity and vagueness associated with these factors. Each main factor is represented as a fuzzy variable with degrees of membership ranging across relevant sets (e.g., *Safety* is categorized into *Poor*, *Fair*, and *Good*). Membership functions are defined for each fuzzy variable, quantifying the degree to which an element belongs to a particular fuzzy set. Fuzzy rules are constructed based on expert input and empirical data; these rules form the core of the fuzzy logic system, linking input variables to labour productivity.

To apply FES in labour productivity using a rule-based approach, it is essential to establish a set of rules within the model. Equation (1) is used to determine the number of fuzzy rules.

$$Number\ of\ rules = (m)^n \quad (1)$$

where m is the number of fuzzy sets and n is the number of input variables.

To reduce complexity and prevent an overwhelming number of rule definitions, rule blocks have been defined. Instead of directly using all sub-factors to evaluate the main factor, the model first assesses the impact of a group of sub-factors on their respective main factor and, subsequently, the influence of these groups on the main factor. For instance, with four input variables and three fuzzy sets, the required rules would be 81. However, by grouping these into two rule blocks, only 18 rules per block and an additional nine rules for combining the outputs of these blocks are needed. This method substantially reduces the number of rules, as Figure 3 illustrates.

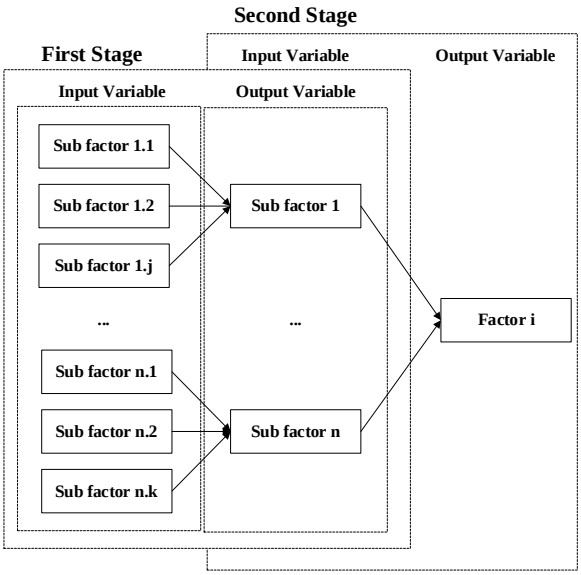


Figure 3. General structure of rule blocks

The input variables have the fuzzy set of x_i and the output variable of y. The input variables are measured on a scale of 0 to 10 and are divided into three categories, represented by the linguistic terms *Poor*, *Fair*, and *Good*. Membership values between 0 and 1 are assigned to each category. The output variable is on a scale of 0 to 100 and is also divided into three categories represented by the linguistic terms *Poor*, *Fair*, and *Good*. Table 1 tabulates the terms of assessment for x_i and y.

As the occurrence of each input variable is independent, FES rules were developed for each category. The outputs were used as input values for the next level. The standard operators *And: Min, Or: Max* implement the rules.

Table 1. Input and output variables of FES

Variable	Numerical scale	Linguistic descriptor	Membership function
x	1–10	Poor	Trapezoidal (0,0,1,4)
		Fair	Triangular (2.5, 4.5, 6.5)
		Good	Trapezoidal (5,8,9,9)
y	0–100	Poor	Trapezoidal (0, 0, 30, 50)
		Fair	Triangular (30, 50, 70)
		Good	Trapezoidal (60, 80, 100, 100)

3.3 Pattern Recognition

Chaos theory is utilized to identify patterns and conduct sensitivity analysis within the productivity system, highlighting its dynamic and highly responsive nature to minor variations in input conditions.

The model's pattern identification process examines productivity measure's non-linear responses to alterations in task characteristics or human factors, demonstrating how these changes can disrupt usual productivity patterns. Run charts can help identify these patterns. A run chart is an effective tool for monitoring data trends over time, displaying data points along a timeline on the *x* axis and their values on the *y* axis. It is used primarily to detect shifts identified by six or more points consistently above or below the median and trends indicated by five or more points consecutively rising or falling. Additionally, a run chart helps identify non-random variations by analyzing the number of runs (sequences of points on one side of the median) against expected ranges. It also shows outliers and unusual data points that merit further investigation.

Figure 4 shows an example run chart for the defuzzified evaluation of FES for a specific factor across various weeks. Tracking a variable over a series of weeks makes it easier to visualize trends, shifts, and patterns. Two distinct trends are indicated in differently coloured boxes. The section in a green box shows an increasing trend where the data points consistently rise over several weeks, suggesting a positive development in the monitored factor. The section in a red box illustrates a decreasing trend where the data points decline over time, signalling a reduction in the analyzed factor.

Another key feature shown in the chart is a "Shift," highlighted in a yellow box, where the data points remain on one side of the median over consecutive weeks. This could point to a significant event or intervention that impacted the process, and it merits closer examination to understand its causes and implications.

Additionally, the chart includes an "Astronomical" point, indicating an outlier. This particular data point is substantially different from others and could represent an anomaly that may skew the overall analysis or a unique event that requires separate consideration.

This run chart exemplifies how visual data analysis tools can help monitor ongoing processes. Identifying increasing and decreasing trends provides insights into the performance dynamics over time.

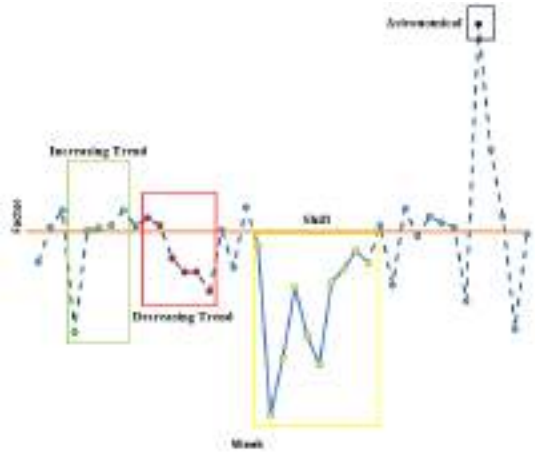


Figure 4. Pattern identification example

3.4 Entropy Analysis

Following identification of potential chaotic patterns, entropy calculations are integrated to measure the degree of uncertainty and disorder within the productivity factor states. This step enhances researchers' understanding of the variability in productivity outcomes and provides a quantitative measure of system complexity.

By applying FES, each productivity factor is evaluated and categorized as *Poor*, *Fair*, or *Good*; the next step entails quantifying the uncertainty embedded in these assessments using entropy calculations. The FES provides membership values for each state of a factor, which are inherently probabilities that sum to one.

Therefore, no additional normalization is required. These membership values directly reflect the system's confidence in each productivity state for each factor considered.

Entropy is calculated for each factor to measure the level of uncertainty or disorder associated with the fuzzy state distributions. Shannon's entropy formula is applied, given by:

$$H = -\sum_{i=1}^{3} p_i \log(p_i) \quad (2)$$

where p_i represents the probability of each state for a particular factor. Where a state's membership value is zero, its contribution to the entropy calculation is also zero, as $0\log(0)$ is mathematically treated as zero. This process is repeated for each factor, and the results provide a measure of the uncertainty associated with each aspect of labour productivity. To gain a holistic view of a project's uncertainty in terms of labour productivity, the entropy values are aggregated across all factors. This can be done by averaging all individual entropy values:

$$H_{average} = \frac{1}{8}\sum_{j=1}^{8} H_j \quad (3)$$

where $H_{average}$ is an overall measure of the uncertainty in the project's labour productivity and indicates the complexity and variability within the project environment.

An entropy value close to 1 indicates greater unpredictability and variability among the factors, which implies a more complex scenario where multiple interdependent factors influence productivity. Conversely, a lower entropy value, closer to 0, indicates more stability and predictability. Understanding these entropy values helps project managers and decision-makers identify periods or conditions of high uncertainty, which enables them to take pre-emptive measures or adjust project management strategies accordingly. This approach aids in better resource allocation, risk management, and scheduling, ultimately leading to more efficient project execution. For a more in-depth discussion, including specific entropy thresholds and their implications across various project phases, the authors refer readers to the detailed journal article by Pourrahimian et. al (2024), which is citation [1]

Figure 5 maps the assessment process of various productivity factors and their sub-factors within a construction project using the proposed framework. Each factor group is evaluated using fuzzy logic to determine each factor's state (*Poor*, *Fair*, *Good*). Subsequently, the distribution of these states across each factor is analyzed to calculate entropy, providing a measure of uncertainty or variability.

3.5 Objectives and Expected Outcomes

The primary objectives of this framework are twofold. First, by integrating fuzzy logic and chaos theory, the framework aims to better understand the factors impacting labour productivity in complex construction environments. Second, it is designed to support decision-making processes by understanding the variability in productivity outcomes and providing a quantitative measure of system complexity. This dual focus enhances

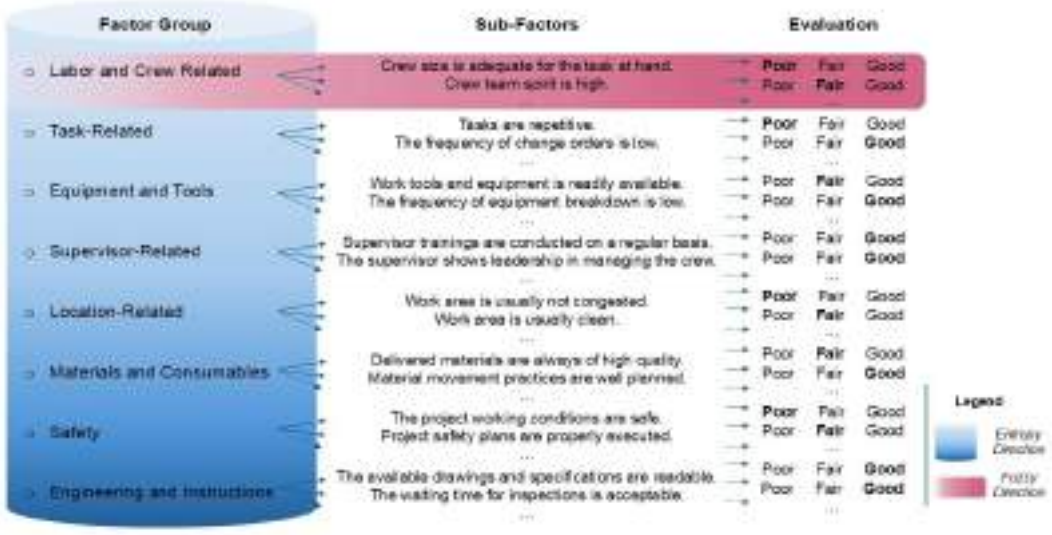


Figure 5. Integrated analysis of labour productivity factors using fuzzy logic and entropy

comprehension of dynamic productivity issues and facilitates more-informed management decisions.

4 Case Study, Results, and Discussion

The case study examines data from a $100-million industrial project, from February 2022 to October 2023, involving approximately 400,000 direct labour hours. The project scope included electrical and instrumentation (E/I) work, which made up 13% of the project, and piping tasks, which comprised 40% to 45%. The project was expected to employ a peak workforce of 280 individuals, comprising 250 workers and staff, plus 30 subcontractors, with 125 workers dedicated specifically to piping tasks. An expert panel featuring structural, electrical, piping, and mechanical engineering professionals was assembled to identify critical productivity factors. The panel included a structural construction manager with 22 years of experience, an electrical superintendent with 15 years, a piping superintendent with 34 years, a mechanical engineer operations manager with 19 years, and an electrical foreperson with 12 years. Following extensive analysis, this panel unanimously identified the essential state variables impacting project productivity.

Table 2 presents two distinct scenarios with the FES for each factor, each representing the labour productivity conditions across two different weeks of the project, to illustrate how varying project environments can significantly impact the entropy calculations and, consequently, the perceived stability and uncertainty within a construction project.

The first scenario explores a week characterized by high variability in project conditions. This includes disruptions in supply chains, unforeseen changes in project specifications, and fluctuations in workforce availability. The FES's evaluations reflect this instability, assigning a mix of Poor and Fair states across most factors, with very few factors rated as Good. This scenario illustrates a week for which entropy calculations reveal high uncertainty, indicating a need for proactive management interventions to stabilize the project environment. Conversely, the second scenario depicts a week where the project environment is stable and predictable. Factors such as weather conditions, material supply, and crew performance are optimal, leading to predominantly Good evaluations from the FES. This week demonstrates a lower entropy value, suggesting a well-managed project with minimal uncertainty and high predictability in labour productivity.

The integration of fuzzy logic and chaos theory within construction management can provide a promising framework for addressing the industry's inherent complexity. Beyond the immediate applications demonstrated in this study, insights from related fields further emphasize the potential of these methodologies to provide an alternative way to enhance the performance of construction projects.

Table 2. Membership value for each factor in two scenarios

	Scenario 1			Scenario 2		
Factor	Poor	Fair	Good	Poor	Fair	Good
Engineering and instructions	0.6	0.4	0	0	0	1
Safety	0.5	0.5	0	0	0.1	0.9
Material and consumables	0.7	0.3	0	0	0.1	0.9
Location-related	0.5	0.5	0	0	0.1	0.9
Supervisor-related	0	0.5	0.5	0	0.1	0.9
Equipment and tools	0	0.6	0.4	0	0.1	0.9
Task-related	0	0.4	0.6	0	0	1
Labour- and crew-related	0.5	0.5	0	0	0	1
Entropy ($\boldsymbol{H_{average}}$)	0.97			0.25		

The authors hypothesize that the fuzzy-chaos model will exhibit superior adaptability and accuracy in forecasting labor productivity due to its unique ability to manage construction projects' inherent uncertainty and non-linearity. This capability stems from the integration of fuzzy logic, which allows for handling ambiguous and imprecise data, and chaos theory, which addresses the sensitive dependence on initial conditions.

The fuzzy chaotic neural network proposed by Zhang and Wang (2012) presents a combination of fuzzy logic, chaos theory, and neural network dynamics to tackle highly non-linear and uncertain systems. This approach mirrors the core challenges faced in construction labour productivity management, where complex interdependencies and persistent fluctuations dominate [26]. Incorporating elements of such neural networks into construction project models could significantly enhance their ability to simulate issues, such as workforce dynamics or unforeseen disruptions. For instance, chaotic feedback loops could be employed to anticipate

cascading delays or cost overruns, providing a proactive management tool.

Similarly, the chaos-adapted multi-verse optimizer algorithm demonstrates the power of integrating chaotic maps and fuzzy logic in optimization tasks. Although applied in benchmark optimization problems, the principles underlying these chaotic adaptations could be valuable for construction resource management [27]. For example, replacing traditional deterministic approaches with chaos-driven algorithms could improve scheduling robustness and enable construction managers to dynamically allocate resources in response to fluctuating conditions.

The use of chaotic maps to replace stochastic components in optimization algorithms offers an appealing analogy for managing uncertainties in construction projects. Just as the algorithm balances exploration and exploitation of solution spaces, construction managers could apply similar principles to balance short-term project needs with long-term strategic goals, which might ultimately enhance efficiency and resilience.

These external applications provide evidence of the power of chaos theory and fuzzy logic across domains, which can be successfully adapted to the construction industry.

This study contributes critical advancement by contextualizing and empirically validating these theories within the construction industry. This not only enhances the understanding of complex project dynamics, but also offers actionable strategies that enhance productivity management and project resilience.

5 Practical Implementation

Project managers can follow a structured approach to facilitate the integration of the fuzzy-chaos model into daily decision-making processes. This includes the initial setup of the model within a decision support system (DSS), regular updates of project data, and interpretation of output for proactive management. Specifically, the authors suggest incorporating the model into existing project management software platforms that support modular customization. This allows for seamless integration without requiring managers to switch between different systems.

Practical steps for adopting the fuzzy-chaos model in construction project management include:

- Initial setup: Integrating the model with existing project management tools or DSS
- Data entry: Regular input of project data into the model to reflect current conditions
- Analysis and interpretation: Using the model's outputs to make informed decisions based on the calculated entropy values, which indicate the level of uncertainty and potential risk
- Adjustments and re-evaluation: Continuously adjust project strategies based on model feedback and re-evaluate the results for improved decision-making

To further enhance the practical utility of the fuzzy-chaos framework, the authors propose its extension to include longitudinal analysis capabilities. This adaptation would allow project managers to track labour productivity trends over weeks and across entire project lifecycles. Continuous monitoring can be facilitated by integrating the framework with real-time data collection tools and advanced analytics platforms.

Such an approach would enable the identification of long-term trends and patterns in labor productivity, which are essential for strategic planning and resource allocation. It would also allow for the adjustment of project management strategies in response to observed trends, thereby enhancing the adaptability and effectiveness of the management process.

6 Conclusions

This study presents a novel methodology that integrates fuzzy logic and entropy calculations to evaluate labour productivity factors in construction projects. The proposed approach utilizes the strengths of FES to classify productivity-related factors into three distinct states: Poor, Fair, and Good. This classification provides an understanding of various productivity factors, enabling project managers to pinpoint specific areas that require attention or demonstrate strength within the project.

Further, entropy calculations are applied to quantify the uncertainty inherent in these assessments. Findings indicate that factors characterized by high entropy values typically reflect significant variability, influenced by external uncertainties such as supply chain disruptions and workforce fluctuations. These high entropy values signal areas requiring closer managerial attention and possibly implementing adaptive management strategies to mitigate potential risks. Conversely, factors with low entropy values exhibited stability and predictability, suggesting that existing management practices were effective and could be maintained to ensure ongoing project efficiency.

This methodology has implications for project management. By identifying factors with high entropy, project managers can better allocate resources, pre-emptively implement risk mitigation strategies, and dynamically adjust project plans to address emerging challenges. This approach also supports continuous monitoring of labour productivity, enabling project teams to respond swiftly to changes and maintain control over

project delivery.

While the fuzzy-chaos model is anticipated to provide significant insights into labor productivity, it is important to acknowledge its limitations. One such limitation is the model's reliance on the quality and availability of input data. Inaccurate or incomplete data can significantly affect the model's predictions and outputs. Additionally, the complexity of integrating chaos theory and fuzzy logic may present a steep learning curve for users unfamiliar with these concepts. Furthermore, the model's novel approach, while robust in handling project complexities, has not yet been tested across a broad spectrum of construction environments, which may limit its generalizability.

Looking ahead, further research could explore the integration of this methodology with advanced predictive analytics and machine learning techniques to enhance the accuracy of productivity assessments and entropy predictions. Additionally, applying this methodology across different construction projects, including residential, commercial, and infrastructure, would help validate its effectiveness and adaptability. Investigating the impact of incorporating real-time data feeds into the fuzzy logic assessments could further refine this tool, offering a more dynamic and responsive resource for project management.

In conclusion, this paper contributes an analytical framework to the field of construction project management, bridging the gap between qualitative assessments and quantitative analytics. It can be used to enhance project managers' ability to understand and manage labour productivity, fostering more resilient and efficient project environments. The methodology informs current project management practices as well as sets the stage for future innovations in project analytics.

References

[1] E. Pourrahimian, D. Salhab, F. Hamzeh, and S. AbouRizk. Entropy-centric framework for understanding and managing project dynamics in construction. *Automation in Construction*, 170:105928, 2025. doi.org/10.1016/j.autcon.2024.105928

[2] N. M. Hewahi. Incorporate doubts in fuzzy logic. *2019 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, pages 1–3, 2019. doi.org/10.1109/3ICT.2019.8910304

[3] E. Pourrahimian, D. Salhab, L. Shehab, F. Hamzeh, and S. AbouRizk. Application of chaos theory in project management. *Construction Research Congress*, pages 601–610, 2024. doi.org/10.1061/9780784485286.060

[4] S. A. Swanson. Keeping chaos out of complexity. *PM Network*, 26(8):56–59, 2012. www.pmi.org/learning/library/keeping-chaos-out-complexity-3842

[5] Z. Li and X. Zhang. On fuzzy logic and chaos theory: from an engineering perspective. In: Wang, P.P., Ruan, D., Kerre, E.E. (eds.), *Fuzzy Logic: A Spectrum of Theoretical & Practical Issues*, pages 79–97, Springer, Berlin, Heidelberg. 2007. doi.org/10.1007/978-3-540-71258-9_5

[6] A. A. Tsehayae and A. R. Fayek. System model for analysing construction labour productivity. *Construction Innovation*, 16(2):203–228, 2016. doi.org/10.1108/CI-07-2015-0040

[7] R. Assaad and I. H. El-adaway. Impact of Dynamic workforce and workplace variables on the productivity of the construction industry: New gross construction productivity indicator. *Journal of Management in Engineering*, 37(1):1–15, 2021. doi.org/10.1061/(asce)me.1943-5479.0000862

[8] M. H. W. Radosavljević. The evidence of complex variability in construction labour productivity. *Construction Management and Economics*, 20(1):3–12, 2002. doi.org/10.1080/01446190110098961

[9] C. Prakash, B. P. Rao, D. V. Shetty, and S. Vaibhava. Application of time and motion study to increase the productivity and efficiency. *Journal of Physics: Conference Series*, 1706(1):012126, 2020. doi.org/10.1088/1742-6596/1706/1/012126

[10] A. A. Javed, W. Pan, L. Chen, and W. Zhan. A systemic exploration of drivers for and constraints on construction productivity enhancement. *Built Environment Project and Asset Management*, 8(3):239–252, 2018. doi.org/10.1108/BEPAM-10-2017-0099

[11] H. R. Thomas, W. F. Maloney, R. M. W. Horner, G. R. Smith, V. K. Handa, and S. R. Sanders. Modeling construction labor productivity. *Journal of Construction Engineering and Management*, 116(4):705–726, 1990. doi.org/10.1061/(ASCE)0733-9364(1990)116:4(705)

[12] A. R. Fayek. Fuzzy logic and fuzzy hybrid techniques for construction engineering and management. *Journal of Construction Engineering and Management*, 146(7):1–12, 2020. doi.org/10.1061/(ASCE)CO.1943-7862.0001854

[13] A. Alshibani, B. E. Hafez, M. A. Hassanain, A. Mohammed, M. Al-Osta, and A. Bahraq. Fuzzy logic-based method for forecasting project final cost. *Buildings*, 14(12):3738, 2024. doi.org/10.3390/buildings14123738

[14] R., Canesi and C. D'Alpaos. A fuzzy logic application to manage construction-cost escalation. *Buildings*, 14(9):3015, 2024.

doi.org/10.3390/buildings14093015

[15] K. Shang and Z. Hossen. Applying fuzzy logic to risk assessment and decision-making. *Casualty Actuarial Society*, pages 1–59, 2013.

[16] A. Minasowicz, B. Kostrzewa, and J. Zawistowski. Construction project risk control based on expertise using fuzzy set theory. *Proceedings of the 28th International Symposium on Automation and Robotics in Construction, ISARC 2011*, pages 101–106, 2011. doi.org/10.22260/isarc2011/0015

[17] O. Moselhi and A. Salah. Fuzzy set-based contingency estimating and management. 2012 *Proceedings of the 29th International Symposium of Automation and Robotics in Construction (ISARC)*, 2012. doi.org/10.22260/isarc2012/0037

[18] B. Lal. A brief study report on the applications of chaos theory in real life. *Middle East Journal of Applied Science & Technology*, 6(4):52–58, 2023. doi.org/10.46431/mejast.2023.6404

[19] D. Watt and M. Willey. Fractal geometry in project bottleneck analysis. *Journal of Complexity in Construction*, 28(3):156–170, 2014.

[20] J. C. Galvez. Sensitivity analysis in project management using chaos theory principles. *Chaos, Solitons & Fractals*, 135, 109–121, 2020.

[21] O. Gomes. The emergence of chaos in productivity distribution dynamics. *Decisions in Economics and Finance*, 47(2):565–596, 2023. doi.org/10.1007/s10203-023-00419-9

[22] S. E. Christodoulou. Entropy-based methods for resource-constrained scheduling in projects. *Journal of Management in Engineering*, 26(4):204–211, 2010. doi.org/ 10.1016/j.autcon.2009.04.007

[23] M. Vanhoucke, and J. Batselier. Entropy in project management: A measure of uncertainty. *Project Management Journal*, 50(5), 555–567, 2019.

[24] M. Raoufi and A. R. Fayek. Framework for identification of factors affecting construction crew motivation and performance. *Journal of Construction Engineering and Management*, 144(9):04018080, 2018. doi.org/10.1061/(ASCE)CO.19437862.0001543

[25] E. Pourrahimian, A. Eltahan, D. Salhab, J. Crawford, S. AbouRizk, and F. Hamzeh. Integrating expert insights and data analytics for enhanced construction productivity monitoring and control: a machine learning approach. *Engineering, Construction and Architectural Management*, ahead-of-print, 2024. doi.org/10.1108/ECAM-02-2024-0268

[26] Y. Zhang, and X. Y. Wang. Fuzzy neural network based on a Sigmoid chaotic neuron. *Chinese Physics B*, 21(2), 2012. doi.org/10.1088/1674-1056/21/2/020507

[27] L. Amezquita, O. Castillo, J. Soria, and P. Cortes-Antonio. New variants of the multi-verse Optimizer algorithm adapting chaos theory in benchmark optimization. *Symmetry*, 15(7):1319, 2023. doi.org/10.3390/sym15071319

Video-Based Productivity Monitoring of Worker and Large-Scale Object Interactions in Construction Sites

Mik Wanul Khosiin[1], Jacob J. Lin[1], Eko Andi Suryo[2], Kartika Puspa Negara[2], Ismiarta Aknuranda[3], and Chuin-Shan Chen[1]

[1]Department of Civil Engineering, National Taiwan University, Taipei City, Taiwan
[2]Department of Civil Engineering, Universitas Brawijaya, Malang City, Indonesia
[3]Department of Information Systems, Universitas Brawijaya, Malang City, Indonesia

mikwk@caece.net, jacoblin@ntu.edu.tw, dchen@ntu.edu.tw

Abstract -

Automating productivity monitoring is crucial for improving the construction industry. To measure productivity, we should identify which worker works for what object and their relationship. The lack of understanding between human and object interaction in a large-scale format from video surveillance has become a significant challenge in construction sites. However, the existing vision-based studies only focus on object detection and activity recognition, which do not recognize workers, objects, and actions simultaneously. This situation makes managers unable to measure the productivity of the workers effectively. To address the issue, this study applies the HOI technique, which consists of object detection tasks and interaction prediction tasks through faster R-CNN and graph neural networks (GNN). There are two groups of actions in this interaction, including productive (installing, preparing, and transporting) and non-productive actions (no interaction) on the formwork structure. Our model achieves 0.674, 0.556, and 0.632 mAP scores of the local area, global area, and average area of the objects sequentially, indicating that the model can monitor construction productivity effectively. For future studies, utilizing more information, such as temporal and body postures of workers, can potentially improve the performance of the HOI model for the productivity monitoring process.

Keywords -

Productivity monitoring; human-object interaction; graph neural networks (GNN)

1 Introduction

The construction industry is less competitive than other sectors. From 2000 to 2022, It only grew 0.4 percent annually and has a 1.6 percent and 2,6 percent gap towards the world total economy and manufacturing industry simultaneously ([1]). Even though global productivity was boosted by China's improvement (4 percent) in 2021, it shows signs of slowing down due to the ineffective workforce monitoring process. The main issue is that managers cannot improve their performance because they lack understanding of the interaction between workers, objects, and their actions ([2]) simultaneously. This situation brings practical problems, such as rework activities, defective products, construction accidents, etc. Moreover, the current technology uptake is mainly focused on controlling systems rather than productivity. This means a comprehensive investigation of workers is essential to help the managers in the decision-making process.

Currently, to understand the interactions between humans and objects in construction sites, managers must conduct real observations consistently ([3], [4]), which is time-consuming and low accuracy. The current technologies applications in construction sites only focus on detection and activity recognition, and their interaction has not yet been addressed ([5]). This means the prompt about which workers work for what objects remains unsolved and becomes a crucial issue in the productivity monitoring system ([6]). Moreover, construction sites have abundant data collected by managers for daily reports from the site's video surveillance. This group of data is dominated by large-scale instances for both humans and objects. However, this data is only used for administrative purposes and has not been organized and utilized optimally for the productivity monitoring process. Thus, the potential technology is required to explore this data and generate valuable information about worker activity.

Human-object interaction (HOI) [7] represents a transformative vision-based methodology that enables simultaneous detection and analysis of humans, objects, and their interactions within an image or video. This technique is particularly valuable on construction sites, where monitoring worker productivity and interactions with various elements is crucial. By employing HOI models, supervisors can gain a detailed understanding of how workers engage with tools and materials, which can significantly enhance operational efficiency and safety protocols. The ability to identify and categorize these interactions in real time offers numerous benefits. It allows for immediate responses to safety hazards or inefficiencies, and the data collected

can be used to refine workflow designs and worker training programs. Moreover, integrating HOI technology into existing surveillance systems transforms passive monitoring into an active management tool. This facilitates a more proactive approach to project management and quality control, ensuring that all activities on site align with safety and productivity standards. Through the precise detection of interactions, HOI technology not only boosts site management but also contributes to the development of smarter, safer construction environments.

To address this challenge, this study proposes the HOI utilizing graph theory in large-scale instances to understand the connection between workers and their elements. This study consists of five sections: first, the Introduction presents the topic and its importance, while Related Studies reviews previous work and contextualizes the research. The Methodology section explains the approaches and techniques used for HOI detection. The experiment details the dataset preparation and results of practical tests conducted. Finally, the Conclusion summarizes the findings and suggests future research directions.

2 Related Studies

To understand the process of the HOI technique ([8], [9], [9], and [10]), we need to compare it with the activity recognition task. In the activity recognition task, there are only two steps, including object detection and activity prediction. Object detection in this task only localizes the humans and uses their spatial and temporal features to determine the type of action. In Figure 6, the activity recognition can identify the worker doing the installing action on the particular object. However, without detecting the objects and generating the interaction with the humans, activity recognition can not address a practical issue in construction sites that requires a comprehensive understanding of workers for the productivity monitoring process.

While in the HOI task, this technique offers more complete detection information, such as humans, objects, connection, and action prediction. To obtain those outputs, the HOI task requires three steps, consisting of object detection, interaction detection, and activity prediction, where this approach has one more additional step than the activity recognition task. After evaluating both tasks, we can conclude that there are two important processes, including object detection and activity prediction. To have an accurate HOI prediction, four elements from the frames must be extracted: humans, objects, interaction, and actions with graph neural networks (GNN)([11]). Thus, we summarize the recent construction worker and objects-related studies in Table 1 below that utilized several approaches to highlight the research gap in this study. In addition, the two important tasks will also be elaborated in this section to highlight the technical gap in this study.

Table 1. The recent studies in monitoring construction workers

Authors	Worker monitoring	Limitations
Q. Zhang ([12])	object detection	no obj., act., and int.
Y. Sun ([13])	object detection	object only
R. Mutegeki ([14])	activity recognition (CNN-LSTM)	no obj., and int.
S. Garcia ([15])	activity recognition (YOLO)	no obj., and int.
M.-Y. Cheng ([16])	action recognition (YOWO)	no obj., and int.
S. Tang ([17])	HOI for safety	safety only

Object detection offers classification and localization tasks. We can identify the type and position of the objects in each frame by using several existing models, such as You Only Look Once (YOLO) ([18]), Faster R-CNN ([19]), Mask R-CNN ([20]), and others. In the current construction practices, Q. Zhang ([12]) detects the construction workers and classifies them based on the color of the hat using Faster R-CNN. However, this study has not considered the materials around the workers and the type of action of the workers, so this study provides only one HOI element. Moreover, Y. Sun ([13]) offers to detect construction materials, such as concrete, brick, metal, wood, and stone, through the VGG16 model. However, they focus solely on the objects and ignore the other two important elements in HOI tasks (humans and actions). Therefore, the object detection task is not enough to fulfill the required information in construction practices. We need to utilize potential techniques to enhance the recognition level in construction sites.

Activity recognition is the deep learning task that mainly focuses on understanding the type of action humans take. The current models can help us conduct action recognition consisting of Long Short-Term Memory (LSTM) ([21]), YOLO, and You Only Watch Once (YOWO) ([22]), and etc. Some studies have implemented these techniques in construction sites, like R. Mutegeki ([14]) recognizing the worker action by combining Convolution Neural Networks (CNN) ([23]) and LSTM models. Furthermore, S. Garcia ([15]) and M.-Y. Cheng ([16]) observed the worker activity by using YOLO and YOWO sequentially. However, Those studies did not provide information about objects and their interaction in construction sites. Consequently, construction practitioners cannot fully understand what materials are being used or processed by which workers.

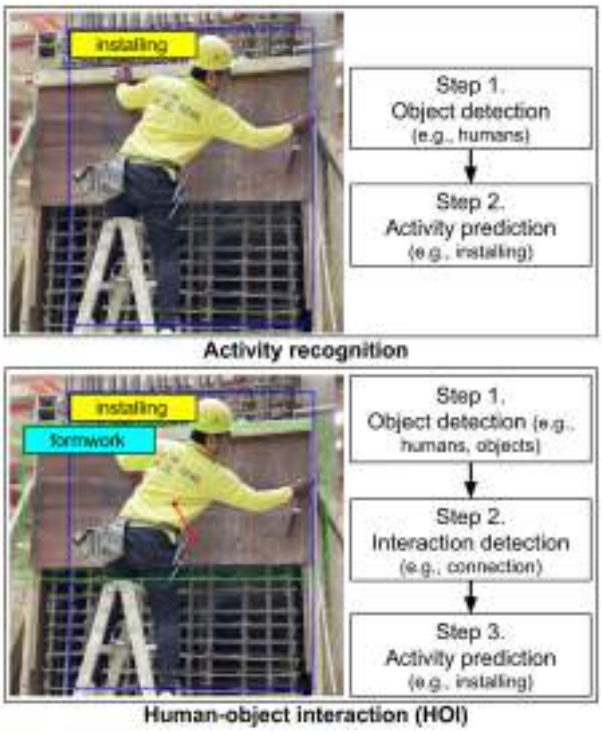


Figure 1. Activity recognition and HOI task comparison

This incomplete situation triggers us to find a model that can provide comprehensive information about workers.

From Table 1 and In Figure 6, we can imply that the current studies lack interaction information among the workers. Moreover, the objects located around workers have not been completely recognized, which adds more challenges. Therefore, identifying humans, objects, actions, and their interaction as a single information is an essential issue to enhance the construction monitoring process.

3 Methodology

Based on the previous discussion, this study also has three steps (Figure 2) in utilizing the HOI technique for productivity monitoring in construction sites, such as Step 1. Object detection, and Step 2. Interaction prediction with GNN, and Step 3. Activity recognition. In Step 1, we utilize Faster R-CNN ([19]) to localize the objects and extract potential features, such as visual features, semantic features, and spatial features. Then, these features will be stored in the fully connected layers (fc7) as input for the GNN ([11]) in the next step. In Figure 2, W1, W2, O1, and O2 stand for worker-1, worker-2, object-1, and object-2, which represent the detection results that consist of two HOI elements, humans and objects. These results will be considered as the node data, the initial information used to build the edge between one node and another. Furthermore, despite using a large-scale format, we still consider many to many detections, which means we can potentially identify crowd humans and objects in each frame. From those two data, we can generate multiple interactions to obtain a better understanding of the worker's activities. In other words, if we detect fewer people and objects in the image, we will also have less interaction prediction, or it relies on the performance of the object detection task. Finally, object detection is crucial in HOI; the higher the object detection performance, the better the productivity monitoring results will be achieved.

In Step 2, the features from the previous step will be processed in the GNN. We identify the potential interaction or edge candidates (red arrow) through attention layers to update the features in each node and construct the final graph structures. We utilize the graph attention networks (GAT) ([24], [25]) as one of the variations of GNN structure that offers a comprehensive node's analysis. Then, we follow these mathematical operations (Eq. 1 and Eq. 2) as the GAT concept to prepare the HOI model.

$$G = (V, E, \{h'_1, h'_2, h'_3, \ldots, h'_n\}) \tag{1}$$

$$h'_i = \sigma\left(\sum_{j \in N(i)} \alpha_{ij} W h_j\right) \tag{2}$$

Where G is defined as a graph, V represents vertices that consist of $v_1, v_2, \ldots, v_n$, and E stands for the edge of the graph $\{w_{ij} \in V\}$. Before we build the Graph, we should calculate h'_i as the updated feature vectors for each node i, obtained by applying a non-linear activation function σ (such as the ReLU function) to a weighted sum of the transformed feature vectors Wh_j of neighboring nodes j in the set $N(i)$. The weights α_{ij} are attention coefficients obtained from the training process through the attention networks that determine the importance of each neighbor's features to node i, computed by the network based on the nodes' relative positions or feature similarities. Once we have the updated node vector h'_i, we can combine the whole features and generate the final graph structure for the prediction process. In addition, since we most likely detect many workers and objects in one frame, this study also proposes the multiple-heterogeneous interaction prediction. Ultimately, we will gain a thorough understanding of the intricate interactions between workers and the objects they handle on construction sites.

In Step 3 of the workflow, the activity recognition utilizes enriched graph data from GNNs to predict specific interactions like "installing" or "transporting" between workers and objects on construction sites. This stage critically assesses interaction dynamics using contextual and relational data of the nodes, which represent workers and

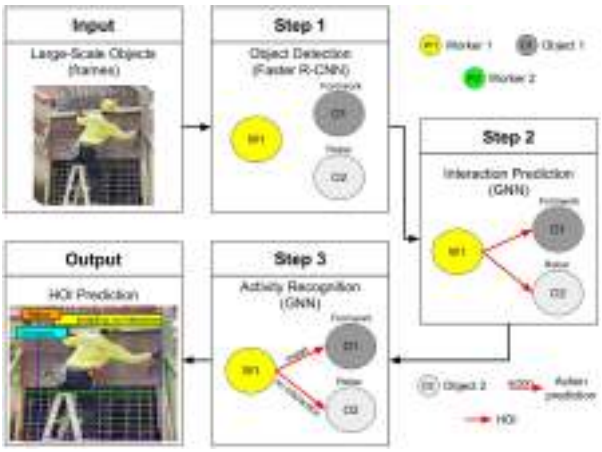


Figure 2. General workflow of productivity monitoring using HOI technique

objects. The GNN refines features and attention weights to highlight significant interactions based on historical and spatial-temporal relationships, providing supervisors with insights for resource allocation and workflow optimization. For evaluation, the model's performance is quantified using the mean Average Precision (mAP) score, which assesses the precision of the predicted interactions across various confidence thresholds. This metric helps ensure the accuracy and reliability of the interaction predictions, supporting effective decision-making in productivity monitoring.

In conclusion, the implementation of HOI techniques using GNNs offers a transformative approach to productivity monitoring on construction sites. This research, structured in three strategic steps: object detection, interaction prediction, and activity recognition, enables precise monitoring and analysis of worker interactions with site elements. Applying GNNs facilitates real-time processing and complex interaction analyses, significantly improving operational efficiency and safety management. The adaptability of this methodology to diverse construction environments underscores its potential to revolutionize project management strategies. By leveraging such advanced technologies, construction managers can optimize workflows, enhance safety protocols, and improve overall resource allocation. Continued advancements and integration of broader data types and interaction scenarios could further elevate the effectiveness of this model, setting new benchmarks in technology-driven construction management and enhancing both efficiency and safety on sites.

4 Experiment

We utilize the supplementary material from M.-Y. Cheng, 2023 [16] provided a set of figures about construction workers who work in installing formwork structures in a large-scale and high-resolution format. We collected 1250 images and divided them into training (1000 images) and testing sets (250 images). We use the Hico-det-UI ([26]) for the labeling process, and we provide two labels (bounding boxes and the action ID) and three classes, which consist of worker, formwork, and steel rebar. Moreover, since those materials are installed in large locations in construction sites, we implement local and global areas that refer to the associated area of objects with direct interaction with the workers. Then, we prepare a list of potential actions, such as preparing and installing the formwork objects (productive actions) collected from a lexical database from English (WordNet)([27]) and no-interaction as non-productive actions. Then, to process the model and dataset, we use the high-performance hardware NVIDIA GeForce RTX 4090 with a GPU memory of 24.5GB and 131GB for RAM.

We trained our model in two steps: object detection and GNN for the HOI tasks. For the Faster R-CNN, we used some training parameters, such as 2500 epochs, batch size 20, and learning rate 0.001, and we achieved 0.83 for the mAP@0.5. Then, in the next step, we train our HOI model with these conditions (learning rates 1e-05, batch size 32, drop probability 0.3, and epochs 300. We evaluate the model with IoU more than 0.5, which means if the intersection over the union of each bounding box is more than 0.5, we consider it as the candidate for building the graphs. Moreover, we define the performance as a true positive if the three predicted HOI elements (humans, objects, and actions) are suitable with the ground truth, and we consider false positives if the model can predict the classes that are not available in our datasets. We achieve the 0.674, 0.556, and 0,632 mAP scores (Table 5) for the local area, global area, and average area simultaneously as the evaluation report for our datasets.



Figure 3. Installing formwork [16]

Table 2. Installing action prediction, Fig 3

Items	Humans	Objects	Interactions
1	H0 worker-1	00. rebar-local area	no interaction (0.001)
2	H0. worker-1	01. formwork-local area	installing (0.008)
3	H1. worker-2	02. formwork-local area	preparing (0.011)



Figure 4. Preparing formwork [16]

Table 3. Preparing action prediction, Fig 4

Items	Humans	Objects	Interactions
1	H0. worker-1	01. rebar-local area	no interaction (0.001)
2	H0. worker-1	02. formwork-local area	no interaction (0.011)
3	H0. worker-1	02. formwork-local area	preparing (0.493)



Figure 5. Transporting formwork[16]

From Figure 3 and Table 2. we can see that the workers interact with the nearest objects; for instance, H0 (worker 1) has an installing action with the formwork but no interaction with the steel rebar in the local area. Then, the worker (H1) interacts with the stacks of formwork as the preparing action. In Figure 4 and Table 3, worker-1 (H0) mainly focuses on the formwork board or preparing the formwork structures, but the surrounding entities lack rebar rods and formwork components are also detected as the preparation, inspection, and no interaction activity. The last qualitative result is the transporting works (Figure 5, and Table 4); the two workers are collaborating in transporting the formwork's elements. Here, worker-1 (H0) transports the formwork-local area (01), while worker-2 (H1) transports the small sheet of board (formwork-local area 02) in a different area. In addition, we also compare

Table 4. Transporting action prediction, Fig 5

Items	Humans	Objects	Interactions
1	H0. worker-1	01. formwork-local area	transporting (0.015)
2	H0. worker-1	04. formwork-global area	preparing (0.010)
3	H1. worker-2	01. formwork-local area	preparing (0.010)
4	H1. worker-2	02. formwork-local area	transporting (0.05)
5	H1. worker-2	03. formwork-local area	preparing (0.266)

Table 5. Performance report of the HOI tasks

Actions	AP local area	AP global area	AP average area
Installing formwork	0.844	0.540	0.681
Preparing formwork	0.644	0.603	0.626
Transporting formwork	0.533	0.525	0.588
mAP score	0.674	0.556	0.632*

Table 6. mAP (%) score comparison with the previous studies in HICO-DET groups

HOI models	mAP average area
HO-RCNN [8]	0.078
PMN [28]	0.212
VS-Gats [10]	0.203
ViPLO [29]	0.372
PVIC-SwinL [30]	0.443
Ours	0.632

the results with the existing HOI model as a validation process in which our average area (mAP 0.632) is higher 0.189 than PVIC-SwinL as the current highest performance HOI model (Table 6), especially in the HICO-DET domain. This result indicates that the model can comprehensively localize the objects on both large and small scales and predict the action connection accurately.

After we obtain and evaluate the results, managers can monitor the productivity of each worker. Productivity can be categorized by the type of detected action in the model. From the predicted results in each frame, we can conclude that they are productive workers with the installing, preparing, and transporting actions for the formwork structure as the major objects. Moreover, worker compliance, such as wearing standard safety attributes, using complete tools, and having a high dedication to working as professional workers, can also be considered by managers to determine the productivity of the workers. However, minor objects that do not have direct interaction with the workers, such as steel rebar or unused formwork, are predicted as no-interaction actions. This means the workers are not productive in minor objects but perform well in major objects. In addition, the prediction in the other frames can have different results; the major objects can be non-productive if the worker has no interaction actions at the prediction level. Therefore, to have an accurate analysis, Figure (6) shows the total number of the detected actions for all workers in the construction sites. Although the productive actions (preparing) are dominated in the frame, the prediction number of no-interaction actions is also high, so managers should arrange a better scenario to address this issue.

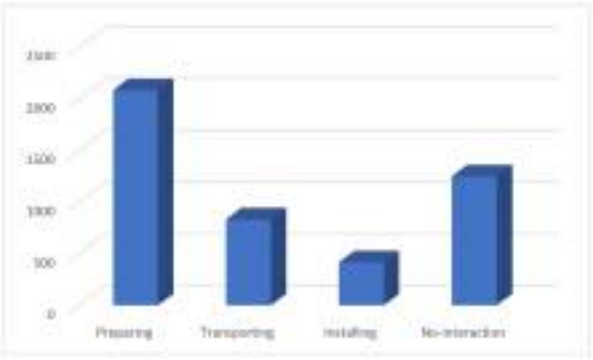


Figure 6. Productivity monitoring results

As the discussion part, even though we achieved excellent performance results, some limitations still exist. Our model has an imbalance data issue that prevents us from reaching better results. The public datasets we obtained do not represent the real conditions in construction sites. The current datasets lack variations, and our datasets might not perform well if a new set of data (unseen) is being used for the testing process. Moreover, our study focuses more on static images in which the dynamic interaction of the workers has not been addressed yet. To conclude, for future studies, we recommend utilizing more features, such as temporal features from the video or real-time scenario and body pose features, to improve the accuracy of the action prediction. Advanced methods, such as vision-language models (VLM), can also be utilized to improve the detection, apply the segmentation, solve the data imbalance, and process the semantic feature much better.

5 Conclusion

This study investigates the missing link or connection between the workers and objects that play an important role in measuring the productivity monitoring process. While the current studies focus solely on object detection and activity recognition only, we propose the HOI approach to address this issue. Our study achieves accurate results in both object detection (Step 1) and interaction prediction (Step 2). Thus, we can identify which workers are productive and non-productive using action recognition. However, the imbalanced data as the main prompt in this study must be addressed with current advanced technology and other effective scenarios. Thus, we can provide the level of the assessment process to improve the productivity level of the construction industry in the future.

Acknowledgment This research received support from four institutions: first, the Ministry of Education (MoE) of Taiwan provided full funding for the doctoral program. Second, the AI Research Center and the Department of Civil Engineering at National Taiwan University (NTU), contributed research funding and high-performance facilities. Third, the AI Center at Brawijaya University (UB) in Indonesia offered research partnership grants (95/UN10.F07.11/KS/2024) and provided annotation teams to assist in completing the study. In addition, the National Science and Technology Council (NSTC) of Taiwan (Case No. 114PGFA0100550) facilitated the international conference expenses. Thus, this study can be finished, accepted, and presented at ISARC 2025.

References

[1] Koen Vermeltfoort-Birgit Biemans Jan Mischke, Kevin Stokvis. Delivering on construction productivity is no longer optional. Technical report, McKinsey Company, 2024. URL `https://www.mckinsey.com/`.

[2] Khosiin, Lin, and Chen. Worker accountability in computer vision for construction productivity measurement: A systematic review.

Korea Institute of Construction Engineering and Management, pages 775–782, july 2024. doi:https://doi.org/10.6106/ICCEPM.2024.0775.

[3] Jiukun Dai, Paul M. Goodrum, William F. Maloney, and Cidambi Srinivasan. Latent structures of the factors affecting construction labor productivity. *Journal of Construction Engineering and Management*, 135(5):397–406, 2009. doi:10.1061/(ASCE)0733-9364(2009)135:5(397).

[4] O. Salem, J. Solomon, A. Genaidy, and I. Minkarah. Lean construction: From theory to implementation. *Journal of Management in Engineering*, 22(4):168–175, 2006. doi:10.1061/(ASCE)0742-597X(2006)22:4(168).

[5] Shang-Hsien Hsieh Mani Golparvar-Fard Arita Pal, Jacob J. Lin. Activity-level construction progress monitoring through semantic segmentation of 3d-informed orthographic images. *Automation in Construction*, RR156-11:105157, 2024. ISSN 0926-5805. doi:https://doi.org/10.1016/j.autcon.2023.105157.

[6] Construction Industry Institute (CII). Engineering productivity measurement. *Construction Industry Institute*, RR156-11:286, 2001.

[7] Georgia Gkioxari, Ross Girshick, Piotr Dollár, and Kaiming He. Detecting and recognizing human-object interactions, 2018. URL `https://arxiv.org/abs/1704.07333`.

[8] Yu-Wei Chao, Yunfan Liu, Xieyang Liu, Huayi Zeng, and Jia Deng. Learning to detect human-object interactions. *CoRR*, abs/1702.05448, 2017. URL `http://arxiv.org/abs/1702.05448`.

[9] Hongsheng Li, Guangming Zhu, Wu Zhen, Lan Ni, Peiyi Shen, Liang Zhang, Ning Wang, and Cong Hua. Spatial parsing and dynamic temporal pooling networks for human-object interaction detection. In *2022 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2022. doi:10.1109/IJCNN55064.2022.9892087.

[10] Zhijun Liang, Junfa Liu, Yisheng Guan, and Juan Rojas. Visual-semantic graph attention networks for human-object interaction detection. In *2021 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pages 1441–1447, 2021. doi:10.1109/ROBIO54168.2021.9739429.

[11] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1):61–80, 2009. doi:10.1109/TNN.2008.2005605.

[12] Qilin Zhang, Zhichen Wang, Bin Yang, Ke Lei, Binghan Zhang, and Boda Liu. Reidentification-based automated matching for 3d localization of workers in construction sites. *Journal of Computing in Civil Engineering*, 35(6):04021019, 2021. doi:10.1061/(ASCE)CP.1943-5487.0000975.

[13] Ying Sun and Zhaolin Gu. Using computer vision to recognize construction material: A trustworthy dataset perspective. *Resources, Conservation and Recycling*, 183:106362, 2022. ISSN 0921-3449. doi:https://doi.org/10.1016/j.resconrec.2022.106362.

[14] Ronald Mutegeki and Dong Seog Han. A cnn-lstm approach to human activity recognition. In *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pages 362–366, 2020. doi:10.1109/ICAIIC48513.2020.9065078.

[15] Sagrario Garcia-Garcia and Raúl Pinto-Elías. Human activity recognition implenting the yolo models. In *2022 International Conference on Mechatronics, Electronics and Automotive Engineering (ICMEAE)*, pages 127–132, 2022. doi:10.1109/ICMEAE58636.2022.00029.

[16] Min-Yuan Cheng, Akhmad F.K. Khitam, and Hongky Haodiwidjaya Tanto. Construction worker productivity evaluation using action recognition for foreign labor training and education: A case study of taiwan. *Automation in Construction*, 150:104809, 2023. ISSN 0926-5805. doi:https://doi.org/10.1016/j.autcon.2023.104809.

[17] Shuai Tang, Dominic Roberts, and Mani Golparvar-Fard. Human-object interaction recognition for automatic construction site safety inspection. *Automation in Construction*, 120:103356, 2020. ISSN 0926-5805. doi:https://doi.org/10.1016/j.autcon.2020.103356. URL `https://www.sciencedirect.com/science/article/pii/S0926580520309365`.

[18] Glenn Jocher, Alex Stoken, Jirka Borovec, NanoCode012, ChristopherSTAN, Liu Changyu, Laughing, tkianai, Adam Hogan, lorenzomammana, yxNONG, AlexWang1900, Laurentiu Diaconu, Marc, wanghaoyang0106, ml5ah, Doug, Francisco Ingham, Frederik, Guilhen, Hatovix, Jake Poznanski, Jiacong Fang, Lijun Yu , changyu98, Mingyu Wang, Naman Gupta, Osama Akhtar, PetrDvoracek,

and Prashant Rai. ultralytics/yolov5: v3.1 - bug fixes and performance improvements, October 2020. URL https://doi.org/10.5281/zenodo.4154370.

[19] Huaizu Jiang and Erik Learned-Miller. Face detection with the faster r-cnn. In *2017 12th IEEE International Conference on Automatic Face Gesture Recognition (FG 2017)*, pages 650–657, 2017. doi:10.1109/FG.2017.82.

[20] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross B. Girshick. Mask R-CNN. *CoRR*, abs/1703.06870, 2017. URL http://arxiv.org/abs/1703.06870.

[21] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8): 1735–1780, 1997. doi:10.1162/neco.1997.9.8.1735.

[22] Okan Köpüklü, Xiangyu Wei, and Gerhard Rigoll. You only watch once: A unified CNN architecture for real-time spatiotemporal action localization. *CoRR*, abs/1911.06644, 2019. URL http://arxiv.org/abs/1911.06644.

[23] Kunihiko Fukushima and Sei Miyake. Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position. *Pattern Recognition*, 15(6):455–469, 1982. ISSN 0031-3203. doi:https://doi.org/10.1016/0031-3203(82)90024-3.

[24] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks, 2018. URL https://arxiv.org/abs/1710.10903.

[25] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *CoRR*, abs/1706.03762, 2017. URL http://arxiv.org/abs/1706.03762.

[26] Yu-Wei Chao, Zhan Wang, Yugeng He, Jiaxuan Wang, and Jia Deng. HICO: A benchmark for recognizing human-object interactions in images. In *Proceedings of the IEEE International Conference on Computer Vision*, 2015.

[27] George A. Miller. Wordnet: a lexical database for english. *Commun. ACM*, 38(11): 39–41, November 1995. ISSN 0001-0782. doi:10.1145/219717.219748.

[28] Zhijun Liang, Junfa Liu, Yisheng Guan, and Juan Rojas. Pose-based modular network for human-object interaction detection. *CoRR*, abs/2008.02042, 2020. URL https://arxiv.org/abs/2008.02042.

[29] Jeeseung Park, Jin-Woo Park, and Jong-Seok Lee. ViPLO: Vision Transformer Based Pose-Conditioned Self-Loop Graph for Human-Object Interaction Detection . In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17152–17162, Los Alamitos, CA, USA, 2023. IEEE Computer Society. doi:10.1109/CVPR52729.2023.01645.

[30] Frederic Z Zhang, Yuhui Yuan, Dylan Campbell, Zhuoyao Zhong, and Stephen Gould. Exploring predicate visual context in detecting of human-object interactions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10411–10421, October 2023.

Development of a Fall Detection and Safety Communication System Using Small Language Models

Malithi Wanniarachchi Kankanamge [1], Abdur R. Shahid[1], Ning Yang[1], S M Jamil Uddin[2], Rahul Biswas[3]

[1]School of Computing, Southern Illinois University, Carbondale, IL, USA
[2]Department of Construction Management, Colorado State University, USA
[3]Faculty of Engineering, Rangamati Science and Technology University, Bangladesh

malithi.mithsara@siu.edu, shahid@cs.siu.edu, nyang@siu.edu, smj.uddin@colostate.edu, rahulbiswas.cse@gmail.com

Abstract -

Fall incidents are a leading cause of injuries and fatalities in the construction industry, significantly impacting worker safety and productivity. Although many AI-based automated fall detection methods have been introduced, existing systems lack continuous communication support and often fail to address critical scenarios such as isolated work zones or high-risk tasks involving limited oversights. To address these limitations, this study proposes a new fall detection system for the construction industry using small language models (SLMs). Incorporating real-time conversation support within the system improves communication with emergency care teams and increases its utility in high-risk environments. We present the system architecture, integrate lightweight machine learning models, and evaluate the system's performance using the TinyLlama and Phi-3 models. Our assessment, which emphasizes relevance, correctness, and response speed, provides essential insights into effectively integrating language models into high-reliability systems for the construction industry.

Keywords -

Small Language Models, Fall Detection

1 Introduction

Recent advancements in Artificial Intelligence (AI) and the Internet of Things (IoT) have significantly transformed safety protocols across industries by enabling innovative monitoring and emergency response systems. In the construction sector, known for its inherently hazardous environments, these technologies offer a unique opportunity to mitigate risks and enhance workers safety. Falls remain one of the most pressing safety concerns in construction, as they are the leading cause of fatalities [1]. With the rapid advancement of wearable sensors, safety monitors, and video surveillance systems, effective AI approaches have revolutionized fall detection in the construction field. These technologies enable the analysis of complex motion and safety data, provide tailored insights into worker safety, and improve communication between workers and safety supervisors, significantly enhancing the efficiency and effectiveness of remote safety monitoring systems on construction sites [2]. According to the Occupational Safety and Health Administration (OSHA), falls accounted for 351 out of 1,008 construction-related fatalities in 2020. In 2021, falls caused 37% of construction-related deaths, and another 8% of fatalities were due to workers being struck by objects that were swinging, falling, or misplaced. Fatalities caused by falls from elevation continue to be a leading cause of death for construction employees, accounting for 395 of the 1,069 construction-related fatalities recorded in 2022, according to Bureau of Labor Statistics (BLS) data [3].

Various methodologies, including video and wearable sensing-based, have been introduced to facilitate automatic fall detection in both research and consumer products. Despite numerous efforts to enhance fall detection in construction fields, state-of-the-art approaches have several limitations[4], which can be summarized as their lack of continuous communication support and insufficient support for construction sites. Once a fall has been detected, systems often fail to provide continuous communication support. This limitation is critical, as timely and ongoing communication with emergency care teams and construction authorities can significantly improve outcomes for the affected individual.

Given these drawbacks, our work suggests improving fall detection system using small language models (SLMs). These approaches can enable rapid and continuous communication with emergency care workers by incorporating real-time conversation support. Moreover, with the recent advancements in language models, this approach enhances the responsiveness of fall detection system and extends their applicability to vulnerable populations, including those in construction sites. We introduce a novel fall detection system using small language models for the construction sites. This system demonstrates how these language models can be integrated with lightweight machine learning models to enhance trust and reliability, addressing the challenges associated with the unreliability of



Figure 1. Some example companies working on the utilization of AI for fall detection in construction sites.

language models in critical applications. Furthermore, we simulate the proposed system using the TinyLlama[5] and Phi-3[6] models and evaluate its performance in terms of accuracy, response time, and relevance of responses. Our experiments provide valuable insights into how these language models can be effectively implemented in systems that require high reliability and responsiveness.

1.1 Contributions of the Paper

This paper makes significant contributions to the advancement of the fall detection and emergency response system for construction applications- enhancing the safety of workers on sites. We propose a fall detection system designed to enhance safety and reliability on construction sites. This system integrates lightweight machine learning models with small language model-based AI agents to provide an efficient solution to the one of the most important in construction. The research identifies and addresses significant limitations of existing fall detection system, such as the lack of continuous communication support and inadequate assistance for vulnerable populations. By incorporating real-time conversation capabilities, the proposed system ensures continuous communication with emergency care teams. By incorporating real-time conversation capabilities, the proposed system ensures continuous communication with emergency care teams. We simulate the proposed system using the TinyLlama and Phi-3 models. We evaluate the system's performance in terms of accuracy, response time, and relevance of responses, providing insights into the practical implementation of language models in high-reliability systems.

2 Related Work

The research presented in this paper represents several key topics, including fall detection and language models. This section reviews some of the state-of-the-art works closely related to our study.

2.1 Video-Based Fall Detection in Construction Sites

In construction sites, ongoing research is being conducted on fall detection. Liu et al. [4] focus on workers' postures and aim to directly detect fall indicators using a computer vision (CV)–based non-contact approach that utilizes the K-means algorithm. Khan et.al.[7] introduce fall from height as a computer vision-based fall detection in construction sites by incorporating inertial measurement unit (IMU) sensors, altimeters, LED indicators, and buzzers to monitor workers' safety at heights. Numan[8] introduces a method that focuses on enhancing the safety of lone workers in construction sites through a deep learning-based fall detection system. These technologies detect falls but do not support continuous communication.

2.2 Industry-Wide Effort to Enhance Construction Site Safety with AI-Driven Fall Detection Systems

We review the current efforts of industries to utilize AI for fall detection on construction sites (Figure 1). Several companies are actively working in this area. For example, Synaedge[9] focuses on real-time fall detection using video surveillance. Similarly, Viact[10] specializes in fall detection through video analysis. Another notable company, Slatesafety[11], focuses on fall detection using wearable sensor data. Their system generates real-time alerts to supervisors, typically via mobile devices, when a "fall is detected". Kwant[12] is another company working on detecting worker falls using smart wearable devices, providing real-time alerts. AlwaysAI[13] is another organization utilizing AI to detect objects and specific activities, including fall detection. Visionify[14] has introduced a slip-and-fall detection system that monitors falls in construction areas in real time using cameras. This system is deployed in risk-prone zones such as ladders, ramps, and scaffolding to ensure immediate alerts to supervisors and rapid responses to potential dangers. IntelliSee[15] upgrades existing security cameras into proactive safety tools capable of identifying falls in real-time. Its AI technology continuously monitors video feeds, analyzing movement and body positioning to detect when a fall has occurred. Once detected, IntelliSee immediately sends alerts to designated personnel, providing precise details about the incident's location.

2.3 Large and Small Language Models and Fine-tuning

Language models come in a wide range of sizes, each offering distinct advantages and challenges. Large-scale models, such as OpenAI's GPT-3 and GPT-4[16], are characterized by extensive parameters and vast training datasets. While increasing model size typically enhances performance, it also complicates fine-tuning, especially

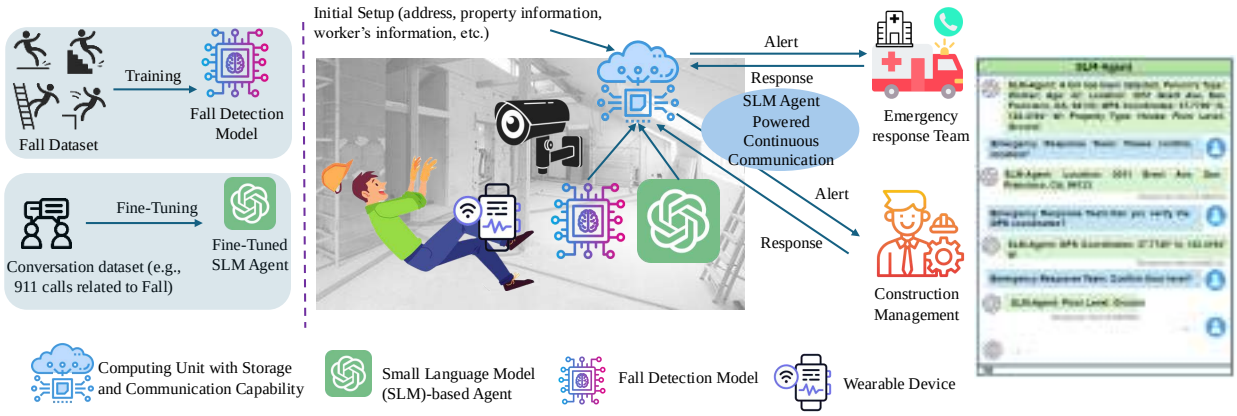


Figure 2. Overall Architecture of the Proposed Framework: Fall Detection in Construction Sites with a camera and server stores the fall detection model, fine-tuned SLM, and initial data. The fine-tuned SLM communicates with an emergency team when a fall incident occurs.

due to the high computational and financial costs associated with platforms like OpenAI's fine-tuning services. In contrast, smaller language models are more suitable for fine-tuning in low-resource settings. Examples of open-source small models include Tiny-LLama[5], GPT-2[17], and Phi-3[6]. For our approach, we chose Tiny-LLama (1.1B) because of its efficiency in low-resource environments and its impressive performance[5]. Fine-tuning is the process of taking a pre-trained model and adapting it to a specific task or domain by training it further on specialized data. This approach adjusts the model's parameters to fit the new dataset better. There are two primary types of supervised fine-tuning: instruction tuning and alignment. Instruction tuning improves a model's ability to handle user queries by training it on instruction-formatted datasets, which include example instructions paired with corresponding input-output pairs[18]. To optimize resource usage, efficient fine-tuning methods focus on adjusting a small subset of parameters instead of the entire model, as fine-tuning all parameters in large language models (LLMs) is computationally expensive[19]. We utilize LoReft, a representation fine-tuning (ReFT) technique for our approach. This method keeps the base model frozen while learning task-specific modifications to hidden representations[20].

3 The Proposed framework

Our proposed framework, illustrated in Figure 2, aims to enhance safety measures on construction sites by ensuring rapid, reliable, and effective communication during emergencies. The system is designed to detect falls in construction areas and promptly alert the emergency response team and construction authorities. To achieve this, the framework incorporates camera feeds for real-time monitoring, wearable devices to assess worker conditions, and a Small Language Model (SLM)-based agent for continuous automated communication. By integrating a fall detection model with the SLM agent, the system ensures timely and efficient responses to emergencies, significantly improving overall safety protocols. This proposed framework offers several benefits that are crucial for the safety of construction sites.

Firstly, the framework provides rapid detection of falls and facilitates prompt communication with emergency care teams and construction management, significantly reducing response time and improving safety outcomes.

Secondly, it is designed with lightweight implementation, which makes it suitable for resource-constrained environments commonly found in construction sites by optimizing computational and storage requirements.

Another important feature is the system's continuous communication support, which provides uninterrupted interactions with various teams and parties.

Lastly, the system's architecture is build for scalability, allowing it to adapt to diverse construction scenarios and cater to varying site requirements and conditions, which makes it versatile solution for enhancing workplace safety.

The process begins with the preparation of two datasets: a fall dataset, featuring images of various fall scenarios, to train a fall detection model, and a conversation dataset, comprising transcripts of emergency communications (e.g., 911 calls), which is used to fine-tune the SLM-based agent for seamless and efficient dialogue handling. The system includes a computing unit with robust storage and communication capabilities and connects with wearable devices worn by workers for continuous monitoring. The workflow starts with an initial setup where key details, such as the worker's identity, property infor-

mation, and GPS location, are registered. A camera feed monitors the worker's movements, and in the event of a fall, the fall detection model generates an alert containing preliminary information, including the worker's location and other critical details. This alert triggers the SLM-based agent, which initiates communication with the emergency response team and construction authority. The agent facilitates real-time interactions, confirming essential details such as the worker's location, GPS coordinates, and floor level. The dialogue showcased in the figure highlights the agent's capability to respond instantaneously, ensuring that emergency teams receive accurate and timely information. Through this proposed system, continuous communication occurs after falls, significantly improving safety management at construction sites.



Figure 3. Fall detection analysis based on video data

4 Proof of Concept Evaluation and Results

In a simulated construction environment featuring a fall detection camera, our evaluation focuses on the performance of the Small Language Model (SLM) during its interactions with the emergency response team. The primary goal is to assess the SLM's ability to process critical information, such as the worker's location, status, and safety during emergencies. Additionally, we analyze the confidence score of the fall detection module. This focus is crucial, as inaccurate or delayed responses could compromise the worker's safety. In this current simulation implementation, we do not consider sensor-based data; only real-time video data processing is taken into account. The evaluation metrics for this experiment are as follows.

Accuracy: The ratio of correct answers provided by the SLM to critical information queries.

Response Time: The time required for the SLM to generate responses to the queries.

Relevance: The extent to which the SLM's responses are contextually appropriate and aligned with the questions posed.

4.1 Results and Analysis

Fall Detection Module: The Fall Detection module utilizes the YOLOv8 model to identify individuals and utilize the UR-Fall-Detection dataset [21] to ensure precise fall detection. The detection mechanism involves calculating a threshold value to determine if a fall has occurred. When the threshold value drops below zero, the system identifies it as a fall event and generates an alert, which is subsequently sent to the SLM for appropriate notification. Figure3 depicts the fall detection using a test video related to construction sites.

Small Language Models: Initially, we attempted to use the Tiny-Llama and Phi-3 models without fine-tuning, but they generated completely incorrect responses, as shown in Fig. 6(a) and Fig. 6(c). To improve performance, we fine-tuned these models using syntactic data generated by ChatGPT. Fine-tuning Tiny-Llama with fifty samples in the specified environment took approximately 56 minutes and 48 seconds, while fine-tuning Phi-3 required around 5 hours and 30 minutes. The loss curves for the fine-tuned models are depicted in Fig. 4, with Tiny-Llama on the left and Phi-3 on the right. Both models exhibit a gradual decrease in loss during training.

Next, we used the fine-tuned Tiny-Llama to generate an initial emergency alert, which successfully followed this format: *"Emergency Alert: A fall has been detected. Here are the critical details: Person's Type:Worker , Age: 40, Location: 914 Lake Avenue, Baltimore, MD, 21202, GPS Coordinates: 38.9796° N, 76.6362° W, Property Type: House, Floor Level: 3, Time of Incident: July 2, 2024, 12:50, Status of the Person: Unresponsive"*

Upon reviewing the generated alert, however, we observed inaccuracies in subsequent responses, as shown in Fig. 5(left). These discrepancies occurred because neither Tiny-Llama nor Phi-3 could recall the initial alert accurately, leading to inconsistencies in details like location and floor level, which differed from the original alert.

To resolve this issue, we stored the initial alert data in a database. When the fall detection module generated an alert, the stored information was retrieved and passed to the small language model (SLM). The SLM used this data to generate a consistent fall alert and provided detailed responses tailored for the emergency response team, as illustrated in Fig. 5(right).

We evaluated the performance of the small language models with and without fine-tuning using human evaluation metrics. The metrics assessed the relevance of the responses to specific questions and the time taken to generate them. Relevant answers were scored as 1, while irrelevant answers were scored as 0. We analyzed Tiny-Llama and Phi-3's performance under these metrics, both

with and without fine-tuning. Additionally, we evaluated the accuracy of critical information, such as location details, GPS coordinates, and person type, when utilizing the local database with the fine-tuned Tiny-Llama and Phi-3 models.

Accuracy: Responses generated by small language models (SLMs) without using a local database, as shown in Fig. 5(left), highlight their inability to retain critical information. This example uses TinyLlama, although Phi-3 produces similar results. In these scenarios, the SLM-Bot generates a fall alert but fails to recall essential details. Conversely, when critical information is stored in a local database, as illustrated in Fig. 5(right), the model consistently generates accurate responses. Ultimately, accuracy without a local database is zero, while leveraging the database ensures 100 percent accuracy.

Answer Relevance: Fig. 6(a) displays TinyLlama's responses without fine-tuning. Initially, the SLM-Bot issues a fall alert, followed by inquiries from the emergency response team. Upon evaluation, we find that while the responses are generally relevant, only one out of four answers is pertinent to the question, "How is the person's blood sugar?" After fine-tuning TinyLlama, Fig. 6(b) reveals a notable improvement, with SLM-Bot providing consistently relevant answers to all questions. Similarly, Fig. 6(c) shows Phi-3's performance without fine-tuning, where only two responses are relevant. However, as seen in Fig. 6(d), the fine-tuned Phi-3 consistently delivers relevant answers. These results indicate that fine-tuning enhances the ability of models to generate accurate and relevant answers for emergency scenarios.

Response Time: When analyzing response times, Phi-3 consistently outperforms TinyLlama, generating answers faster in both fine-tuned and non-fine-tuned states, as shown in Fig. 6. Additionally, fine-tuned versions of TinyLlama (Fig. 6(b)) and Phi-3 (Fig. 6(d)) exhibit slightly longer response times compared to their non-fine-tuned counterparts (Fig. 6(a) and Fig. 6(c)). Despite the slight increase in response time, the fine-tuned models provide more accurate and relevant answers, underscoring their overall effectiveness.

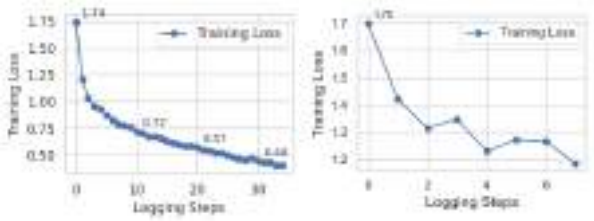


Figure 4. Loss curve in fine-tuning (left) TinyLlama and (right) Phi-3

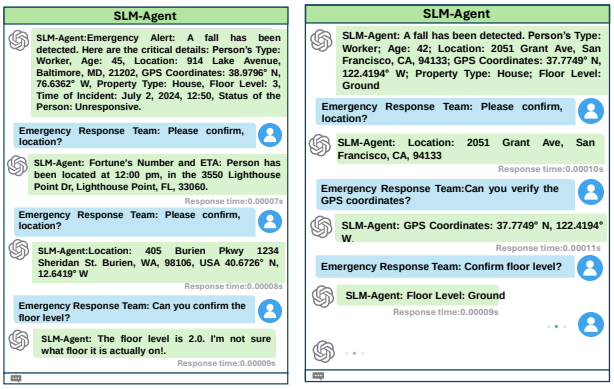


Figure 5. Responses from Small Language Models (SLM): (left) TinyLlama without local-database, (right) TinyLlama with local-database

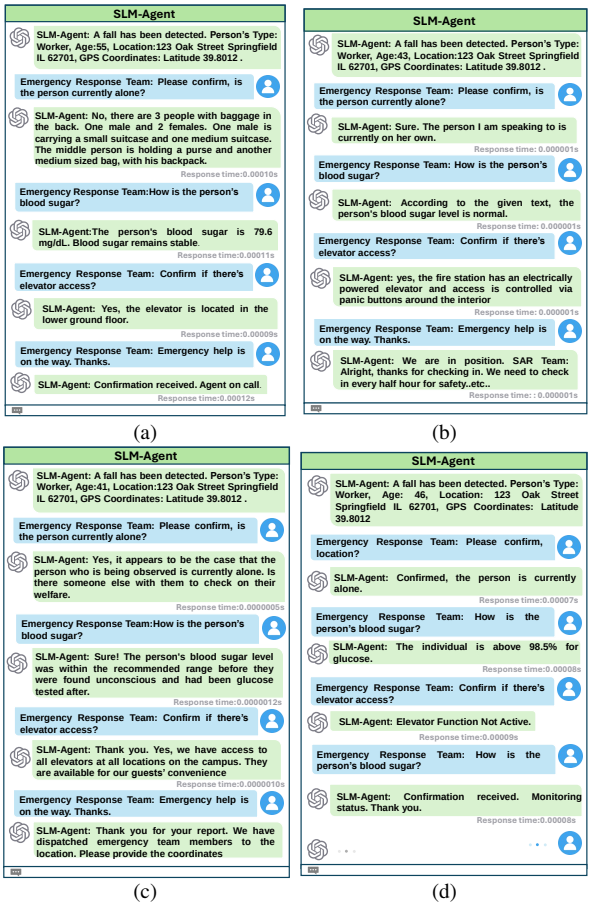


Figure 6. Responses from Small Language Models (SLM): (a) TinyLlama without fine-tuning, (b) TinyLlama with fine-tuning, (c) Phi-3 without fine-tuning, and (d) Phi-3 with fine-tuning.

5 Conclusion

This paper presents a novel fall detection system designed for construction sites, integrating small language models (SLM) to improve real-time communication and provide improved support for construction safety. By combining SLMs with lightweight machine learning models, we addressed key limitations of traditional fall detection systems, such as insufficient communication capabilities and limited assistance for critical construction sites. Through simulations using TinyLlama and Phi-3 models, we demonstrate how SLMs enhance accuracy, response times, and the relevance of responses in fall detection scenarios.

While the results are encouraging, there are several opportunities for future development to further refine the system. In particular, we are focusing on the following directions as part of our future work. We plan to integrate wearable sensors data, as outlined in the proposed framework, in the implementation to study context awareness and detection accuracy. Furthermore, we intend to expand the training dataset with diverse and synthetic to increase the model's robustness and generalizability. Finally, comparing the performance of this system with larger language models could offer valuable insights into balancing model size, accuracy, and computational efficiency.

References

[1] Sarfraz Iqbal. Artificial intelligence tools and applications for elderly healthcare - review. In *Proceedings of the 2023 9th International Conference on Computing and Artificial Intelligence*, ICCAI '23, page 394–397, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9781450399029. doi:10.1145/3594315.3594347. URL https://doi.org/10.1145/3594315.3594347.

[2] Malithi Wanniarachchi Kankanamge, Syed Mhamudul Hasan, Abdur R. Shahid, and Ning Yang. Large language model integrated healthcare cyber-physical systems architecture, 2024. URL https://arxiv.org/abs/2407.18407.

[3] U.S. Bureau of Labor Statistics. Construction deaths due to falls, slips, and trips increased 5.9 percent in 2021. On-line: https://www.bls.gov/opub/ted/2023/, Accessed: 01/11/2025.

[4] Xiaoyu Liu, Feng Xu, Zhipeng Zhang, and Kaiyu Sun. Fall-portent detection for construction sites based on computer vision and machine learning. *Engineering, Construction and Architectural Management*, ahead-of-print(ahead-of-print), Jan 2023. ISSN 0969-9988. doi:10.1108/ECAM-05-2023-0458. URL https://doi.org/10.1108/ECAM-05-2023-0458.

[5] Peiyuan Zhang, Guangtao Zeng, Tianduo Wang, and Wei Lu. Tinyllama: An open-source small language model. *arXiv preprint arXiv:2401.02385*, 2024.

[6] Marah Abdin, Sam Ade Jacobs, Ammar Ahmad Awan, Jyoti Aneja, Ahmed Awadallah, Hany Awadalla, Nguyen Bach, Amit Bahree, Arash Bakhtiari, Harkirat Behl, et al. Phi-3 technical report: A highly capable language model locally on your phone. *arXiv preprint arXiv:2404.14219*, 2024.

[7] Muhammad Khan, Rabia Khalid, Sharjeel Anjum, Si Van-Tien Tran, and Chansik Park. Fall prevention from scaffolding using computer vision and iot-based monitoring. *Journal of Construction Engineering and Management*, 148(7):04022051, 2022.

[8] Numan Khan, Sharjeel Anjum, Rabia Khalid, Jun-Sung Park, and Chansik Park. A deep learning-based predication of fall portents for lone construction worker. In *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, volume 38, pages 419–426. IAARC Publications, 2021.

[9] Synaedge. Synaedge - multifunctional ai video surveillance. On-line: https://synaedge.com/en/, Accessed: 01/11/2025.

[10] viact. viact - we automate monitoring. On-line: https://www.viact.ai/, Accessed: 01/11/2025.

[11] slatesafety. slatesafety - the 1 heat stress prevention wearable. On-line: https://slatesafety.com/, Accessed: 01/11/2025.

[12] kwant. kwant - end-to-end construction workforce management platform. On-line: https://www.kwant.ai/, Accessed: 01/11/2025.

[13] Alwaysai. Computer vision solutions for enterprises. On-line: https://alwaysai.co/, Accessed: 01/11/2025.

[14] Visionify. Visionify - nextgen workplace safety for manufacturing. On-line: https://visionify.ai/, Accessed: 01/11/2025.

[15] Intellisee. Intellisee - ai-powered risk detection surveillance solutions. On-line: https://intellisee.com/, Accessed: 01/11/2025.

[16] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

[17] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.

[18] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, et al. A survey of large language models. *arXiv preprint arXiv:2303.18223*, 2023.

[19] Zeyu Han, Chao Gao, Jinyang Liu, Sai Qian Zhang, et al. Parameter-efficient fine-tuning for large models: A comprehensive survey. *arXiv preprint arXiv:2403.14608*, 2024.

[20] Zhengxuan Wu, Aryaman Arora, Zheng Wang, Atticus Geiger, Dan Jurafsky, Christopher D Manning, and Christopher Potts. Reft: Representation finetuning for language models. *arXiv preprint arXiv:2404.03592*, 2024.

[21] Bogdan Kwolek and Michal Kepski. Human fall detection on embedded platform using depth maps and wireless accelerometer. *Computer methods and programs in biomedicine*, 117(3):489–501, 2014.

SVR and GA Aided Lean Six Sigma Method for Planning in Modular Construction

Angat Bhatia[1], Osama Moselhi [1,2] and SangHyeok Han[1,2]

[1]Department of Building, Civil and Environmental Engineering, Concordia University, Montreal, QC, Canada
[2]Centre for Innovation in Construction and Infrastructure Engineering and Management (CICIEM), Concordia University, Montreal, QC, Canada
angatpalsingh.bhatia@concordia.ca, moselhi@encs.concordia.ca, sanghyeok.han@concordia.ca

Abstract –

Modular construction presents a strong alternative to traditional construction, offering advantages such as improved productivity, and better quality. However, the prefabrication of module components follows a make-to-order process, resulting in customized module components. This design customization, along with various factors such as worker skill levels, and defects in shop drawings causes significant variability in the process times for prefabricating module components at workstations. This variability leads to imbalanced production line, and idle time at workstations, which increases the overall completion time of fabricating module components. To address these challenges, this paper develops a Lean Six Sigma based method that comprises three modules. In the first module, the production line that requires improvements is identified and project objectives are defined. In the second module, the process time data of module components at workstations are collected to identify and analyse inefficiencies in the production line utilizing six sigma performance metrics. The third module focuses on improving and controlling the production line process using support vector regression (SVR) and meta-heuristic optimization. A light gauge steel (LGS) wall panel production line in Edmonton, Canada was analysed to demonstrate the use of the developed method and test its performance. The results show that, after addressing the production line bottlenecks, the sigma level improves to 1.85 σ compared to 1.41 σ earlier. This method can help production managers identify wastes and bottlenecks in the production line, enabling them to plan their processes more efficiently.

Keywords –

Modular Construction; Lean; Six Sigma; Support Vector Regression; Optimization

1 Introduction

For many years, construction professionals have drawn to the idea of making construction practices after those in manufacturing, as this approach leads to substantial improvements in productivity and quality [1]. In this regards, modular construction has been applied to a wide range of projects, from large-scale industrial buildings to residential housing. The modular construction process is based on two strategies: (i) the product approach, which seeks to minimize on-site construction by turning buildings into unique products (i.e., module components) that can be prefabricated in a factory environment; and (ii) the process approach, which applies a production line principle to the prefabrication process [2]. As a result, companies that have adopted prefabrication have seen significant productivity gains within a few years. Despite the advantages highlighted above, the high degree of design customization in module components (i.e., variation in design specifications of module components), often result in each module being a unique product. For example, on the same production line not all produced wall panels are identical. Wall Panels are rather customized to accommodate dynamically changing owners' requirements and is identified as a key factor preventing modular construction from realizing its full potential benefits. This design customization results in differing process times for module components at each workstation, in turn leading to different production rates, imbalanced production lines, and an increased total completion time (i.e., makespan) [3-4]. To reduce completion time, it is crucial to identify sources of waste (e.g., overproduction and waiting) within a production line so that improvement efforts can be directed towards these areas. Lean manufacturing principles, originally developed in the manufacturing industry, have proven to be applicable to identify waste in modular construction

[2]. Although lean focuses on eliminating waste in production, it does not specifically address process variation such as the time taken to prefabricate wall panels in the production line and employ statistical analysis [5]. Meanwhile, Six Sigma can identify/reduce variations, eliminate waste and is a supplement to lean manufacturing principles. Moreover, Six Sigma DMAIC method provides a standardized and step-by-step approach unlike lean for: (i) comprehensive analysis to identify root causes to the production line waste, (ii) reducing process variations and (iii) providing solutions to achieve process improvements (i.e., minimizing completion time) [6-7].

In this respect, the primary goal of this study is to minimize the makespan by identifying and eliminating production line waste through integration of Six Sigma's DMAIC approach with machine learning and optimization techniques. The developed method encompasses: (i) defining the production line that requires improvements; (ii) identifying factors that causes variations in process time at workstations and developing a Six Sigma production line performance (PL_P) metric to assess performance; and (iii) developing a prediction model using support vector regression (SVR) and implementing GA for hyperparameter tuning. Additionally, the 5 Whys technique to reach to the root cause of the issues and simulated annealing optimization for determining optimal sequences of module components are employed to eliminate waste and minimizing makespan.

2 Literature Review

2.1 Application of Six Sigma

The Six Sigma principles are applied using a problem-solving approach like DMAIC (Define, Measure, Analyze, Improve, Control). The DMAIC method aligns with strategy for addressing problems by recognizing the root causes of variations in the production line and utilizing statistical methods to maintain continuous improvements [8]. In the construction industry, the application of Six Sigma principles for performance evaluation, especially for controlling variability, began in 2000. Buggie [9] introduced Six Sigma as a method to enhance productivity by focusing on reducing cycle time and eliminating defects in processes. Meanwhile, Han et al. [5] utilized a Six Sigma-based approach to enhance performance in construction operations by establishing quantitative metrics. Despite the numerous benefits of the Six Sigma approach, its application in modular construction has been limited. To the authors' knowledge, there have been almost no project-specific case studies in modular construction aimed at enhancing production performance while achieving appropriate sigma levels. Additionally, previous studies have not identified the list of possible factors that could impact the process time for prefabricating module components at workstations.

2.2 Data Analytics

The highly customized nature of module components, along with factors such as worker skills, on-site change orders, and defects during prefabrication, leads to uneven distribution of process times across workstations. This unevenness results in various forms of waste (e.g., overproduction and waiting) and increases the total completion time for prefabricating module components [10]. Therefore, it is crucial to accurately estimate process times for production line planning considering the design variations and other factors. In this context, Six Sigma's DMAIC method has been used to remove bottlenecks and enhance the production line process by employing statistical techniques such as t-test and linear regression. However, these techniques are not effective when dealing with large, complex and non-linear data [4]. As a result, researchers have recommended combining Six Sigma with machine learning techniques for accurate predictions of process times, enabling production line managers to make better decisions [10].

Support Vector Regression (SVR), a machine learning algorithm that adapts the principles of Support Vector Machines (SVM), has been studied by various researchers for regression tasks with non-linear datasets [11]. For instance, Tsarouhas et al. [12] combined the Six Sigma method with machine learning techniques to predict outcomes for future injection moulding processes. On the other hand, Wauters and Vanhoucke [12] used SVR in a project control setting to develop more accurate time and cost forecasts by utilizing periodic data from Earned Value Management. However, these methods have overlooked the hyperparameter tuning. Several researchers have employed genetic algorithms (GA) for hyperparameter tuning in the fields of construction and manufacturing engineering [13-14] to improve the accuracy of their prediction model. Given the advantages of GA and its effectiveness for fine tuning parameters, this paper applies GA to identify the optimal parameters for an SVR algorithm in order to efficiently predict process times of module components in the production line.

3 Developed Method

Fig. 1 shows the Six Sigma-based method for enhancing production line processes in modular construction. This method emphasizes the use of DMAIC (Define, Measure, Analyse, Improve, and Control), a core component of Six Sigma focusing on continuous

improvement. This method integrates machine learning (Support Vector Regression) and optimization (Genetic Algorithm and Simulated Annealing) within DMAIC process to identify production line wastes and enhance the production line performance.

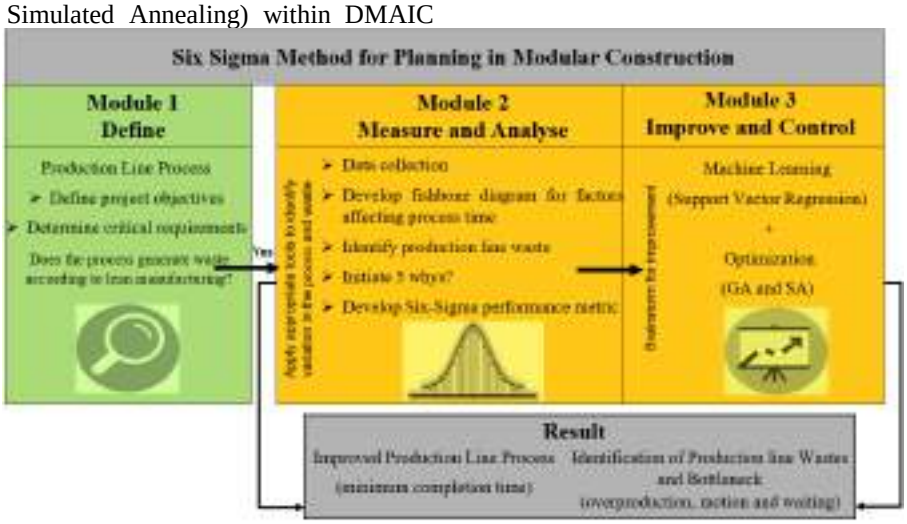


Fig 1: Main components of Six Sigma DMAIC method

The developed method includes: (i) define module, which focuses on identifying the problem that requires improvement and setting the project objectives. In this paper, the process selected is the wall panel production line, with the aim of reducing the total completion time for prefabricating module components; (ii) measure and analyse module to identify production line inefficiencies (waste) and determine their root causes. The process starts with collecting the process time data for module components at each workstation using the C-track app [19]. Statistical analysis is performed to assess the extent of variations and deviations in the process times for module components at workstations. Additionally, a fishbone diagram is created to list the factors affecting the process times and pinpoint potential causes of these variations. Analyzing these root causes and their impacts is crucial for reducing completion times. This module also identifies potential wastes (such as overproduction, waiting, and motion) and bottlenecks within the production line. Along with fishbone diagram, the '5 Whys' technique is used to uncover issues and solutions within the production line such as workstations, that takes more time to prefabricate module components compared to other workstations. By continuously asking 'why' for each identified issue, this approach helps to pinpoint the root cause of the problem, ensuring that solutions address the fundamental issues. To assess the sigma level and measure performance, this module includes a Six Sigma Production Line Performance (PL_P) metric (Eq. (1) and (2)). This PL_P metric is modified from the process capability index metrics commonly found in Six Sigma literature. Production line performance is expressed by following equation:

$$PL_P = \frac{(USL-CT)}{(3\times STDEV)} \quad (1)$$

$$\sigma = 3 \times PL_P \quad (2)$$

where USL= upper specification limit; CT= completion time; and STDEV= standard deviation of the process time data.

For instance, if the completion time and standard deviation are 50 and 10 minutes, respectively, and the USL for the completion time is set at 55 minutes, the current PL_P is calculated to be 0.16, with a sigma level of 0.5 σ. This indicates a low performance level, suggesting that improvements are needed in the production line. (iii) the third module, which is improve and control, address the issue of variability in process time at workstations to enhance the production line process. To achieve this, it is crucial to predict process times based on design factors and the number of workers at each workstation, which significantly influence the process. This research utilizes SVR to develop prediction models for workstations in the production line. The data is pre-processed and split into training (80%) and testing (20%) subsets. Additionally, to avoid overfitting, K-fold cross-validation is employed, to assess model performance. Initially, default parameter values, including C (regularization), gamma, and the kernel, are used in the SVR model. Subsequently, the GA is applied for hyperparameter tuning (C and Gamma) with the goal of reducing the mean absolute error (MAE). Given the successful implementation of GA in the literature for tuning the parameters of machine-learning models and avoiding local optima [4], the present study implements GA to determine the optimal parameters for an SVR-based process time prediction model. It is worth mentioning that in the SVR model, C and gamma play crucial roles with regards to controlling the error rate and model complexity; therefore, tuning the C and gamma is crucial to achieve optimal predictive performance in SVR. Fig. 2 illustrates the GA-based hyperparameter tuning process, which begins by creating an initial

population of chromosomes, each representing different configurations of C and gamma. This population is generated randomly, with the size of the population determining the number of chromosomes. The structure of the generated chromosome (Fig. 2), represents a potential solution. For example, in the chromosome "0,1,0,0,0,1,1,0,1,0" the first five digits encode the value for C, while the remaining digits represent gamma.

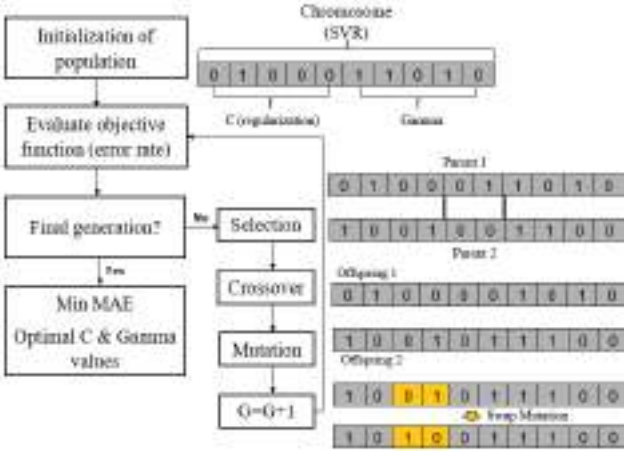


Fig 2. GA based hyperparameter tuning

These binary strings are then decoded using Eq. (3) to convert the encoded parameters back into their real-world values.

$$D = (\Sigma bit \times 2^{i}) \times P + \alpha \quad (3)$$

$$P = \frac{b-a}{2^{l}-1} \quad (4)$$

Where D denotes the decoding process, and P represents the parameter (as indicated in Eq. (4)). The variables a and b specify the lower and upper bounds of the parameter (e.g., gamma), while l refers to the length of the chromosome. For instance, to decode the variable Gamma, which has lower (a) and upper (b) bounds of 0.05 and 0.99, respectively, and a chromosome length of 5 (Fig. 3), the decoded value (D) is 0.81, which falls within the acceptable range (i.e., 0.05 to 0.99).

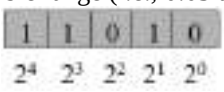


Fig 3. Decoding of chromosome (Gamma parameter)

Later, the process of evaluation, where fitness values of each chromosome are evaluated based on the optimization criteria (i.e., minimum error rate); Third, the selection process involves choosing the best chromosomes (i.e., potential parents) from the population as the best solution (i.e., the best combination of Gamma and C) based on their fitness value. This paper uses tournament selection due to its efficiency, ease of implementation, and relatively low computational time compared to other methods (e.g., roulette wheel selection) [15]. In the tournament selection technique, a number of chromosomes (represented as k, i.e., the tournament size) are randomly selected from the population, and the chromosome with the best fitness score among the k chromosomes is selected as parent 1. This process is repeated until the desired number of parents for crossover is selected. For example, in a population of 10 with a tournament size (k) of 3, chromosomes 1, 2, and 3, with fitness values of 0.75 min, 0.82 min, and 0.68 min, respectively, are randomly picked. The chromosome with the best fitness (i.e., minimum error rate)—in this case, chromosome 3 (fitness 0.68)—is selected as the first parent, and the process continues with the identification of a second parent. In the next step, finally, after selecting two parents, crossover is performed using two-point crossover to create new chromosomes for the next generation. Mutation is subsequently performed, and this process (i.e., the sequence of selection, crossover, and mutation) continues, until the termination criteria (i.e., maximum number of generations) are met.

The predictive process times generated by the SVR-based model are used as input for SA optimization. This optimization process focuses on solving the bottleneck, in the production line by optimizing the sequence of module components in with the goal of minimizing the completion time (as indicated in Eq. (5)) and achieve the desired sigma level. Although, the effectiveness of optimization algorithms varies depending on the scheduling problem and objective, the positive reviews of SA algorithms, known for providing effective solutions and good computing capabilities in solving production line scheduling problems, justifies their selection to optimize the sequences in this study [4].

$$Min\ (Max\ C_{m,w};\ \forall\ m \in M, w \in W) \quad (5)$$

Where $C_{m,w}$ is the completion time of module component at workstation. M = {1, 2…… m} and W = {1, 2…… w} are set of 'm' number of module components to be scheduled at a set of 'w' workstations.

4 Case Study and Results

The developed method was applied to a wall panel production line consists of three primary workstations: (i) assembly, (ii) framing, and (iii) sheathing (Fig. 4). The fabricator produces a range of LGS wall panels for multi-storey residential buildings. Process times for each workstation, comprising 200 data series, were recorded using the C-track app.

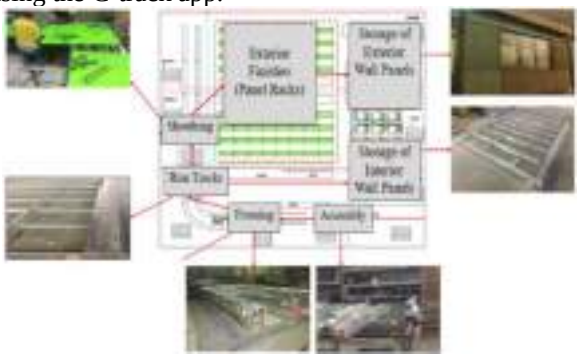


Fig. 4. Production line layout

The statistical details, including the mean and standard deviation, for each workstation on the production line were calculated. For instance, the average process time at the assembly workstation is 37.86 minutes, with a standard deviation of 23.47 minutes. A high standard deviation relative to the mean suggests high variation in the process times, with data points dispersed over a broader range. Additionally, at the assembly station, wall panel processing times range from 7 to 102 minutes. This considerable variation is attributed to factors such as design specifications (e.g., number of windows and studs), and the number of workers at each workstation, which impact the daily productivity and overall completion time of the workstations. Therefore, to assess performance, the Six-Sigma production line performance (PL_P) metric was utilized to calculate the sigma level. The current production line has a sigma level of 1.41 σ, which indicates that the high variation in the process time of prefabricating module components at workstations is causing waste in the production line and leading to low sigma level. In this respect, a fishbone diagram (Fig. 5) was created to illustrate the various factors contributing to process time variation at workstations. The high variation in process times was attributed to factors such as customized panel designs, worker-related issues, on-site conditions, and inventory management. The diagram further details specific causes within these factors, such as suboptimal worker allocation, inadequate skills or experience, learning curves, and space congestion around workstations, which affects worker coordination. This space congestion was also reported by Zhang et al. [16], which disrupts workflow and reduces productivity at the workstation level.

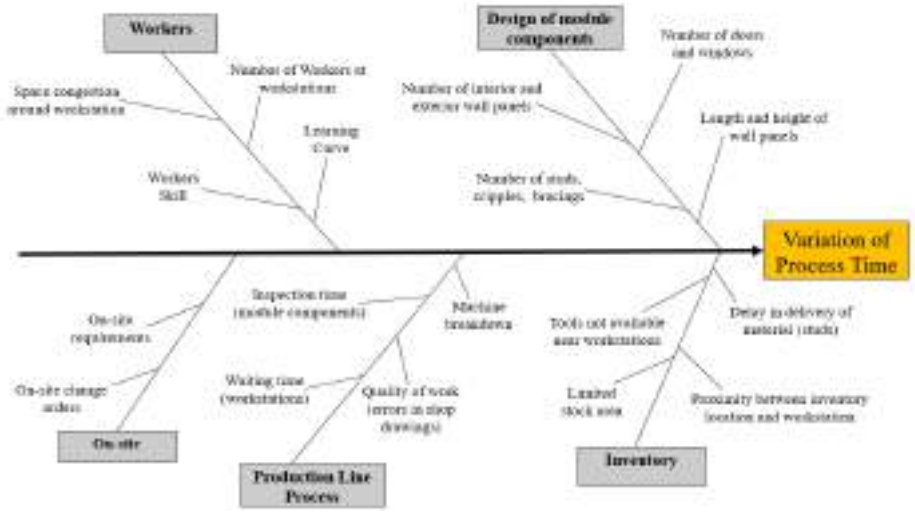


Fig. 5: Fishbone diagram for factors affecting process time

Based on the identified factors affecting process time and observations over several months, various lean manufacturing waste were detected, including overproduction, waiting, motion, and overprocessing. For example: (i) workers at the assembly workstation had to frequently move to retrieve materials (e.g., studs and bracings), which increased motion waste and prolonged process times; (ii) the workload was uneven, with the framing station having longer process times compared to the other workstations, leading to an imbalanced production line, waiting time of workers at the sheathing station, and overproduction at the assembly station; (iii) the cutting machine was located far from the framing workstation, causing workers to frequently move to cut bracings for wall panels, resulting in motion waste; and (iv) quality issues, such as errors in shop drawings (e.g., incorrect stud spacing or missing components), causing defects that required corrections at downstream stations, contributing to increased process times. In this respect, this study utilized the 5 Whys iterative questioning technique to identify solutions to some of the problems. By repeatedly asking 'why' a problem occurred, the root causes were determined, and potential solutions were documented. Table 1 provides an example related to the waste of motion observed at the assembly workstation. The first question posed was, why are some wall panels taking so long in the assembly process? The documented reason was, because workers take time to find specific materials (e.g., header and sill track, studs). The follow-up question was, why do workers take time to find specific materials? The reason noted was, because materials were not arranged properly in their designated locations. This questioning process continued through five iterations, ultimately leading to the proposed solution of 5S implementation, efficient layout design, and standardization. To further reduce production line waste (e.g., waiting and overproduction) and enhance performance by reducing completion time, it was crucial to accurately predict process times. To achieve this, design factors (e.g., the number of studs, cripples) and the

number of workers at each workstation were used as inputs in SVR algorithm to forecast process times. Before developing the prediction model, data preprocessing steps were conducted, including removing outliers, normalizing the data, and identifying significant design factors that impact process time. In this study, correlation analysis was employed to identify significant design factors. Factors were selected if their correlation coefficient, indicating the strength of the relationship between the dependent variable (duration) and independent variables, was greater than 0.65 (Fig. 6). For instance, at the assembly workstation, four factors (the number of studs, studs with foam, cripples, and plates) had correlation coefficients exceeding 0.65.

Table 1: 5 whys example

Why 1?	Why 1 Reason	Why 2?	Why 2 Reason	Why 3?	Why 3 Reason	Why 4?	Why 4 Reason	Why 5?	Why 5 Reason
Why the assembly process taking so long for some of the wall panels?	Because workers spend time searching for specific materials (e.g., sill track, studs).	Why workers take time to search a specific material?	Because the materials are not properly organized in their designated locations.	Why material is not arranged properly and placed at specific location?	Because there are no designated storage areas for some of the material	Why there are no designated storage areas for some of the material	Because of poor plant layout	Why the storage and plant layout are poor?	Because there is no 5S implementation and standardized workflow

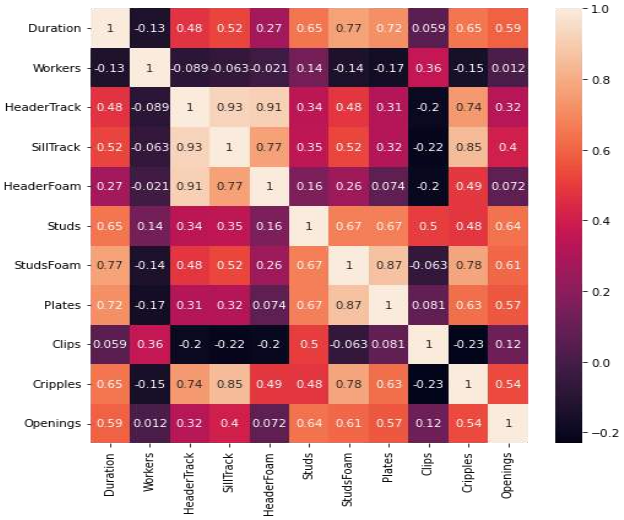


Fig.6: Correlation heatmap for assembly station

SVR models were developed on training dataset (80%) utilizing the significant design factors. As a result, shown in Table 2, MAE for the assembly station was 5.25 minutes. It is important to note that the initial SVR model was created using default parameters (i.e., C=1, Gamma (γ)=scale).

Table 2: Mean absolute error measurement for SVR

Workstation	C	γ	SD (min)	MAE (min)	MAPE (%)
Assembly	1	Scale	23.4	5.25	27.8
Framing	1	Scale	22.1	5.29	23.3
Sheathing	1	Scale	24.3	12.22	30.2

To further reduce the MAE in the prediction model, hyperparameter tuning of C (regularization) and Gamma was performed using a GA. The parameter C was selected because it controls the penalty for errors, while Gamma is crucial as it ensures the model captures the data's underlying structure accurately without being overly sensitive to noise. For SVR, there are various types of kernel functions such as linear, polynomial, RBF and sigmoid. However, it should be noted that in this research only linear and RBF were tested and compared due to their (i) good generalization capabilities; (ii) better computation and (iii) capability to avoid overfitting in nonlinear data set [17]. Table 3 shows the results, indicating a reduced MAE and MAPE after hyperparameter tuning. For instance, at the assembly workstation, the MAE was reduced to 1.78 minutes. The GA parameters include: (i) a population size of 40; (ii) a maximum of 20 generations; (iii) a mutation probability of 0.3; (iv) a crossover probability of 1; (v) a tournament size of 3; (vi) C ranging from 10 to 1000; and (vii) Gamma ranging from 0.05 to 0.99. The GA parameter values are determined by experimenting with various

values and selecting the one yielding the best results. Additionally, it is worth noting that, C range (10–1000) allows control over model complexity, avoiding extreme underfitting or overfitting; and Gamma range (0.05–0.99) ensures non-linearity is captured effectively. These ranges cover a broad spectrum of hyperparameter values for SVR in order to achieve near-optimal solution in a reasonable run time (i.e., explore broad search space, while achieving computational efficiency).

Table 3: MAE values for SVR after parameter tuning

Workstation	C	γ	Kernel	MAE (min)	MAPE (%)
Assembly	977.5	0.9	Linear	1.7	13.3
Framing	110.5	0.5	RBF	5.1	21.2
Sheathing	13.8	0.2	RBF	11.1	26.4

As previously mentioned, the framing workstation acts as the bottleneck in the production line. To address this, SA optimization was used to reflect the practical scenario when determining the optimal sequence of panels by considering an additional framing station to the production line to reduce completion time (i.e., improving sigma level), allowing two wall panels to be prefabricated concurrently. To reflect this practical case of capturing parallel workstations in the optimization algorithms, Eq. (6, 7 and 8) outlines the constraint concerning workstation capacity (X) and start time of wall panel. For example, Eq. (7) outlines the constraint, which stipulates that the start time of a wall panel (i.e., panel 'm') at parallel workstations must be equal to the completion time of the same wall panel at previous workstation. This can reduce the waiting times of wall panels between workstations, and allows the parallel workstation to work on two wall panels simultaneously. The detailed discussion regarding SA implementation for optimization is detailed in Bhatia [4].

$$\text{Max } X_2 \leq 2; \forall S_{m,2} \quad (6)$$

$$S_{m,\,2} = C_{m,\,1} \quad (7)$$

$$S_{m,\,2} \geq \min (C_{m\text{-}1,\,2}; C_{m\text{-}2,\,2}) \quad (8)$$

Figure 6 shows the completion time results for 30 wall panels: with one framing station (7 a), the completion time was 1069.59 minutes, which was reduced to 1057.2 minutes with a parallel framing workstation (7 b). In other words, optimizing the sequencing of wall panels and addressing bottlenecks at the framing workstation improved the sigma level (i.e., 1.85 σ with two framing stations compared to 1.41 σ with one framing station). It is worth noting that the factors affecting the process time in this study are a function of the given module component (wall panel) design specifications and tasks performed at the workstations of the case study production line. For other modular construction cases, practitioners would need to modify the data analysis phase based on the given design specifications in order to identify the factors and developing process times prediction models based on those inputs. Additionally, the practitioners need to identify the bottlenecks and sources of waste according to their given production line to improve the sigma level.

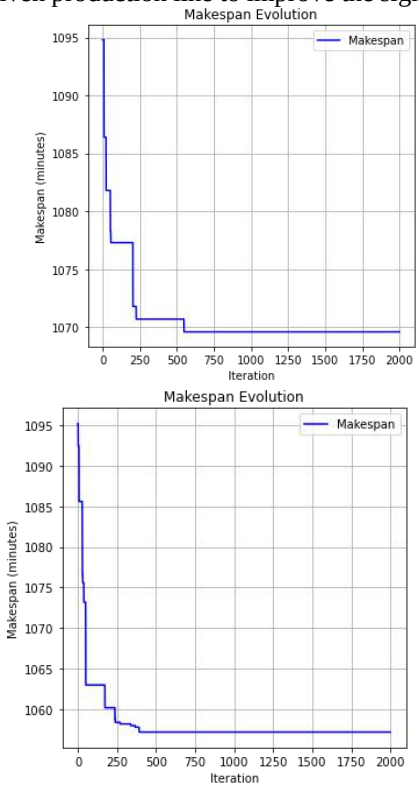


Fig 7: completion time: (a) one framing workstation and (b) two framing workstations

Conclusion

In a modular construction production line, process time is influenced by several factors, including the customized design of wall panels and worker-related variables. These factors cause a high level of variation in process time, resulting in production line waste, such as unbalanced production line, overproduction, and idle time at some workstations. This increases total completion time for prefabricating modules as demonstrated through the case study. Therefore, to ultimately minimizing makespan, it is significant to identify the list of factors impacting the process time for prefabricating module components at workstations, and utilize them to identify and eliminate production line waste. Moreover, application of six-sigma to the project-specific case studies in modular construction is required in order to enhance production performance while achieving appropriate sigma levels. In this regards this paper introduces a novel method that incorporates historical process time data from a wall panel production factory into the Six Sigma DMAIC framework. First, process time prediction models were developed where design factors and the number of workers were used as inputs into SVR model. Subsequently, the GA

optimization technique was applied for hyperparameter tuning, specifically to determine the optimal values for C and Gamma. The hyperparameter tuning in the wall panel production line showed a reduction in the MAE to 1.78, 5.10, and 11.15 minutes compared to 5.25, 5.29, and 12.22 for the assembly, framing, and sheathing workstations, respectively. Additionally, this paper provides a solution to eliminate a bottleneck by considering an optimization scenario where an additional framing station allows two wall panels to be prefabricated concurrently. The optimal sequences of wall panels were generated using the SA algorithm, which reduced the total completion time and improved the sigma level to 1.85 σ. The contributions of this study are: (i) implementing GA to find the optimal hyperparameters for SVR; (ii) identifying the various factors (e.g., workers, inventory and design of module components) influencing process time at workstations, which leads to production line waste and low sigma level; and (iii) aiding production managers in gaining insights into the production line, thereby reducing the prefabrication time of module components at workstations.

However, this study is subject to some limitations. First, it does not explore means for eliminating other types of production line waste (e.g., motion, inventory and defects). Therefore, it is recommended for future work to collect data regarding these production line waste and applying lean principles along with simulation, where the former will be used for gaining a deeper understanding of the production line to reduce these production line waste, while the latter will be used for evaluating various production scenarios before implementing them in real production line. Second, future research could investigate other metaheuristic algorithms, such as particle swarm optimization (PSO) for hyperparameter tuning and compare their outcomes with those of GA. Third, the current optimization does not consider fine tuning different kernel functions. The future research would consider hyperparameter tuning of kernel functions in order to test their performance with regards to minimizing MAE.

References

[1] Koskela, L., 2000. An exploration towards a production theory and its application to construction. VTT Technical Research Centre of Finland.

[2] Yu, H., Al-Hussein, M., Al-Jibouri, S. and Telyas, A., 2013. Lean transformation in a modular building company: A case for implementation. Journal of management in engineering, 29(1), 103-111.

[3] Salama, T., Salah, A. and Moselhi, O., 2021. Integrating critical chain project management with last planner system for linear scheduling of modular construction. Construction innovation, 525-554.

[4] Bhatia, A.P.S, 2024. Automated Planning and Scheduling Method for Modular Construction Manufacturing. PhD thesis, Concordia University, Montreal, QC, Canada.

[5] Han, S.H., Chae, M.J., Im, K.S. and Ryu, H.D., 2008. Six sigma-based approach to improve performance in construction operations. Journal of management in Engineering, 24(1), 21-31.

[6] Stewart, R.A. and Spencer, C.A., 2006. Six-sigma as a strategy for process improvement on construction projects: a case study. Construction Management and Economics, 24(4), 339-348.

[7] McDermott, O., ODwyer, K., Noonan, J. Trubetskaya, A. and Rosa, A., 2023. The development of a lean six sigma and BIM framework for enhancing off-site manufacturing. International Journal of Lean Six Sigma, 15(8), 50-69.

[8] Harry, M. and Schroeder, R., 2006. Six sigma: the breakthrough management strategy revolutionizing the world's top corporations. Crown Currency.

[9] Buggie, F. D. 2000. “Beyond six sigma.” J. Manage. Eng., 164, 28– 31

[10] Mohsen O, Mohamed Y, and Al-Hussein M, 2022. A machine learning approach to predict production time using real-time RFID data in industrialized building construction. Advanced Engineering Informatics, 52, 101631.

[11] Wauters, M. and Vanhoucke, M., 2014. Support vector machine regression for project control forecasting. Automation in Construction, 92-106.

[12] Tsarouhas P, Sidiropoulou N. 2023. Application of Six Sigma methodology using DMAIC approach for a packaging olives production system: a case study. Int J Lean Six Sigma.

[13] Koc, K., Ekmekcioğlu, Ö. and Gurgun, A.P., 2021. Integrating feature engineering, genetic algorithm and tree-based machine learning methods to predict the post-accident disability status of construction workers. Automation in Construction, 131, 103896.

[14] Zaalouk, A., Moon, S. and Han, S., 2023. Operations planning and scheduling in off-site construction supply chain management: Scope definition and future directions. *Automation in construction*, *153*, 104952.

[15] Razali, N.M. and Geraghty, J., 2011. Genetic algorithm performance with different selection strategies in solving TSP. Proceedings of the world congress on engineering, Hong Kong, China. 1-6.

[16] Zhang, Y., Lei, Z., Han, S., Bouferguene, A. and Al-Hussein, M., 2020. Process-oriented framework to improve modular and offsite construction manufacturing performance. Journal of Construction Engineering and Management, 146(9), 04020116.

[17] Ramedani, Z., Omid, M., Keyhani, A., Shamshirband, S. and Khoshnevisan, B., 2014. Potential of radial basis function-based support vector regression for global solar radiation prediction. Renewable and Sustainable Energy Reviews, 39, 1005-1011.

Falling Objects/Debris Detection System Using Surveillance Camera at Construction Site

Songyuan Zheng, Meta Soy, Jinwoong Lee, Kyubyung Kang, Ph.D., Jiansong Zhang, Ph.D., and Emad Elwakil, Ph.D.

School of Construction Management Technology, Purdue University, USA
zheng776@purdue.edu, soym@purdue.edu, lee4067@purdue.edu, kyukang@purdue.edu, zhan3062@purdue.edu, eelwakil@purdue.edu

Abstract–

Falling objects and debris at construction sites pose significant safety hazards, often going unnoticed during operations. Identifying these incidents in post-construction reviews can enable safety managers to develop preventive action plans for future projects. This study uses existing surveillance cameras equipped with computer vision technology to detect, track, and count any fallen objects or debris. The proposed method achieved 100% accuracy across all metrics in Dataset 1, which was captured using a fixed camera. In contrast, Dataset 2, simulated with camera shake to mimic real-world conditions, showed decreased precision for smaller objects, with overall precision dropping to 95% and an F1 score of 98%. The system could be used to support the lesson-learned plan for future site safety project meetings, which could further minimize the potential objects/debris falling in construction projects.

Keywords –

Falling object; Falling debris; Safety; Computer vision.

1 Introduction

The construction site is a complex environment that could cause potential dangers to people working in the area. Especially with vertical structures such as buildings, falling debris could seriously harm the construction workers or other personnel. According to the International Safety Equipment Association (ISEA), a worker in the United State gets stuck by a falling object about every 10 minutes [1]. Those falling-debris or objects at construction sites could cause safety hazards resulting from potential injuries to casualties. To prevent those cases, the occupational safety and health administration (OSHA)'s guideline requires employers to equip workers with falling object protection gears, install necessary falling object protection systems, and guide on how to secure objects from falling [2]. Despite the effort, there are still accidents from falling objects happening.

With advancement in technology and availability of surveillance cameras in general, there are many studies trying to minimize construction site accidents by various means. For example, Yang et al. combined BIM with computer vision technology to detect the potential danger of falling when working near floor openings [3]. For the case of Personal Protection Equipment (PPE) detection, Zou et al. tested a light weight (You Only Look One, YOLOv5) model for helmet wearing detection when attached to the surveillance system [4]. In general, most research focuses on the prevention of accidents through developing a real time warning system [5], detecting the absence of PPEs [6], collision warning [7], monitoring falling of the workers [8] and so on.

Those studies mainly emphasize the planning and real-time case of the construction site safety. However, during the post-construction review, the feedback from all safety activities and data is critical for the safety manager to improve for the next project. Especially for the undetected or unreported accidents such as unnoticed falling debris/object was not mainly in the scopes of existing research, but that information is crucial in supporting the future safety action plan.

To address the gap, this study developed a falling objects/debris detection system using the OpenCV open-source tools, to detect, track and count the falling objects.

Two datasets are tested to validate the system. Dataset 1 is collected from a stable surveillance camera, while Dataset 2 is collected from a shaking surveillance camera to imitate the real construction site. The results show that for the Dataset 1, the precision, recall and F1 scores achieved 100%. However, for the Dataset 2, the system remained 100% recall, with a small drop in the precision (95%) and F1 score (98%). The results justify the system's effectiveness and its potential to support construction site safety strategies.

2 Literature Review

The study focuses on computer vision technology coupled with camera input to develop a system to count falling objects from the video data while the information is used for lessons learned in the post-construction review. Therefore, it is essential to investigate the technologies which were used for construction safety with computer vision applications. Based on the study done by Fang et al. (2020) [9] and Lee et al. (2023) [10], a variety of technologies have been developed for improving the safety in construction sites. In addition, the main enabling technology of computer vision is deep learning which is extensively used in many aspects including object detection, semantic segmentation and so on.

Deep learning is a multi-layer neural network architecture and in the case of computer vision, Convolutional Neural Networks (CNNs) allow computers to interpret and understand visual data. Fang et al. (2018) [6] used R-CNN architecture to input video from CCTV at construction site to detect workers who are not wearing safety hardhat. The proposed system obtains 95.7% precision and 94.9% recall respectively. Another popular object detection deep learning architecture is You Only Look One (YOLO). YOLO with various custom-made integrated layers is used in different safety conditions at construction sites. Zou et al. (2024) [4] proposed a YOLOv5 based light weight algorithm to detect workers wearing hats. The result suggests that the integration of the YOLOv5, Ghostbottle, and DWConv surpass the YOLOv5 model with 30% less model size. Despite good results in both accuracy and recall, the use of deep learning architecture requires massive amounts of datasets to train on. Therefore, regarding the falling object, Tu et al. (2024), inspired by the Moving Object Detection Dataset (MOD), was generated the called Falling Object Detection around Buildings (FADE) and FADE-net were proposed and trained on FADE dataset which contained fallen items such as bottles, packaging boxes, books, shoes, and so on [11].

While there is not much research using computer vision to detect falling objects and tracking them without the use of deep learning and dataset, there are some studies used OpenCV library to detect movement of any moving objects from video input. OpenCV library is a python open-source library which allows computers to visualize the object, allow the image processing, and is used with different machine leaning application for image detection, objection's motion, and so on. Pandey et al. [12] utilize contours to detect and track the motion of any moving object using OpenCV. Contours refer to the imaginary points on the boundaries of the focused object, highlighting two stages: object detection and object tracking. Through these stages, the authors can detect real-time objects frame by frame, and by connecting successive frames, they check the object's location in each frame and correlate them.

According to the above literature, and in the case of no dataset, using the computer vision with OpenCV approach is suitable for falling object/debris detection.

3 Methodology

The study aims to detect falling objects and count their number from surveillance camera footage. Unnoticed/unreported falling object/debris are recorded during the whole construction duration. The proposed algorithm enables the safety manager to easily review and count unreported falling object/debris after the construction by simply running algorithm with the video input. The information includes the falling object location based on the camera's field of view, number of objects, object types, and others can be used for post-construction safety analysis that will be key lessons learned for future projects.

The proposed algorithm employs the OpenCV package and consists of three steps, as depicted in Figure 1: Background Elimination, Object Detection, and Object Tracking.

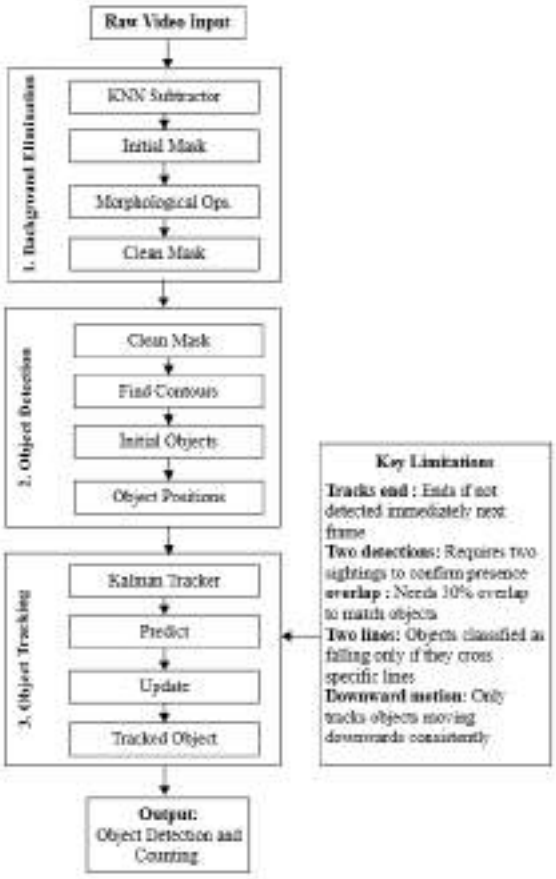


Figure 1. Proposed system architecture.

3.1 Background Elimination

This module aims to enhance the foreground scene for better analysis. The process begins with a background subtraction algorithm that extract moving foreground elements from a static background. This is followed by morphological operations—specifically, an opening to reduce noise and enhance object features within the mask by removing small, isolated elements. External contours of objects are then extracted from the refined mask to identify potential objects. Contours smaller than a predefined area threshold will be discarded to focus on significant objects. Bounding boxes are computed to save the detected objects in the remaining contours. The process isolates dynamic objects, enabling the following module: object detection and object tracking. Figure 2 shows how object's background is eliminated.



Figure 2. The effects of background elimination

3.2 Object Detection

In the object detection module, the cv2.findContours function is used to extract contours from the refined mask, specifically focusing on external contours to isolate significant object boundaries while ignoring internal variations. To avoid incorrect detections caused by camera vibration, the minimum area for contours is set to 10 pixels. This threshold ensures that smaller, potentially erroneous detections are disregarded, concentrating on more substantial and relevant objects. For the objects with the sizes beyond the minimum area, their bounding boxes (Figure 3) will be calculated and added to a list. This list will be saved and used in the next module.



Figure 3. The effects of object detection.

3.3 Object Tracking

To track the falling objects, an object tracking module is developed based on the Kalman filter with an IoU threshold of 0.3.

The objects which show in the list, saved in the last step will be chosen to be tracked. In this process, the existing KalmanBoxTracker algorithm from OpenCV package will be used with the following parameters: track termination window will be one frame, meaning that the object tracking will be terminated if the module did not detect the falling objecting in the next frame. And the objects and its trajectory will be saved in another if they are detected in the continuous two steps.

Furthermore, to avoid some mistakes caused by the shaking camera, a two-line detection, where the objects could cross two lines will be detected as a falling object, is added to this module to ensure the detected objects with downward direction. the module also evaluates the angle of movement by analyzing the trajectory of the bounding box centers over ten frames. Only objects that consistently show a downward direction are detected as falling, ensuring precise detection and tracking of objects under specific motion conditions. An object is counted as falling only (shown in Figure 4) if it satisfies both the line-crossing and consistent downward trajectory criteria. These conditions are designed to accurately identify and track falling objects, avoiding the impact of a vibrating camera (or wind and other external forces).

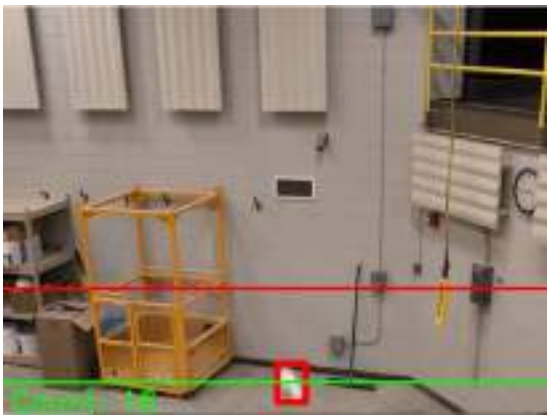


Figure 4. Detection and counting performance.

3.4 Evaluation Metrics

During the experiments, it was observed that some background elements were incorrectly identified as falling objects. In this context, evaluating the system's performance solely based on the ratio of detected objects to true falling objects would be misleading. Thus, to accurately assess the experimental results, Precision, Recall (also known as sensitivity), and F1 score are selected. The formulas for these metrics are provided in equations (1), (2), and (3). Correctly identified falling objects are True Positives (TP), while incorrect detections are False Positives (FP). Additionally, any falling objects that were not detected are considered False

Negatives (FN).

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

$$F1 = 2 * (\frac{Precision*Recall}{Precision+Recall}) \quad (3)$$

4 Implementation and Results

The proposed system was evaluated using a dataset collected from the Moss Lab at Purdue University (Figure 5). The dataset consists of two subsets, totaling 20 videos. One subset (Dataset 1) was captured using a stable camera mounted on a tripod, providing controlled conditions. The other subset (Dataset 2) simulated camera vibration to mimic real-world construction site conditions. Objects were intentionally dropped from a height of 15 feet (as shown in figure 6) without controlling their speed to mimic the unplanned and accidental dropping of items on construction sites, recorded with a Google Pixel 9 Pro XL smartphone at 12MP resolution and 60 frames per second, were analyzed for detection and tracking accuracy. Each video featured five distinct objects, as shown in Figure 7.



Figure 5. Example test of dropping object in Moss Lab.

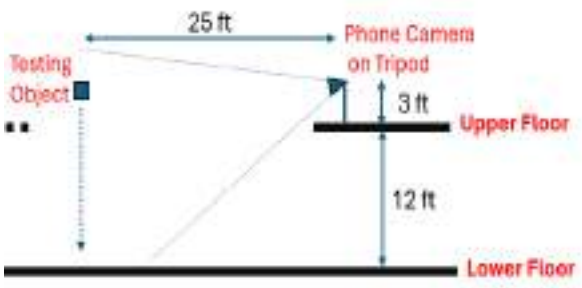


Figure 6. Testing setup layout (not to scale).



Figure 7. Object used for falling tests.

Under stable conditions (Dataset 1), the system achieved 100% precision, recall, and F1 scores for all objects as illustrated in Table 1. This demonstrates its high reliability in controlled environments, where object boundaries and movements are clearly distinguishable. The effectiveness of the system in Dataset 1 highlights its potential for detecting falling objects in situations where camera vibrations or environmental disturbances are minimal.

In dynamic conditions, Dataset 2, simulated through camera vibration, the system's performance slightly decreased for smaller objects. In Table 1, Precision for these objects dropped to 95% for the fourth object and 83% for the fifth object, while recall remained at 100% for all cases. This resulted in an overall precision of 95% and an F1 score of 98%. These results underscore the challenges posed by unstable camera conditions and the complexities of detecting smaller objects.

Figure 8 illustrates the challenges of false detections caused by background noise and camera vibration in Dataset 2. Subfigure 8a shows a static background element mistakenly identified as a falling object due to slight shifts in contour boundaries caused by vibrations. This misclassification demonstrates the system's sensitivity to motion artifacts, which can mimic object movement in the video. Subfigure 8b depicts another scenario where the system incorrectly identifies non-moving objects as falling, further highlighting the difficulties in isolating genuine object motion from dynamic background disturbances. These cases emphasize the importance of incorporating advanced filtering mechanisms to better handle background noise and reduce false positives.

Figure 9 provides an example of incorrect object trajectory classification under dynamic conditions. The figure shows an object misclassified as falling due to overlapping motion and inconsistent detection caused by vibrations. The yellow box and arrow highlight a non-falling object that was erroneously counted, revealing a limitation in the current trajectory analysis criteria. This issue is particularly prevalent in smaller objects, where fast movements and background inconsistencies can lead to misinterpretation. Enhancements such as refined trajectory angle analysis and improved stability algorithms could mitigate these errors.

Table. 1: Performance Evaluation on Dataset 1 and Dataset 2

Number of Objects Size(m^3)		object 1 0.0106	object2 0.00295	object3 0.00378	object 4 0.0005	object5 0.000145	Total
Dataset 1	count (by our system)	10	10	10	10	10	50
	True Positive	10	10	10	10	10	50
	False Positive	0	0	0	0	0	0
	False Negative	0	0	0	0	0	0
	Precision	100%	100%	100%	100%	100%	100%
	Recall	100%	100%	100%	100%	100%	100%
	F1 score	100%	100%	100%	100%	100%	100%
Dataset 2	count (by our system)	20	20	20	21	24	105
	True Positive	20	20	20	20	20	100
	False Positive	0	0	0	1	4	5
	False Negative	0	0	0	0	0	0
	Precision	100%	100%	100%	95%	83%	95%
	Recall	100%	100%	100%	100%	100%	100%
	F1 score	100%	100%	100%	98%	91%	98%



Figure 8. False detection due to camera movement.

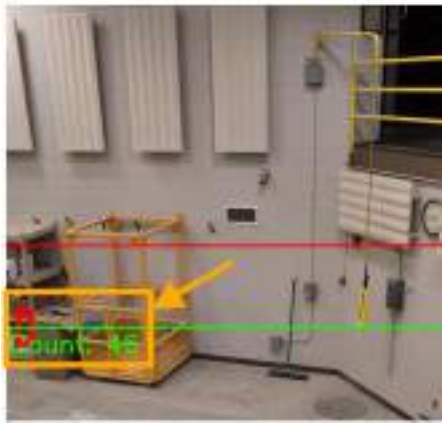


Figure 9. False detection sample.

The results demonstrate the system's robustness in detecting and tracking falling objects under stable conditions, achieving excellent performance metrics. While the system remains effective under dynamic conditions, challenges such as false detections and reduced precision for smaller objects highlight areas for improvement. Addressing these limitations will be critical for deploying the system in real-world construction environments.

A complete and robust algorithm can be used by the safety manager for post construction analysis by simply detecting and counting unreported failing object/debris from input surveillance footage. The system can also be integrated in real-time with the live CCTV camera running all day. The detected failing object/debris can be considered as almost real time or daily feedback to safety managers which enable them to investigate or create any necessary action plan.

5 Limitation and Future Work

Despite achieving high performance on both testing datasets, the following limitations of the proposed systems are acknowledged: (1) the study experimented on five household objects, which might not accurately represent most objects at construction sites, which can vary in size and weight. (2) The experiment was set up indoors in the construction lab, which was susceptible only to the minimum wind, which could affect the motion of the falling object. (3) The camera set up and illumination conditions in the experiments, including the height and orientation of the camera as well as the laboratory lighting, could also vary.

The authors recommend future studies on conditions of different falling objects mainly taken or selected from construction sites. Direct testing in the construction site environment with existing CCTV cameras at different frame rates will be ideal. A real-time warning system could be developed based on the results of the proposed detection system. Specifically, the system would issue alerts via mobile devices whenever a falling object is detected by cameras positioned at various heights, ensuring that construction workers in different locations receive timely warnings. Additionally, the speed and number of falling objects recorded could be used to assess the risk level at the construction site, thereby enhancing safety measures and protocols.

6 Conclusion

Data about falling objects/debris during construction is crucial for lessons learned in the safety framework for construction projects. By leveraging the existing information from surveillance cameras and computer vision technology, the paper successfully proposed a system that can detect, track, and count falling objects/debris at construction sites. Based on the location of the CCTV camera field of view and the trajectory of the falling objects, the safety managers could analyze the cause of the falling case to improve the prevention/action plan further. According to the experimental results, the system shows good object tracking abilities with the counting 100 % precision, recall, and F1 for non-vibrating camera. In the camera vibration scenario, the system achieves 95% precision, 100% recall, and 98% F1 score, approximating real-world conditions more closely. Given the limitations of this study in the testing scenarios and testing environment, future study on different weight/types of falling objects, higher fame rate of the input video, and the use of different moving object detection algorithms are recommended.

7 References

[1] N. Randall. ISEA Release Updated Dropped Object Standard. On-line: https://safetyequipment.org/isea-releases-new-dropped-objects-standard/, Accessed: 12/22/2024.

[2] Occupational Safety and Health Administration, Fall Protection Systems and Falling Object Protection, OSHA Standard 1910.29, 2023. On-line: https://www.osha.gov/laws-regs/regulations/standardnumber/1910/1910.29. Accessed: Dec. 22, 2024.

[3] B. Yang, B. Zhang, Q. Zhang, Z. Wang, M. Dong, and T. Fang, Automatic detection of falling hazard from surveillance videos based on computer vision and building information modeling, Structure and Infrastructure Engineering, vol. 18, no. 7, pp. 1049-1063, 2022.

[4] X. Zou, Z. Chen, and Y. Zhou, A Lightweight Model Based on YOLOv5 for Helmet Wearing Detection, Proceedings of SPIE, 2024. doi: 10.1117/12.2627279

[5] Y. Jeon, D. Q. Tran, A. S. Kulinan, T. Kim, M. Park, and S. Park, Vision-based motion prediction for construction workers safety in real-time multi-camera system, Advanced Engineering Informatics, vol. 102898, 2024.

[6] Q. Fang, H. Li, X. Luo, L. Ding, H. Luo, T. M. Rose, and W. An, Detecting non-hardhat-use by a deep learning method from far-field surveillance videos, Automation in Construction, vol. 85, pp. 1–9, 2018.

[7] W. Gan, K. Gu, J. Geng, C. Qiu, R. Yang, H. Wang, and X. Hu, A Novel Three-Stage Collision-Risk Pre-Warning Model for Construction Vehicles and Workers, Buildings, vol. 14, no. 8, p. 2324, 2024.

[8] A. Danilenka, P. Sowiński, K. Rachwał, K. Bogacka, A. Dąbrowska, M. Kobus, K. Baszczyński, M. Okrasa, W. Olczak, P. Dymarski, I. Lacalle, M. Ganzha, and M. Paprzycki, Real-Time AI-Driven Fall Detection Method for Occupational Health and Safety, Electronics, vol. 12, no. 20, p. 4257, 2023.

[9] W. Fang, L. Ding, P. E. D. Love, H. Luo, H. Li, F. Peña-Mora, B. Zhong, and C. Zhou, Computer vision applications in construction safety assurance, Automation in Construction, 110: 103013, 2020.

[10] J. Lee and S. Lee, Construction Site Safety Management: A Computer Vision and Deep Learning Approach, Sensors, 23(2): 944, 2023.

[11] Z. Tu, Z. Gao, Z. Zhang, C. Zhou, J. Yuan, and B. Du, "FADE: A Dataset for Detecting Falling Objects around Buildings in Video," arXiv, 2024.

[12] H. Pandey, P. Choudhary, and A. Singh, Detect and Track the Motion of Any Moving Object Using OpenCV, in R. Agrawal, J. He, E. Shubhakar Pilli, and S. Kumar, Eds., Cyber Security in Intelligent Computing and Communications, Studies in Computational Intelligence, vol. 1007, Springer, Singapore, 2022.

Blockchain-based Framework for Securing Construction Worker Certification and Task Authorization

Jaewook Jeong[1*], Louis Kumi[1], Jaemin Jeong[2], Hyeongjun Mun[1], Minji Kim[1], Sangho Yun[1], Jaehui Son[1], Minsang Gu[1], Minwoo Song[1], Jiwon Hwang[1]

[1] Department of Safety Engineering, Seoul National University of Science and Technology, Seoul, 01811, Republic of Korea
[2] Department of Civil & Mineral Engineering, University of Toronto, Toronto, Canada
jaewook.jeong@seoultech.ac.kr*

Abstract –

The construction industry is widely recognized for its high risk of workplace accidents, making effective safety management a critical priority. Traditional safety systems often rely on manual processes and fragmented documentation, leading to inefficiencies and human error in managing worker certifications and task authorizations; Therefore, this study presents a blockchain-based decentralized application (DApp) for managing worker certifications and task authorizations using smart contracts on the Ethereum blockchain. The proposed system automates certification verification, ensuring that only certified workers are authorized for specific tasks, thereby reducing accidents caused by unqualified personnel. The DApp features a dual-application interface: one for worker registration and task authorization, and another for on-site certification verification. Testing and validation in a simulated environment demonstrated enhanced data integrity, real-time certification checks, and secure storage of certification records, preventing unauthorized modifications. This research highlights the potential of blockchain technology to revolutionize construction safety practices by offering an efficient, transparent, and tamper-proof certification management system. The findings contribute to the growing body of knowledge on blockchain applications in construction and lay the foundation for future advancements in safety compliance technologies.

Keywords – Blockchain; Construction Safety; Smart Contracts; Certification Management; Decentralized Applications (DApp)

1 Introduction

The construction industry is globally recognized as one of the most hazardous sectors, consistently leading in workplace fatalities, severe injuries, and occupational health risks [1–4]. Although many countries have implemented strict safety regulations and mandatory certification requirements, the industry continues to struggle with persistent safety challenges [5,6]. Ensuring that only properly trained and certified personnel are authorized to perform specific tasks is essential for mitigating safety risks on construction sites [5,7]. However, traditional systems, such as paper-based records or isolated databases, are prone to errors and inefficiencies, creating safety gaps [8]. Paper-based systems are prone to documentation loss, storage challenges, and tampering, while conventional digital databases face limitations in data security, cross-organizational verification, and real-time accessibility [9]. These shortcomings allow uncertified or improperly certified workers to undertake high-risk tasks, often resulting in preventable accidents and fatalities [7]. The growing complexity of construction projects, characterized by multiple contractors, diverse workforces, and varying regional safety standards, further complicates the task of effectively managing and verifying worker credentials [10].

Blockchain technology offers a promising approach to address these challenges by providing immutability, decentralization, and automation, which are critical for ensuring secure and tamper-proof certification management [11]. Unlike AI, which can be used for certification validation but still relies on centralized databases, blockchain eliminates single points of failure and reduces trust dependency by ensuring that all transactions are recorded transparently and cannot be altered retroactively [11]. The decentralized nature of blockchain ensures that worker credentials remain verifiable across multiple organizations without the risk of fraudulent modifications [11]. Additionally, smart contracts provide self-executing compliance mechanisms, automatically verifying worker certifications before authorizing hazardous tasks, thereby reducing the risk of manual oversight or manipulation [12]. By eliminating

manual checks and enabling real-time compliance, blockchain-based systems can streamline safety management and reduce the risk of accidents. By introducing a blockchain-based decentralized application for real-time certification verification, this study addresses critical gaps in traditional safety systems, such as inefficiency, human error, and limited scalability, presenting a novel approach to safety compliance in construction.

This study explores the development of a blockchain-based decentralized application (DApp) for managing worker certifications and authorizing tasks in the construction industry. The proposed system uses smart contracts on a private Ethereum blockchain to securely store worker data, authorize tasks based on certification validity, and enable on-site certification checks. Additionally, a dual-application interface is designed to facilitate worker registration, task authorization, and on-site certification verification. Through this approach, the study aims to demonstrate how blockchain can enhance safety compliance, improve efficiency in authorization processes, and reduce human error in high-risk environments.

By combining blockchain technology with automated compliance checks, this research contributes to the advancement of safety management practices in construction. It presents a model that construction companies and safety managers can adopt to strengthen their oversight of worker qualifications, minimize unauthorized task assignments, and promote a safer, more efficient construction environment.

2 Literature Review

The construction industry is known for its high-risk nature, with worker safety remaining a significant concern globally. Many studies have explored various aspects of safety management, highlighting the critical role of proper training, certification, and task authorization in minimizing accidents [5,13–15]. This section provides an overview of key research in the areas of worker safety, certification management, and the role of blockchain technology in enhancing safety practices in the construction sector.

2.1 Worker Safety in Construction

Construction workers are exposed to a range of hazards, from physical injuries due to heavy machinery and equipment to long-term health risks like exposure to harmful chemicals and dust. According to a study by Newaz et al. (2021), effective safety management on construction sites requires comprehensive strategies that address both the physical and organizational aspects of safety [16]. These strategies include safety training, hazard identification, risk assessment, and the use of safety equipment. However, the effectiveness of these strategies often depends on the qualifications and competencies of the workers [16].

Proper training and certification have been identified as fundamental in mitigating risks [14]. Research by Başağa et al. (2018) emphasize that certified workers are significantly less likely to be involved in accidents, as they possess the requisite knowledge to identify hazards and follow safety protocols [14]. However, traditional certification systems often fail to address challenges like discrepancies in worker qualifications and the verification of training completion. These limitations create a pressing need for secure, real-time, and scalable certification management solutions.

2.2 Blockchain Technology in Construction

Blockchain technology has emerged as a promising solution for addressing challenges in various industries, including construction [11,12,17]. Blockchain is essentially a distributed ledger that records transactions in a secure, transparent, and tamper-proof manner. The decentralized nature of blockchain ensures that no single party has control over the data, thus reducing the risk of fraud or manipulation.

A review by Li and Kassem (2021) outlines several potential applications of blockchain in the construction sector, including supply chain management, project financing, and safety compliance [11]. While blockchain's utility in supply chain transparency and contract enforcement is well-documented, its application to worker certification management and task authorization remains underexplored. In the context of safety management, blockchain can significantly improve the accuracy and reliability of information [18–20]. A study by Yang et al (2022) demonstrated how blockchain can be used to manage scaffolding work information [21].

Smart contracts are one of the most promising applications of blockchain technology in construction [12,22]. These are programmable contracts that automatically execute predefined actions when certain conditions are met. In the context of worker certification and task authorization, smart contracts can automatically verify a worker's certification status and authorize or deny access to specific tasks based on pre-defined criteria.

Research by Ciotta et al. (2021) demonstrates that smart contracts can be used to automate various processes in the construction industry, such as managing payments, enforcing legal agreements, and verifying qualifications [12]. In the context of safety management, smart contracts can be used to ensure that workers are only assigned tasks they are qualified to perform, based on their training and certification records. The use of smart contracts can greatly reduce the administrative burden of manual checks, ensuring that safety standards

are consistently met across construction projects. Building on this foundation, our research focuses on using smart contracts for real-time certification verification, a novel application that directly addresses safety compliance gaps in construction.

2.3 Research Gap and Contribution

This study contributes to the literature by addressing several key gaps in the current research on construction safety and certification management. Firstly, it explores the use of blockchain technology for real-time worker certification and task authorization, which is largely underexplored in the existing literature. Secondly, it tackles the challenges of scalability and data security by implementing a blockchain-based solution in a simulated environment, providing insights into its feasibility for large-scale construction projects.

The novelty of this research lies in its approach to applying blockchain with smart contracts for real-time safety compliance, providing a new method for ensuring that only certified workers are authorized to perform hazardous tasks. By automating the verification of worker qualifications and integrating this process with on-site certification checks, the proposed system offers a practical tool for improving safety management in the construction sector. This solution not only reduces administrative overhead but also enhances transparency and accountability, potentially reducing accidents and improving overall project safety.

3 Methodology

3.1 Research Design

This study employed a blockchain-based system design to enhance worker safety in the construction industry by ensuring only certified and trained workers were authorized for specific tasks. The system was developed as a decentralized application (DApp) using smart contracts on the Ethereum blockchain [19,23]. The research design involved the following steps as shown in Figure 1:

- System Development: Developed a smart contract to record worker certifications, training completion, and task authorizations.
- User Interface: Created two separate frontend applications, one for worker registration and task authorization, and another for certification checking at construction sites.
- Testing and Validation: Performed functional testing to ensure the system accurately registered workers, authorized tasks, and verified certifications.

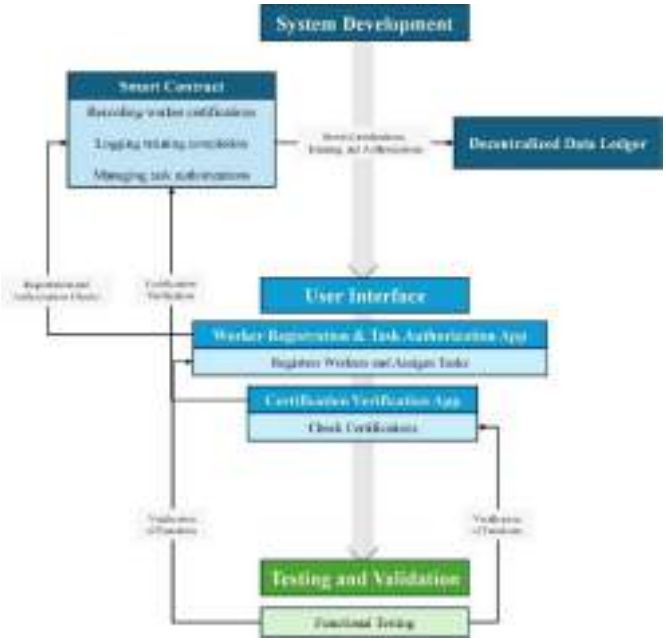


Figure 1. Research framework

3.2 System Architecture

The architecture consisted of four main components:

1. Blockchain Network: A private Ethereum network was set up using Ganache, simulating a blockchain environment for development and testing.
2. Smart Contract: A smart contract was written in Solidity and deployed on the Ethereum blockchain, maintaining data on worker IDs, certifications, training status, and authorized tasks.
3. Frontend Applications: Two React.js applications were developed to interact with the smart contract. The first application managed worker registration and task authorization, while the second was used at construction sites to verify certifications.
4. MetaMask Integration: MetaMask was used as the web3 provider to securely connect the frontend applications with the Ethereum network. It allowed users to interact with the smart contract by signing and authorizing transactions directly through the MetaMask wallet. This provided a secure and user-friendly interface for managing blockchain-based interactions.

3.3 Smart Contract Development

The smart contract was developed using the Solidity 0.8.0 programming language on the remix integrated development environment (IDE) [24] as shown in Table 1, and key functionalities were included as follows:

- Worker Registration: The contract enabled

administrators to register workers, recording details such as worker ID, name, certification type, and certification status.

- Task Authorization: The contract allowed administrators to authorize specific tasks for workers based on their certification and training status.
- Certification Verification: A function was provided for querying a worker's certification status and authorized tasks to confirm their eligibility for on-site activities.

The contract code was tested using Remix IDE to ensure it met functional requirements and handled error scenarios, such as unauthorized access.

Table 1 Code for the Smart Contract for Worker Certification and Task Authorization

```
// SPDX-License-Identifier: MIT pragma solidity ^0.8.0;
contract WorkerCertification {
    // Worker data structure
    struct Worker {
        string workerId;
        string name;
        string certificationType;
        bool certificationValid;
        bool trainingCompleted;
    }
    // Mappings for workers and their authorized tasks
    mapping(string => Worker) public workers;
    mapping(string => string[]) public authorizedTasks;
    // Event for certification checks
    event CertificationChecked(string workerId, string taskId, bool isCertified);
    // ================================
    // Worker Registration
    // ================================
    function registerWorker(
        string memory _workerId,
        string memory _name,
        string memory _certificationType,
        bool _certificationValid,
        bool _trainingCompleted
    ) public {
        workers[_workerId] = Worker(
            _workerId,
            _name,
            _certificationType,
            _certificationValid,
            _trainingCompleted
        );
    }
    // ================================
    // Task Authorization
    // ================================
    function authorizeTask(string memory _workerId, string memory _taskId) public {
        authorizedTasks[_workerId].push(_taskId);
    }
    // ================================
    // Certification Verification
    // ================================
    function checkCertification(string memory _workerId, string memory _taskId) public {
        bool isCertified = workers[_workerId].certificationValid && workers[_workerId].trainingCompleted;
        emit CertificationChecked(_workerId, _taskId, isCertified);
        require(isCertified, "Worker is not certified for this task");
    }
    function getWorker(string memory _workerId)
        public
        view
        returns (
            string memory name,
            string memory certificationType,
            bool certificationValid,
            bool trainingCompleted
        )
    {
        Worker memory worker = workers[_workerId];
        return (
            worker.name,
            worker.certificationType,
            worker.certificationValid,
            worker.trainingCompleted
        );
    }
}
```

3.4 Frontend Development

Two separate frontend applications were built using React.js to interact with the smart contract [25]:

1. Worker Registration and Authorization Application: This application allowed administrators to register workers, authorize tasks, and manage certifications. It connected to the smart contract using Web3.js, enabling interaction with the blockchain.
2. Certification Verification Application: This application allowed on-site managers to verify a worker's certification status and task authorization by entering the worker ID and task ID. If the worker was certified and authorized, they were allowed to proceed with the task.

4 Results and Discussion

4.1 System Functionality

The blockchain-based certification management system was successfully implemented and tested in a simulated environment. The system achieved key functionalities, including worker registration, task authorization, and real-time certification verification at construction sites. Testing revealed that both frontend applications, the worker registration and authorization application and the certification verification application as shown in Figures 2 and 3, performed as intended, providing accurate, secure, and reliable data retrieval. The decentralized approach demonstrated a clear benefit by preventing data tampering and ensuring that only verified certifications were used to authorize tasks.



Figure 2. User interface for worker registration and task authorization



Figure 3. User interface for certification check

4.2 System Performance

The system's performance was evaluated by examining transaction execution in the Ganache environment, as shown in Figure 4. Ganache, a personal Ethereum blockchain, provided insight into the resource utilization and transaction handling efficiency of the Worker Certification DApp.

In Figure 4, each transaction involved deploying and interacting with the smart contract for worker registration and certification checking. The transaction logs indicate multiple interactions, including contract deployment, worker registration, and certification verification. MetaMask facilitated these interactions by enabling users to review and confirm each transaction before execution, providing a layer of security and transparency. Key aspects of the transactions are summarized below:

- Address and Balance: The Ethereum accounts displayed in Ganache each had initial balances, with balances updated after each transaction. For instance, the primary account used for transactions shows a slight reduction in balance due to gas consumption during contract calls.
- Gas Usage: The system consumed varying amounts of gas for different contract functions. Notably, the register Worker and check Certification functions incurred gas usage of approximately 68,487 units for certification-related transactions, suggesting the efficiency of these operations. This gas cost is reasonable for a private blockchain environment, highlighting the contract's resource efficiency.

The system was tested using a single worker registration, one task authorization, and one certification verification transaction. The purpose of this test was to validate the core functionality of the smart contract rather than stress-test scalability. The simulation was conducted in a controlled private Ethereum network using Ganache, where block mining was instantaneous, and gas fees were fixed. This setup ensured that transactions were processed efficiently, allowing us to verify the correctness of the smart contract logic.

The DApp successfully performed task authorization and certification verification within Ganache's simulated environment. Transactions were processed almost instantaneously due to Ganache's high-speed block mining, while MetaMask allowed secure and seamless interaction with the blockchain. In a real-world scenario, additional testing with multiple workers and varying workloads will be required to evaluate system performance under more complex conditions. Future work will focus on scalability analysis, transaction efficiency, and potential optimizations for deployment in an industry setting.



Figure 4. Transaction record in Ganache blockchain

4.3 Safety and Data Integrity

A major advantage of the blockchain-based approach is the secure storage of certification and authorization data, preventing unauthorized modifications. The smart contract's immutable nature ensured that worker data could not be altered once registered, enhancing data integrity. This feature is particularly valuable for construction safety, where accurate certification records are critical for regulatory compliance and risk management.

The results showed that the decentralized structure also effectively restricted access to only certified workers for specific tasks, reducing the likelihood of human error or intentional breaches in safety protocol. By enabling verification on a blockchain, the system provided transparency and accountability, which are essential in high-risk industries such as construction.

To address human error in certificate management, the worker registration and certification data entry process is managed by designated administrators. Before a new worker arrives on-site, their certification details are manually recorded into the system, verified against official records, and then uploaded to the blockchain. While manual data entry may introduce potential human errors, blockchain ensures that once recorded, data cannot be altered or manipulated.

4.4 Implications for Construction Safety

The proposed blockchain-based system introduces transformative changes to safety compliance in construction. By ensuring that only certified and trained personnel are authorized for hazardous tasks, the system directly addresses one of the primary causes of construction accidents: unqualified workers. The transparency of blockchain transactions further supports accountability, as immutable records provide a verifiable audit trail of certifications and authorizations.

In addition to improving safety, the system offers economic benefits by reducing administrative overhead associated with manual certification checks and preventing costly accidents caused by uncertified workers. To quantify these benefits, key performance metrics can be considered. These include the reduction in time spent on manual certification verification, which in traditional systems can be time-intensive, particularly for large projects involving multiple subcontractors. Furthermore, a reduction in administrative labor costs can be measured by comparing the staff hours required for certification management before and after blockchain implementation. Additionally, based on existing safety literature, a decrease in accident rates attributed to unauthorized or uncertified workers could provide an estimate of potential cost savings related to reduced compensation claims, legal liabilities, and project delays. While the implementation of blockchain-based certification management incurs initial development and integration costs, the long-term economic advantages can outweigh these expenses. The trade-off between cost and benefits can be analyzed through a cost-benefit framework, where initial investments in blockchain infrastructure, training, and deployment are weighed against projected savings in administrative labor, accident-related costs, and regulatory compliance efforts. This discussion underscores the system's feasibility and provides a foundation for future empirical validation in real-world construction environments. These findings suggest that blockchain technology has the potential to become an industry standard for safety compliance in construction.

5 Conclusion

This research demonstrates the potential of blockchain technology to revolutionize safety management in the construction industry by enhancing

the verification of worker qualifications and task authorizations. By developing a blockchain-based decentralized application (DApp) that utilizes smart contracts on the Ethereum blockchain, this study successfully addresses key challenges in traditional safety management systems, such as inefficiency, human error, and unauthorized task assignments. The system provides a secure, transparent, and tamper-proof method for managing worker certifications and training completion, ensuring that only qualified workers are authorized for specific tasks.

The results from system testing show that the blockchain-based approach not only improved data integrity by preventing unauthorized modifications but also facilitated real-time certification checks at construction sites, offering immediate verification of worker eligibility for hazardous tasks. The two frontend applications developed, one for worker registration and task authorization and another for certification verification, performed effectively, demonstrating the practical feasibility of blockchain in the construction sector.

This study makes the following contributions. Firstly, in terms of technical contribution, this study provides a secure, transparent, and automated method for managing worker certifications using smart contracts, enhancing data integrity and reducing human errors. Secondly, regarding research contribution, the study demonstrates the feasibility of blockchain for safety compliance, laying the foundation for integrating blockchain with other construction systems and highlighting potential applications in high-risk industries.

Despite the promising results, this study has some limitations that should be addressed in future research. Firstly, the usability of the system was not evaluated, as this study focused primarily on developing the framework and testing its technical functionalities. Secondly, the system was tested in a simulated environment using a private Ethereum blockchain, which may not fully replicate the complexities and scalability challenges of real-world construction sites. Lastly, the user interface developed for this study was designed for administrative and supervisor-level interactions, but worker-facing applications were not explored in detail.

Future research should also consider key implementation requirements for real-world adoption, including scalability, interoperability with existing construction management systems, security, and regulatory compliance. Ensuring the system can handle a large volume of transactions efficiently is crucial for scalability, while integration with current digital safety platforms will enhance its usability. Additionally, robust security measures must be in place to prevent unauthorized access or tampering. Finally, legal and regulatory considerations should be addressed to align the system with construction industry safety standards and compliance frameworks.

Acknowledgement

This research was financially supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2023-00213165).

References

[1] L. Kumi, J. Jeong, J. Jeong, Systematic Review of Quantitative Risk Quantification Methods in Construction Accidents, Buildings 14 (2024). https://doi.org/10.3390/buildings14103306.

[2] L. Kumi, J. Jeong, J. Jeong, Data-driven automatic classification model for construction accident cases using natural language processing with hyperparameter tuning, Autom Constr 164 (2024) 105458. https://doi.org/10.1016/J.AUTCON.2024.105458.

[3] J. Jeong, J. Jeong, Quantitative Risk Evaluation of Fatal Incidents in Construction Based on Frequency and Probability Analysis, Journal of Management in Engineering 38 (2022) 4021089. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000998.

[4] L. Kumi, J. Jeong, J. Jeong, Assessing Dust Emissions, Health Impacts, and Accident Risks in Prefabricated and Conventional Construction: A Comprehensive Comparative Study, Buildings 13 (2023). https://doi.org/10.3390/buildings13092305.

[5] M. Kim, D. Jo, J. Jeong, Evaluation of Mobile Risk Perception Training System for Improving the Safety Awareness of Construction Workers, Buildings 13 (2023). https://doi.org/10.3390/buildings13123024.

[6] A. Al-Otaibi, A.F. Kineber, Identifying and Assessing Health and Safety Program Implementation Barriers in the Construction Industry: A Case of Saudi Arabia, Applied Sciences 13 (2023). https://doi.org/10.3390/app13042630.

[7] Q. Fang, H. Li, X. Luo, L. Ding, T.M. Rose, W. An, Y. Yu, A deep learning-based method for detecting non-certified work on construction sites, Advanced Engineering Informatics 35 (2018) 56–68. https://doi.org/10.1016/J.AEI.2018.01.001.

[8] A.P. Chassiakos, S.P. Sakellaropoulos, A web-based system for managing construction information, Advances in Engineering Software

39 (2008) 865–876. https://doi.org/10.1016/J.ADVENGSOFT.2008.05.006.

[9] X. Zhao, Z. Zhang, R. Hu, J. Liu, X. Yang, R. Zhang, H. Gao, Blockchain Technology Based Digital Document Management System Design, in: 2022 7th Asia Conference on Power and Electrical Engineering (ACPEE), 2022: pp. 440–446. https://doi.org/10.1109/ACPEE53904.2022.9783819.

[10] Z. Lafhaj, S. Rebai, W. AlBalkhy, O. Hamdi, A. Mossman, A. Alves Da Costa, Complexity in Construction Projects: A Literature Review, Buildings 14 (2024). https://doi.org/10.3390/buildings14030680.

[11] J. Li, M. Kassem, Applications of distributed ledger technology (DLT) and Blockchain-enabled smart contracts in construction, Autom Constr 132 (2021) 103955. https://doi.org/10.1016/J.AUTCON.2021.103955.

[12] V. Ciotta, G. Mariniello, D. Asprone, A. Botta, G. Manfredi, Integration of blockchains and smart contracts into construction information flows: Proof-of-concept, Autom Constr 132 (2021) 103925. https://doi.org/10.1016/J.AUTCON.2021.103925.

[13] C.M. Tam, S.X. Zeng, Z.M. Deng, Identifying elements of poor construction safety management in China, Saf Sci 42 (2004) 569–586. https://doi.org/10.1016/J.SSCI.2003.09.001.

[14] H.B. Başağa, B.A. Temel, M. Atasoy, İ. Yıldırım, A study on the effectiveness of occupational health and safety trainings of construction workers in Turkey, Saf Sci 110 (2018) 344–354. https://doi.org/10.1016/J.SSCI.2018.09.002.

[15] A.A. Tabassi, A.H.A. Bakar, Training, motivation, and performance: The case of human resource management in construction projects in Mashhad, Iran, International Journal of Project Management 27 (2009) 471–480. https://doi.org/10.1016/J.IJPROMAN.2008.08.002.

[16] M.T. Newaz, M. Ershadi, M. Jefferies, P. Davis, Assessing safety management factors to develop a research agenda for the construction industry, Saf Sci 142 (2021) 105396. https://doi.org/10.1016/J.SSCI.2021.105396.

[17] B. Zhong, X. Pan, L. Ding, Q. Chen, X. Hu, Blockchain-driven integration technology for the AEC industry, Autom Constr 150 (2023) 104791. https://doi.org/10.1016/J.AUTCON.2023.104791.

[18] J. Xu, W. Lu, L. Wu, J. Lou, X. Li, Balancing privacy and occupational safety and health in construction: A blockchain-enabled P-OSH deployment framework, Saf Sci 154 (2022) 105860. https://doi.org/10.1016/J.SSCI.2022.105860.

[19] S. Ahmadisheykhsarmast, S. Aminbakhsh, R. Sonmez, F. Uysal, A transformative solution for construction safety: Blockchain-based system for accident information management, J Ind Inf Integr 35 (2023) 100491. https://doi.org/10.1016/J.JII.2023.100491.

[20] C. Baek, D. Lee, C. Park, Blockchain based Framework for Verifying the Adequacy of Scaffolding Installation, in: "Furuya Hiroshi" "Tateyama Kazuyoshi" "Osumi Hisashi" (Ed.), Proceedings of the 37th International Symposium on Automation and Robotics in Construction (ISARC), International Association for Automation and Robotics in Construction (IAARC), Kitakyushu, Japan, 2020: pp. 425–432. https://doi.org/10.22260/ISARC2020/0060.

[21] J. Yang, D. Lee, C. Baek, C. Park, B.Q. Lan, D. Lee, Leveraging Blockchain for Scaffolding Work Management in Construction, IEEE Access 10 (2022) 39220–39238. https://doi.org/10.1109/ACCESS.2022.3165614.

[22] S. Ahmadisheykhsarmast, R. Sonmez, A smart contract system for security of payment of construction contracts, Autom Constr 120 (2020) 103401. https://doi.org/10.1016/J.AUTCON.2020.103401.

[23] W. Cai, Z. Wang, J.B. Ernst, Z. Hong, C. Feng, V.C.M. Leung, Decentralized Applications: The Blockchain-Empowered Software System, IEEE Access 6 (2018) 53019–53033. https://doi.org/10.1109/ACCESS.2018.2870644.

[24] M. Wohrer, U. Zdun, Smart contracts: security patterns in the ethereum ecosystem and solidity, in: 2018 International Workshop on Blockchain Oriented Software Engineering (IWBOSE), 2018: pp. 2–8. https://doi.org/10.1109/IWBOSE.2018.8327565.

[25] S.C. Panda Sandeep Kumar and Satapathy, An Investigation into Smart Contract Deployment on Ethereum Platform Using Web3.js and Solidity Using Blockchain, in: S.C. and T.-G.C.M. and A.V.N.M. Bhateja Vikrant and Satapathy (Ed.), Data Engineering and Intelligent Computing, Springer Singapore, Singapore, 2021: pp. 549–561.

Utilization of Artificial Intelligence in HAZOP Studies and Reports

Ehab Elhosary[1], Osama Moselhi [1,2]

[1]Department of Building, Civil and Environmental Engineering, Concordia University, Montreal, QC, Canada
[2]Centre for Innovation in Construction and Infrastructure Engineering and Management (CICIEM), Concordia University, Montreal, QC, Canada
ehab.elhosary@mail.concordia.ca, moselhi@encs.concordia.ca

Abstract –

The Hazard and Operability (HAZOP) study is the most widely used hazard analysis technique in the process industry, aimed at enhancing safety and preventing accidents. It identifies potential hazards and operability malfunctions by dividing the design into small sections, called nodes, and analyzing each node's process separately. This paper briefly describes HAZOP studies and their role in enhancing operational safety across process industries. It also describes the advancements made to facilitate the performance of these studies and their challenges. To address these challenges, a new framework integrating BERTopic into HAZOP studies is proposed for enhanced efficiency and accuracy. The framework leverages historical data to categorize HAZOP elements into topics and extract node process and equipment descriptions to generate an intelligent pre-populated HAZOP analysis table. This paper focuses on categorizing causes into main risk factors for each HAZOP node and prioritizing them based on the likelihood of occurrence for each factor. The BERTopic model, incorporating embedding generation, dimensionality reduction, clustering, and topic representation, was applied to 1,574 HAZOP records from an oil pump station. The model achieved coherence and diversity scores of 80% and 92.4% respectively, outperforming Latent Dirichlet Allocation (LDA) model at 45.4% and 88.8%. It identified 13 topics, validated against hazard causes in oil pump stations and pipelines from literature. This model can be extended to categorize consequences and countermeasures, prioritizing them by severity and risk levels to generate a prepopulated table. This table can guide participants during sessions, significantly reducing the time required for the final HAZOP report.

Keywords –

HAZOP; Artificial intelligence; BERTopic; LDA; Coherence Score; Topic Diversity Score.

1 Introduction

Hazard analysis encompasses identifying potential hazards, assessing scenarios that could result in adverse outcomes, and developing countermeasures to eliminate or mitigate these hazards [1]. In this regard, hazard and operability (HAZOP) analysis has been the most widely used technique in the process industry. This is due to its simplicity, systematic approach, thoroughness, structured brainstorming process, and applicability to a wide range of systems [2]. The HAZOP concept dates to 1974 following a large explosion in England, which resulted in the deaths of 28 workers and injuries to 36 others. An international standard for HAZOP studies [BS IEC 61882] was published in 2001 and updated in 2016.

A HAZOP study is typically conducted during the design stage to identify and assess potential hazards, their causes, consequences, safeguards, and recommendations. The primary inputs are Piping and Instrumentation Diagrams (P&IDs) and the expertise of participants. P&IDs provide detailed design and engineering information about the plant. Participants contribute specialized knowledge of processes, drawing upon their professional experience [3]. The process begins by dividing the P&IDs into clearly defined sections, referred to as nodes, to ensure that each piece of equipment in the process is analysed thoroughly. The study employs guidewords (e.g., no, more, less of) in conjunction with parameters (e.g., pressure, temperature, flow) to identify deviations from the normal operating conditions. This systematic approach is applied within a specific node. Once deviations are identified, the team investigates their potential causes and consequences and identifies safeguards and recommendations to prevent or mitigate the hazardous situation [4]. The HAZOP study has gained regulatory acceptance in the chemical industry, and has been extended to other industries, including oil and gas, petrochemical plants, nuclear power, environmental engineering, and infrastructure projects [5]. However, it often faces quality issues stemming from complexity,

time consumption, reliance on expert judgment and knowledge loss. Completing the analysis for a typical chemical process takes 1–8 weeks and for large-scale processes, such as plants with over 200 P&IDs, the required time increases significantly [6]. To overcome these issues, intelligent systems with various levels of automation have been developed to reduce time, cost, and human bias [5].

In this respect, this paper introduces the evolution of HAZOP automation within the process industry and highlights the challenges of existing systems. To address these limitations, it proposes a novel framework that leverages artificial intelligence (AI) tools to generate an intelligent pre-populated HAZOP analysis table, designed to assist participants during workshop sessions.

2 Literature Review

HAZOP automation has been a research focus for over 30 years, developing intelligent systems, including knowledge-based systems, model-based approaches, and data-driven models

Knowledge-based systems are computer systems that use knowledge representation to provide expert knowledge or draw conclusions [7]. These systems, include expert systems, computer simulations, integrated based tools, ontologies, and Bayesian networks [5].

Expert systems emulate human logic through knowledge databases that store domain knowledge, inference engines to deduce conclusions based on predefined rules, and user interfaces, enabling diagnosis, fault-finding, and problem-solving. Notable examples include Stateflow [3] and preHAZOP [8]. While these systems automate routine tasks like identifying common causes and consequences, they are often industry-specific and struggle with adapting to new technologies or dynamic process changes or changing environments [4].

Ontologies, represented as directed acyclic graphs (DAGs) with nodes as concepts and edges as relationships. For instance, Yan et al. [9] applied named entity recognition (NER) to extract knowledge from HAZOP reports, organized it into an ontology knowledge base via Protégé and web ontology language (OWL), and employed the HermiT inference engine for automated analysis. This approach improved scenario coverage and safeguard descriptions, though its effectiveness depends on the ontology's scope and detail. Notably, the model focuses on text analysis and lacks automated risk assessment. Additionally, ontologies is time-consuming and biased as they rely on human's experience [4].

Bayesian networks (BNs) are acyclic graphs where nodes represent process variables and arcs depict cause–effect relationships. Each node has probabilistic states used to extrapolate the likelihood of other events. BNs have been applied in HAZOP to quantify risk and event probabilities, with experts assigning prior and conditional probabilities [10]. However, building BNs relies heavily on expert input for structure and prior probabilities, which can be subjective and challenging.

The model-based approaches use detailed mathematical models, combining the HAZOP technique with dynamic simulations to understand system behaviour during failures. This integration enhances hazardous scenario identification and risk assessment, reducing subjectivity in evaluating event severity and likelihood. Dynamic HAZOP simulation uses two main approaches: custom mathematical models tailored to specific units, offering full control over simulation methods or commercial process simulators, such as Aspen Plus, Aspen HYSYS, and k-Spice, which allow flexible adjustments to process flow sheets [11]. While these tools provide robust environments for integrating simulation with process hazard analysis, HAZOP studies rely on teams, and analysing new units requires developing and verifying new equations for each case.

Knowledge-based and model-based approaches have supported HAZOP studies, but acquiring and updating domain-specific knowledge remains challenging, subjective, and time-consuming. Recently, data-driven models have advanced HAZOP studies by leveraging historical data to improve safety outcomes and reduce manual effort [5]. Various classification models have been applied to automate specific aspects of HAZOP reports. For example, the BERT-BiLSTM-Attention model was used to predict severity from consequence descriptions, achieving a precision of 88.6% [12]. Bag of Words (BOW) was combined with ML to predict deviations based on causes, reaching 92% accuracy [13]. Peng et al. [14] proposed an ELMo-DCNN-BiLSTM-CRF model for NER to identify material and equipment terms, achieving higher recall for equipment (93.52% recall) than materials (86.08% recall). Term Frequency Inverse Document Frequency (TF-IDF) was applied with Naïve Bayes to predict likelihood, severity, and risk levels using causes, and consequences, with accuracies exceeding 80% [15]. Ekramipooya et al. [16] utilized Bidirectional Encoder Representations from Transformers (BERT) and multi-layer perceptron (MLP) to predict recommendations using two separate models, one using causes and the other using consequences.

These classification models face some limitations. HAZOP studies involve interconnected elements, such as process parameters, guidewords, causes, consequences, safeguards, and recommendations. Effective automation requires considering all these elements and addressing their interdependencies to avoid inaccuracies. Additionally, they lack prediction of safeguards and recommendations, based on these interconnected HAZOP elements.

Topic modelling methods have been developed to uncover latent topics in HAZOP reports. Wang et al. [15] used Latent Dirichlet Allocation (LDA) to uncover hidden cause and consequence from HAZOP data. TF-IDF, LDA, Part-of-Speech (POS) tagging, and the Apriori algorithm were used to explore correlations between causes and consequences, clustering them into 20 topics [17]. Despite its strengths, LDA does not consider the semantic relationships between words and order of words. BERT technique address this by generating contextual word and sentence vector representations. Ekramipooya et al. [16] leveraged BERT, Uniform Manifold Approximation and Projection (UMAP), and Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) to cluster recommendations as a first step for prediction models. However, they did not use topic representation algorithms, limiting the analysis of the generated topics.

The topic modelling literature has focused on identifying the main topics of causes and consequences for entire plants. However, HAZOP teams typically analyse nodes individually, necessitating the identification of these topics at the node level and their corresponding risk levels for effective risk assessment. Additionally, existing studies fail to categorize safeguards and recommendations, which are essential for guiding preventive measures. Moreover, previous research does not map the identified topics to their associated P&ID nodes, which is crucial since the division of P&IDs into nodes depends on the perspectives of HAZOP participants, based on process flow or main equipment. To address these gaps, this study leverages historical data to generate an intelligent HAZOP table that assists participants during workshops in developing the final report. In 2022, BERTopic, a state-of-the-art topic modelling algorithm built on pretrained sentence transformers, was introduced by Grootendorst [18]. It has shown superior performance in capturing semantic context across various domains. For its ability to provide a comprehensive semantic view and generate coherent topics representing different factors of causes, consequences, and countermeasures, BERTopic was chosen for this study to enhance the efficiency and accuracy of topic modelling and analysis.

3 Proposed Method

Figure 1 illustrates an intelligent analysis method, designed to enhance the efficiency and accuracy of conducting HAZOP studies and aid less experienced engineers by leveraging historical HAZOP data (reports and P&IDs' nodes), closely matching the plant under investigation. Using BERTopic, the method identifies the most frequent topics of root causes, consequences, and countermeasures for each HAZOP node and prioritizes them by likelihood, severity, and risk levels. These topics are also linked to their corresponding deviations. A multimodal model like Gemini [19] extracts detailed process and equipment descriptions from P&ID nodes, providing a deeper understanding of the process and associated risks. Mapping these topics to the P&ID nodes offers participants a clearer understanding of hazard identification and assessment for each node under investigation. By combining these insights, the method generates an intelligent, pre-populated HAZOP analysis table to guide participants during workshops, reducing time and manual effort for similar projects to produce the final HAZOP report. Due to space limitations, this paper will focus on categorizing causes into main risk factors, identifying the most frequent factors for each node and prioritizing them based on the likelihood of associated hazards.

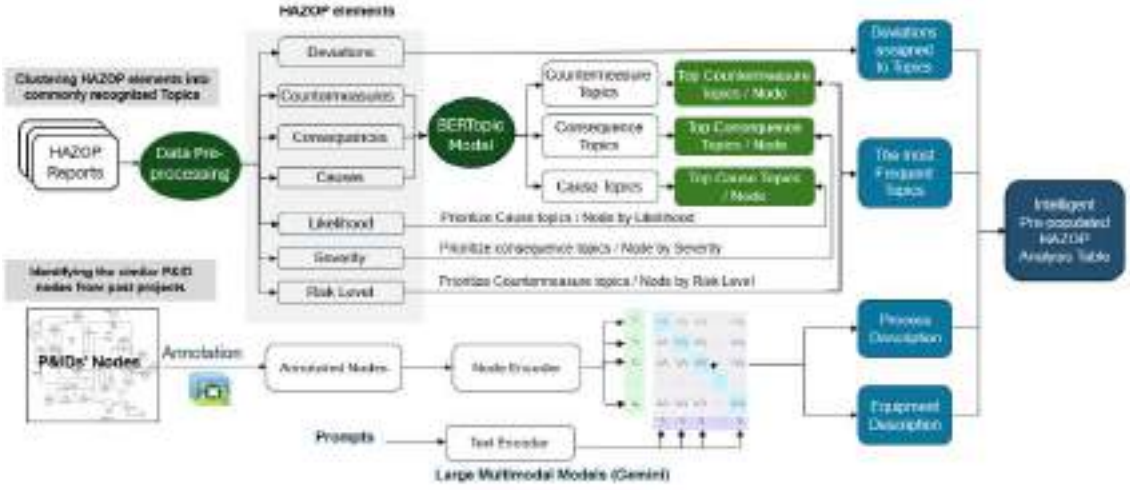


Figure 1. Flowchart of intelligent HAZOP analysis method using Topic Modeling.

The process starts with collecting HAZOP reports and P&IDs' nodes. HAZOP reports include deviations, causes, consequences, safeguards and recommendations, likelihood, severity, and risk. The dataset is shuffled to minimize order bias and enhance model generalization [20]. The preprocessing step involves tokenization, text cleaning, and normalization. Tokenization breaks text into smaller units (tokens) for granular analysis. Text cleaning removes irrelevant elements like stopwords, punctuation, and special characters while normalization uses techniques like lemmatization and lowercase conversion to standardize text [21].

Following preprocessing, the BERTopic model is applied as shown in Figure 2, including vector embedding, dimensionality reduction, clustering, tokenization, and topic representation. It begins with embedding extraction using Sentence-BERT (S-BERT) to generate dense -size embeddings that capture semantic meaning. These embeddings include token embeddings, segment embeddings, and position embeddings. To produce a single-size embedding for each sentence, S-BERT applies mean pooling across token embeddings, resulting in 768-dimensional vectors [22].

To optimize the HDBSCAN process, UMAP is employed for dimensionality reduction, which preserves both local and global structures of high-dimensional embeddings in a lower-dimensional space, facilitating efficient clustering [23]. Key hyperparameters, such as the number of neighbours, the number of components, and the distance metric (e.g., cosine similarity), are fine-tuned to maintain semantic relationships and improve clustering performance [24].

The datapoints are then tokenized and vectorized. Tokenization breaks the text into smaller units like words or phrases. CountVectorizer was used to transform these textual units into numerical vectors [25]. Topic representation is then refined using class-TF-IDF, that identifies distinguishing features between clusters. KeyBERT further enhances topic representation by selecting the most relevant keywords based on their cosine similarity to cluster embeddings [25].

Topic quality is evaluated using coherence and diversity metrics. The coherence score (C_V) assesses the semantic relatedness of topic words, with higher scores (e.g., 0.6–1.0) indicating well-defined and interpretable topics [26]. The Coherence score is calculated as follows:

$$C_(T) = 2 / (|T|(|T| - 1)) * \Sigma_\{i=1\}^\{|T|\} \Sigma_\{j \neq i\} sim(Word_i, Word_j) \quad (1)$$

The coherence score (C_V (T)) for a given topic T is calculated based on the number of words in the topic (|T|), the representation of two distinct words (Wordi and Wordj) within that topic, and the similarity measure (sim (Wordi, Wordj)) between these word pairs.

The diversity score measures the uniqueness of topic words across clusters, with values closer to 1 reflecting minimal redundancy. While metrics provide quantitative insights, human evaluation remains essential for validating the interpretability of topics [27].

To enhance topic exploration and reduce outliers, Zero-shot BERTopic uses predefined topics and cosine similarity thresholds to guide clustering. By adjusting the similarity threshold, the model flexibly assigns documents to predefined topics or generates new topics as needed [27]. This iterative process, paired with human judgment, ensures the final topics are actionable.

The model was evaluated against the LDA model, a probabilistic method for uncovering latent topics in text. LDA uses a document-term matrix, treating each document as a mixture of topics and attributing words to these topics based on co-occurrence patterns [28]. TF-IDF was used to enhance feature representation, as it weights terms based on their frequency within documents and importance across the corpus. The resulting TF-IDF matrix serves as input for LDA, enabling it to analyse term significance and distribution to uncover topics.

The models were implemented on Google Colab (CPU runtime), with Key libraries including NLTK and SpaCy for preprocessing, PyTorch, BERTopic, and scikit-learn for topic modelling, and Gensim for evaluation metrics. Core Python packages like NumPy, pandas, Matplotlib, and Seaborn supported data manipulation and visualization.



Figure 2. Methodology to apply BERTopic Model on HAZOP Causes.

4 Case Study and Results

The data utilized in this study comprised node name, causes, and likelihood derived from HAZOP reports focused on oil pump stations, which play a vital role in transporting oil through pipelines between various locations. The dataset was sourced from the thesis of Kilincet al. [29] report of an oil pump station located in Sivas, Turkey, designed to ensure safe and efficient transfer of oil via pipelines. At the station, oil enters from the pipeline and first reaches the pig receiving system. It then passes through the main pump filters before reaching the pumps, where it is energized and sent through the pig launching section back into the pipeline toward its next destination. Additionally, the wax handling system, used to remove wax (solid oil), represents one of the most challenging processes in the operation. The dataset encompasses six nodes and 1,574 data points required for safe oil transfer within the plant.

The BERTopic model was fine-tuned by experimenting with various hyperparameters to optimize topic generation. Key parameters included min_topic_size (20–50) to control the minimum documents per topic, affecting topic granularity; top_n_words (15) to emphasize the most representative terms for interpretability; n_neighbors (10–15) in UMAP for balancing local and global data structures during dimensionality reduction; n_components (3–5) to set the dimensionality of reduced embeddings. These configurations resulted in topics ranging from 6 to 43. Figure **3** illustrates topic coherence and diversity scores across multiple experiments, providing a quantitative assessment of topic quality. Coherence scores peaked at 32 topics, achieved 0.898, while diversity score was highest at 9 topics, with 0.967. Variations in the number of top words had no impact on the scores.

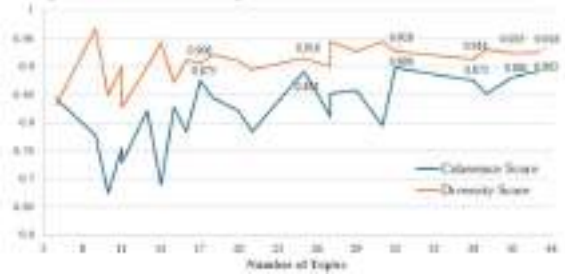


Figure 3. Coherence and Topic Diversity Scores.

As highlighted earlier, the resulting topics should be evaluated through human judgment [30]. Consequently, the topics and their top words were manually reviewed, focusing on experiments with higher coherence and topic diversity scores, such as those with 17, 25, 32, 38, 41, and 43 topics. The evaluation process was guided by the main causes of hazards identified in oil pump stations and pipelines in the literature, including operator error, inadequate pipe management, corrosion, insufficient maintenance, ignored alarms, fire, explosions, and equipment issues [31], [32]. This review assessed whether each cluster contained distinct topics and how well-separated these topics were, or if overlapping top words indicated a lack of distinction among them.

Following this review, 41 topics were initially selected because they exhibited less overlap. However, further analysis revealed several groups of similar topics that required merging. Topics 1, 2, 3, 12, 19, and 21 were found to be closely related and grouped together, as were topics 0, 11, 22, 24, 31, and 33, topics 8, 9, 10, 16, 18, 34, 35, and 36, topics 4, 29, 30, and 39, topics 7, 17, and 27, topics 6, and 14, topics 20, 23, and 28, topics 26, and 38, and topics 13 and 25. These similar topics were merged, and additional experiments were conducted to optimize the coherence and topic diversity scores, resulting in 13 topics, achieving coherence and diversity scores of 80.0% and 92.4%, respectively. Table 1 presents the final topics, including the top frequent words (1-3 grams) associated with each topic and their corresponding suggested labels. Cause topic generation is based on topic words and manual induction. Thus, the topic's formation needs knowledge, experience, and analysis. This approach ensured a robust and interpretable topic modelling outcome while maintaining high performance metrics.

It is worth noting that this model resulted in 71 outliers, stemming from several factors. Some data points contain very short descriptions and general statements, providing insufficient semantic information for clustering. Additionally, the dataset exhibited high variability in text length across data points, further contributing to the outliers. Zero-shot BERTopic was applied to explore potential new topics for documents that did not align with predefined topics. Experiments were conducted with cosine similarity values ranging from 0.5 to 0.95. Values between 0.5 and 0.8 did not yield new topics, while values above 0.8 introduced new topics that closely resembled the predefined ones. Based on these findings, a cosine similarity value of 0.8 was selected, which effectively reduced the number of outliers from 71 to just 32 documents. The frequency of topics is presented in Table 1.

Table 1. The Cause topics and their corresponding labels and Frequency.

Topic	Top Words	Category	Frequency
0	valve, operate valve, motor operate valve, operate, error, close omit, main oil line, completely omit, oil line pump	Omitted or wrong operations	43
1	slop tank, slop tank pump, failure, reinjection, reinjection pump, switch, interlock, flow, block, slop tank block	Pump failure / malfunctions.	200
2	valve, pass, open, valve skid, skid, check valve, check, open pass, valve fail open, safety valve fail, fail open pass	Valve errors or failures.	754

3	maintenance, closure, close maintenance case, blind leave close, maintenance case block, leave close maintenance	Inadequate maintenance	58
4	emergency, emergency shutdown, error false signal, alarm may fail, shutdown valve, error, level, faulty level alarm	Safety system failure	60
5	multiple, multiple simultaneous, multiple drain operation, multiple drain, simultaneous, simultaneous thermal	Simultaneous operations	77
6	roof, complete, complete loss, gauge lead unintended, surge valve gauge, valve gauge, valve gauge lead, valve	Relief devices error/ failure	35
7	insulation, corrosion, corrosion insulation, insulation cui, corrosion insulation cui, cui, cui tank wall, insulation low	Corrosion	87
8	external fire, fire external, external fire external, fire external fire, fire, external	External fire	110
9	underway leak heating, also source, heating coil may, leak heating, leak heating coil, event underway leak,	Leakage	39
10	mode, mistakenly put, put back, put, pump leave manual, tank pump leave, mode mistakenly put, manual mode	Human errors	5
11	pipe, bore, pipe rupture within, pipe rupture, rupture, large bore pipe, large bore, large, bore pipe rupture	Pipe failure or error.	50
12	transmitter fail, transmitter, failure level, failure level transmitter, level transmitter, level transmitter fail	Transmitter failure /error	24

The identified topics were validated against the 11 main causes identified manually by Kilincet al. [29], including omitted operations, fires, inadequate maintenance, operational failures, out-of-range, wrong equipment, incorrect operations, misplaced equipment, equipment failures, safety system failures, and simultaneous operations. The validation confirmed a strong alignment with traditional methods, while the model also provided more granular insights, such as pump malfunctions, pipe rapture, valve failures (the most frequent category), and relief device failures, which align with categories like equipment failure or wrong equipment. Additionally, the model revealed extra causes, such as human errors, leakage, and corrosion, expanding the depth and detail of the traditional HAZOP analysis. Identifying main causes for the entire plant provides valuable insights, but the HAZOP team evaluates nodes individually. Therefore, it is crucial to pinpoint the key causes for each node and determine which risk factors are included, excluded, and prioritized. Figure 4 illustrates the topics and their frequency for each node, highlighting the most critical causes. For instance, in Node 1 (Pig Receiving), the primary causes include valve failures, pipe ruptures, and omitted operations. Notably, Node 1 accounts for 11 out of the 13 identified causes, excluding pump malfunctions and relief device failures.

In addition to those identified main causes for each node, prioritizing them based on likelihood (e.g., 1 to 8, where 1 indicates an event has not yet occurred in the industry and 8 indicates frequent occurrence at the facility) ensures that critical issues are addressed first. High-frequency topics within each likelihood category highlight recurring or critical hazards, guiding the implementation of effective safeguards.

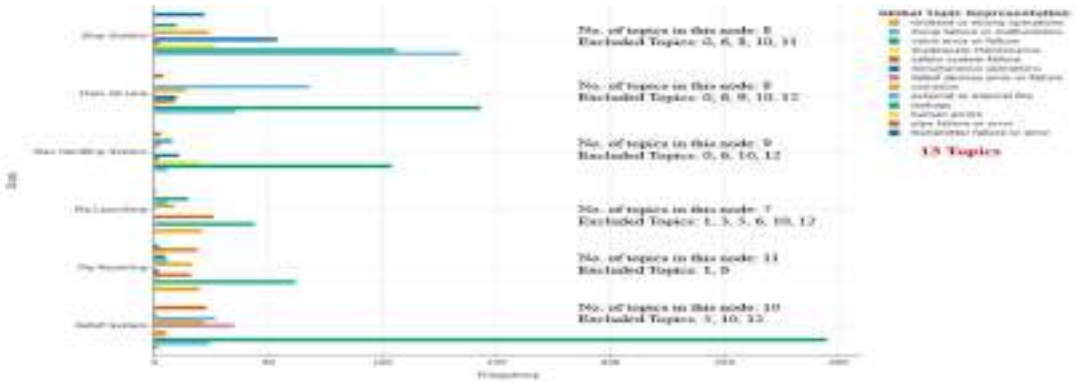


Figure 4. The included and excluded topics associated with each P&ID's node

For instance, in node 2 (Pig Launching), as shown in Figure 5**,** the frequency distribution of topics across probability categories "2," "3," "4," "5," "6," and "7" identifies specific areas requiring mitigation strategies. Topic 9 (leakage) has the highest frequency in category "7," indicating a high probability of occurrence and making it a priority for risk assessment. In contrast, Topic 0 (omitted or wrong operations), Topic 2 (valve errors or failures), and Topic 4 (safety systems failure) are the most frequent topics in categories "4," "3," and "2," respectively. Although these categories represent lower likelihoods, their recurrence across multiple deviations within the same node could cumulatively escalate the overall risk. This emphasizes the need to evaluate whether the frequency of such hazards justifies earlier intervention to prevent escalating if left unaddressed.

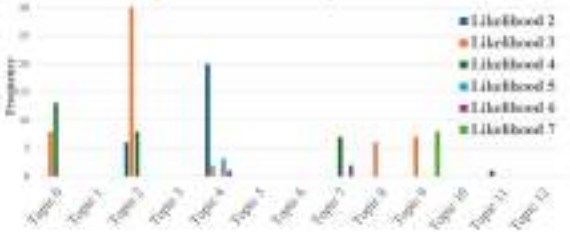


Figure 5. Categorization of topics by likelihood for Node 2 (Pig Launching).

The BERTopic model was compared to the LDA model to evaluate its effectiveness. LDA was selected as it has been extensively applied in HAZOP studies. Since LDA requires a predefined number of topics, a range from 6 to 43 topics (matching BERTopic's output) was tested. Various hyperparameters were explored, including passes (10–100), iterations (1000–5000), and alpha and beta values (0.01). Multiple experiments were conducted to identify the optimal number of topics (K), determined using perplexity and coherence metrics, as smaller perplexity and higher coherence typically indicate better topic recognition [28]. Figure 6 shows perplexity was lowest at 42 topics, while coherence peaked at 12 topics with a score of 45%. However, low and fluctuating coherence scores indicated LDA's challenges in producing stable and interpretable results. Topic diversity was highest (89%) at 8 topics. A detailed review of 42, 12, 8, and 41 topics revealed overlapping clusters, complicating the selection of optimal topics.

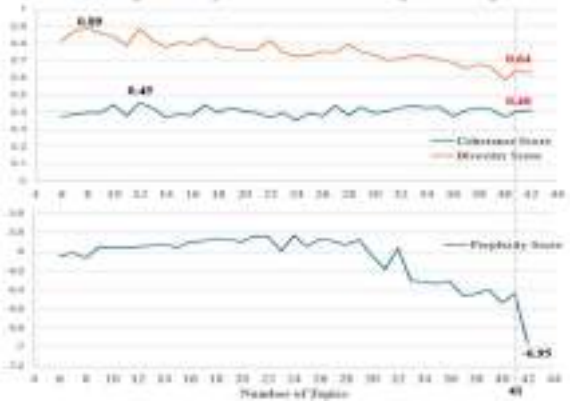


Figure 6. LDA Model Evaluation Metrices.

5 Conclusion

This paper briefly reviewed current developments of intelligent systems in HAZOP studies, emphasizing the potential of utilizing historical HAZOP data to streamline the process. The review reveals several gaps including inadequate incorporation of all HAZOP elements in classification models, prediction of countermeasures based on these interconnected elements, and categorization of causes, consequences, and countermeasures by node, likelihood, severity, and risk levels. This study applied the BERTopic model to categorize causes into 15 risk factors, achieving a coherence score of 80% and a topic diversity score of 92.4%, which outperforms LDA's 45.4% coherence and 88.8% diversity. The model identified key topics by node and likelihood. This approach can be extended to consequences and countermeasures, generating a prepopulated HAZOP table to assist workshop participants in verifying or identifying additional issues.

The main limitation of this study is its focus on oil pump stations, using data from a single HAZOP report, which may affect the model's generalizability. Expanding the dataset is recommended to improve model performance. Moreover, hyperparameter tuning for BERTopic was time-consuming, requiring multiple trials for high coherence and diversity. Optimization techniques are suggested to streamline this process.

References

[1] A. Jensen and T. Aven, "Hazard/threat identification: Using functional resonance analysis method in conjunction with the Anticipatory Failure Determination method," *Proc Inst Mech Eng O J Risk Reliab*, vol. 231, no. 4, pp. 383–389, Aug. 2017, doi: 10.1177/1748006X17698067.

[2] F. Joubert, E. Steyn, and L. Pretorius, "Using the HAZOP Method to Conduct a Risk Assessment on the Dismantling of Large Industrial Machines and Associated Structures: Case Study," *J Constr Eng Manag*, vol. 147, no. 1, Jan. 2021, doi: 10.1061/(asce)co.1943-7862.0001942.

[3] M. F. Chia and P. K. Naraharisetti, "HAZOP using Stateflow software: Methodology and case study," *Process Safety and Environmental Protection*, vol. 179, pp. 137–156, Nov. 2023, doi: 10.1016/j.psep.2023.09.005.

[4] J. Single, J. Schmidt, and J. Denecke, "Computer-aided hazop studies: Knowledge representation and algorithmic hazard identification," in *WIT Transactions on the Built Environment*, WITPress, 2019, pp. 55–66. doi: 10.2495/SAFE190061.

[5] E. Elhosary and O. Moselhi, "Automation for HAZOP study: A state-of-the-art review and future research directions," *Journal of Information Technology in Construction*, vol. 29, pp. 750–777, Sep. 2024, doi: 10.36680/j.itcon.2024.033.

[6] H. J. Pasman and W. J. Rogers, "How can we improve HAZOP, our old work horse, and do more with its results? an overview of recent developments," *Chem Eng Trans*, vol. 48, pp. 829–834, 2016, doi: 10.3303/CET1648139.

[7] J. I. Single, J. Schmidt, and J. Denecke, "Ontology-based support for hazard and operability studies," *International Journal of Safety and Security Engineering*, vol. 10, no. 3, pp. 311–319, Jun. 2020, doi: 10.18280/ijsse.100302.

[8] J. Oeing, T. Holtermann, W. Welscher, C. Severins, M. Vogel, and N. Kockmann, "preHAZOP: Graph-Based Safety Analysis for Early Integration into Automated Engineering Workflows," *Chem Ing Tech*, vol. 95, no. 7, pp. 1083–1095, Jul. 2023, doi: 10.1002/cite.202200222.

[9] H. Yan, S. Wu, L. Yuan, S. Wang, and Y. Cao, "A New HAZOP Automated Method for the Oil and Gas Complex Equipment," *Chemistry and Technology of Fuels and Oils*, vol. 59, no. 4, pp. 872–879, Sep. 2023, doi: 10.1007/s10553-023-01592-8.

[10] P. Gao and W. Li, "Integration of HAZOP and Bayesian network in city gas explosion emergency response processes," *Emergency Management Science*

and Technology, vol. 2, no. 1, pp. 1–9, 2022, doi: 10.48130/emst-2022-0019.

[11] J. Janošovský, M. Danko, J. Labovský, and Ľ. Jelemenský, "Software approach to simulation-based hazard identification of complex industrial processes," *Comput Chem Eng*, vol. 122, pp. 66–79, Mar. 2019, doi: 10.1016/j.compchemeng.2018.05.021.

[12] X. Feng, Y. Dai, X. Ji, L. Zhou, and Y. Dang, "Application of natural language processing in HAZOP reports," *Process Safety and Environmental Protection*, vol. 155, pp. 41–48, Nov. 2021, doi: 10.1016/j.psep.2021.09.001.

[13] A. Ekramipooya, M. Boroushaki, and D. Rashtchian, "Application of natural language processing and machine learning in prediction of deviations in the HAZOP study worksheet: A comparison of classifiers," *Process Safety and Environmental Protection*, vol. 176, pp. 65–73, Aug. 2023, doi: 10.1016/j.psep.2023.06.004.

[14] L. Peng, D. Gao, and Y. Bai, "A study on standardization of security evaluation information for chemical processes based on deep learning," *Processes*, vol. 9, no. 5, 2021, doi: 10.3390/PR9050832.

[15] F. Wang and W. Gu, "Intelligent HAZOP analysis method based on data mining," *J Loss Prev Process Ind*, vol. 80, Dec. 2022, doi: 10.1016/j.jlp.2022.104911.

[16] A. Ekramipooya, M. Boroushaki, and D. Rashtchian, "Predicting possible recommendations related to causes and consequences in the HAZOP study worksheet using natural language processing and machine learning: BERT, clustering, and classification," *J Loss Prev Process Ind*, vol. 89, Jul. 2024, doi: 10.1016/j.jlp.2024.105310.

[17] F. Wang, W. Gu, Y. Bai, and J. Bian, "A method for assisting the accident consequence prediction and cause investigation in petrochemical industries based on natural language processing technology," *J Loss Prev Process Ind*, vol. 83, Jul. 2023, doi: 10.1016/j.jlp.2023.105028.

[18] M. Grootendorst, "BERTopic: Neural topic modeling with a class-based TF-IDF procedure," Mar. 2022, [Online]. Available: http://arxiv.org/abs/2203.05794

[19] M. Imran and N. Almusharraf, "Google Gemini as a next generation AI educational tool: a review of emerging educational technology," *Smart Learning Environments*, vol. 11, no. 1, p. 22, May 2024, doi: 10.1186/s40561-024-00310-z.

[20] B. Li, Y. Esfandiari, M. N. Schmidt, T. S. Alstrøm, and S. U. Stich, "Synthetic data shuffling accelerates the convergence of federated learning under data heterogeneity," *Transactions on Machine Learning Research*, Apr. 2024, doi: https://doi.org/10.48550/arXiv.2306.13263.

[21] G. Liu, M. Boyd, M. Yu, S. Z. Halim, and N. Quddus, "Identifying causality and contributory factors of pipeline incidents by employing natural language processing and text mining techniques," *Process Safety and Environmental Protection*, vol. 152, pp. 37–46, Aug. 2021, doi: 10.1016/j.psep.2021.05.036.

[22] E. Aytaç and M. Khayet, "A Topic Modeling Approach to Discover the Global and Local Subjects in Membrane Distillation Separation Process," *Separations*, vol. 10, no. 9, p. 482, Sep. 2023, doi: 10.3390/separations10090482.

[23] L. McInnes, J. Healy, N. Saul, and L. Großberger, "UMAP: Uniform Manifold Approximation and Projection," *J Open Source Softw*, vol. 3, no. 29, p. 861, Sep. 2018, doi: 10.21105/joss.00861.

[24] R. Zhou, W. Gao, D. Ding, and W. Liu, "Supervised dimensionality reduction technology of generalized discriminant component analysis and its kernelization forms," *Pattern Recognit*, vol. 124, p. 108450, Apr. 2022, doi: 10.1016/j.patcog.2021.108450.

[25] M. Khayet, E. Aytaç, and T. Matsuura, "Bibliometric and sentiment analysis with machine learning on the scientific contribution of Professor Srinivasa Sourirajan," *Desalination*, vol. 543, p. 116095, Dec. 2022, doi: 10.1016/j.desal.2022.116095.

[26] D. Bretsko, A. Belyi, and S. Sobolevsky, "Comparative Analysis of Community Detection and Transformer-Based Approaches for Topic Clustering of Scientific Papers," in *Computational Science and Its Applications – ICCSA 2023: 23rd International Conference, Proceedings, Part I*, Athens, Greece, Jul. 2023, pp. 648–660. doi: 10.1007/978-3-031-36805-9_42.

[27] J. Kim and S. Lee, "Technology Opportunity Analysis for Creating Innovative Solutions: Applying Semi-supervised Topic Modelling on Patent Data," in *2024 PICMET conference*, IEEE, Aug. 2024, pp. 1–9. doi: 10.23919/PICMET64035.2024.10653159.

[28] H. Xu *et al.*, "Cause analysis of hot work accidents based on text mining and deep learning," *J Loss Prev Process Ind*, vol. 76, p. 104747, May 2022, doi: 10.1016/j.jlp.2022.104747.

[29] M. O. , Kilinc, G. A. , Ciftcioglu, and M. N. , Kadirgan, "SIS Application at a Petroleum Crude Oil Pipeline Pump Station After HAZOP Study," MARMARA University, ISTANBUL, 2020.

[30] C. Doogan and W. Buntine, "Topic Model or Topic Twaddle? Re-evaluating Semantic Interpretability Measures," in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2021, pp. 3824–3848. doi: 10.18653/v1/2021.naacl-main.300.

[31] S. Sendy and F. T. Basaria, "A Comparative Analysis of Hazard Analysis Methods for Sustainable Energy Development," *E3S Web of Conferences*, vol. 388, p. 01037, May 2023, doi: 10.1051/e3sconf/202338801037.

[32] A. Waqar, I. Othman, N. Shafiq, and M. S. Mansoor, "Evaluating the critical safety factors causing accidents in downstream oil and gas construction projects in Malaysia," *Ain Shams Engineering Journal*, vol. 15, no. 1, p. 102300, Jan. 2024, doi: 10.1016/j.asej.2023.102300.

A Case Study on Applying Machine Learning on Blockchain (MLOB) for Modular Integrated Construction Inspections

Zhiming Dong[1], Weisheng Lu[1]

[1]Department of Real Estate and Construction, Faculty of Architecture, University of Hong Kong, Hong Kong
zmdong@connect.hku.hk, wilsonlu@hku.hk

Abstract

Modular Integrated Construction (MiC) is an innovative off-site building method that improves productivity, safety, and sustainability by manufacturing standardized modules in factories for on-site assembly. MiC inspections demand high levels of accuracy, efficiency, and security due to the low fault tolerance in factory environments. However, existing approaches fail to comprehensively address these combined requirements. This research explores the application of the Machine Learning on Blockchain (MLOB) framework to real-world MiC inspection tasks, assessing its feasibility and effectiveness. Additionally, a scientific evaluation is conducted to quantitatively measure the performance of MLOB-based MiC inspections. The results demonstrate that MLOB significantly enhances inspection performance by improving decision-making, streamlining processes, and ensuring the integrity of both inspection logic and data. To the best of the authors' knowledge, this is the first study to integrate MLOB into MiC inspections to consider the requirement on both accuracy, efficiency and security, providing a scalable and innovative solution for quality assurance in the construction industry.

Keywords

Machine Learning; Blockchain; Modular Integrated Construction; Inspection

1 Introduction

Modular Integrated Construction (MiC) is an innovative building method that transitions most construction activities from traditional on-site settings to controlled, off-site factories [28]. This method involves the production of standardized, 3D volumetric modules, which are manufactured and preassembled in factories before being transported for on-site assembly [8]. MiC offers numerous advantages, including improved productivity, enhanced safety, reduced waste, and lower carbon emissions. It follows a manufacturing plan that prioritizes precision and standardization, with minimal tolerance for errors in the factory environment [11]. However, any defects or inaccuracies during production can lead to severe complications during on-site assembly, resulting in costly rework and project delays [26]. Consequently, stringent quality control measures are essential to ensure the success of MiC projects.

Recent research underscores the potential of advanced technologies such as machine learning (ML) and blockchain to enhance inspection processes in MiC [22,23]. ML significantly improves inspection accuracy by analyzing large datasets to precisely detect defects, deviations, or anomalies, thereby minimizing errors associated with human oversight. Additionally, ML enhances efficiency by automating repetitive tasks, streamlining the inspection process, and enabling real-time decision-making without continuous human involvement [2]. Meanwhile, blockchain technology bolsters inspection processes by ensuring data integrity, transparency, and security [12]. By creating an immutable record of inspection data, blockchain enhances accountability and fosters trust among stakeholders.

While ML and blockchain independently enhance inspection processes, relying on these technologies in isolation reveals notable limitations. ML lacks robust security and transparency mechanisms, leaving it vulnerable to data tampering and manipulation [16]. Conversely, blockchain ensures data integrity and accountability but does not inherently offer intelligent data analysis or adaptive decision-making capabilities [4]. As a result, current approaches fail to comprehensively address the combined requirements of accuracy, efficiency, and security in MiC inspections.

Machine Learning on Blockchain (MLOB), as a novel computation paradigm, integrates machine learning (ML) with blockchain technology to ensure data and computational security [4]. Unlike conventional ML and blockchain isolated solutions [15,21,29,31], MLOB embeds ML logic directly on-chain, executing computations as smart contracts while preserving execution records on a distributed ledger. This framework offers distinct advantages to the Architecture, Engineering, and Construction industry. From a technical perspective, it mitigates risks associated with external

interference, unauthorized alterations, and adversarial manipulations, ensuring the reliability of critical computations such as project progress monitoring, quality assurance, and risk assessment. Practically, its robust security mechanisms foster greater confidence in adopting advanced ML tools, leading to more precise, efficient, and secure project management.

This research adopts the MLOB paradigm for MiC inspection tasks. The primary objective is to leverage MLOB to significantly enhance the efficiency, accuracy, and security of MiC inspection workflows while assessing the feasibility of its application in real-world construction scenarios. To the best of the authors' knowledge, this study is the first to integrate the MLOB paradigm into MiC inspection tasks.

The structure of the paper is organized as follows: Section 1 introduces the research background and objectives. Section 2 provides a literature review, discussing current research on ML for MiC inspection, blockchain technology for MiC inspection, and the development of the MLOB framework. Section 3 defines the case, offering background and essential details of the MiC project. Section 4 describes the prototype implementation, outlining the four key steps of the MLOB process: ML acquisition, ML conversion, ML safe loading, and consensus-based computing. Section 5 presents the experimental results, while Section 6 discusses the contributions, limitations, and directions for future research. Finally, Section 7 concludes the study.

2 Literature review

2.1 Machine learning for MiC inspection

Liu, et al. [9] evaluates ML object detection algorithms for modular construction, demonstrating faster RCNN's superior accuracy and offering insights into ML's applicability and workflow for improving construction quality and safety. Tusnin, et al. [23] proposes adjustable ML models to efficiently analyze and design modular steel building blocks, achieving accurate stress–strain state predictions with a determination coefficient of 0.88–0.92, overcoming limitations of traditional numerical and engineering methods. Tan, et al. [22] proposes an automated BIM and LiDAR-based inspection method for modular box outer walls and rebars, achieving high accuracy with errors of 2.6 mm for flatness and 2.8 mm for protruding length measurements. Baek, et al. [2] presents an automated framework for monitoring on-site MiC module installation, combining object detection, activity classification, and productivity analysis, achieving 89% accuracy and enabling productivity measurement and improvement across diverse construction sites. Pan, et al. [14] explores future directions for applying artificial intelligence and robotics (AIR) in prefabricated and modular construction through a systematic literature review, proposing five key advancements to enhance interdisciplinary, integrated, and intelligent construction practices. Harichandran, et al. [5] proposes a hierarchical ML framework to enhance activity recognition in Automated Construction Systems (ACS), demonstrating improved precision, robustness, and noise tolerance compared to conventional approaches for monitoring and ensuring safe construction operations.

Moreover, numerous studies have explored the application of ML in inspection-related tasks. [19] applies ML to risk-based inspection (RBI) screening, achieving 92.33% accuracy and 84.58% precision, reducing variability, and improving the reliability and integrity of conventional RBI screening assessments. Ravikumar, et al. [20] explores ML-based visual inspection for defect identification, comparing C4.5 and Naïve Bayes classifiers, with results showing C4.5 achieving superior classification accuracy for automated machine vision system design. Park, et al. [17] proposes a CNN-based method for automated visual inspection of surface defects, demonstrating superior accuracy, efficiency, and cost-effectiveness compared to manual inspection, with applicability to both textured and non-textured surfaces.

The literature demonstrates the versatility and effectiveness of ML in construction and inspection tasks, offering enhanced accuracy, efficiency, and applicability across diverse contexts, from modular construction quality assessments to automated defect detection. However, ML lacks robust security and transparency mechanisms, leaving it vulnerable to data tampering and manipulation.

2.2 Blockchain technology for MiC inspection

Li, et al. [8] proposes a blockchain-enabled IoT-BIM platform (BIBP) for modular construction, achieving decentralized, secure, and efficient supply chain management, validated through a case study demonstrating improved transparency, traceability, and performance in storage, throughput, and latency. Olawumi, et al. [13] critically reviews digital tools and technologies (DTT) in modular integrated construction (MiC), identifying gaps in prefab transportation, highlighting leading countries, and proposing future opportunities for blockchain and integrated DTT adoption in MiC projects. Wu, et al. [26] develops a blockchain-enabled IoT-BIM platform (BIBP) for modular construction, addressing IoT network failures and BIM provenance issues, demonstrating improved reliability, reduced storage costs, and advancing practical applications in off-site production management. Wu, et al. [27] explores blockchain technology to enhance information-sharing accuracy in onsite modular

construction assembly, developing and validating a prototype system that improves information endorsement and synchronization through blockchain's decentralized consensus mechanism.

Likewise, blockchain technology is widely applied in other inspection tasks. Wu, et al. [25] proposes a blockchain-based framework for on-site construction quality inspection (OCQI), integrating IoT and ontology technologies to enhance accountability, automation, and data integrity, validated through a prototype system and analyzed using the TOE theory. Pincheira, et al. [18] presents an end-to-end system for mining asset inspections, integrating IoT devices with blockchain to enhance data integrity, trust, and transparency, validated through a real case study in a mining inspection scenario. Wu, et al. [24] develops a blockchain-enabled framework for tower crane safety inspections, integrating smart contracts and consensus algorithms to ensure automation, transparency, and traceability, validated through a prototype system using Hyperledger Fabric architecture. Ning, et al. [12] proposes a blockchain-enabled secure transmission system (BEST) and improved market matching (IMM) algorithm for remote grid fault inspections, enhancing network performance and ensuring secure, efficient signal transmission using advanced optimization and deep reinforcement learning techniques. Zhang, et al. [30] proposes a blockchain-based performance-security balanced safety inspection framework (PSB-SIF), integrating smart contracts and a novel SIBFT consensus algorithm, ensuring authenticity, security, and efficiency in safety inspections with improved performance over conventional consensus methods.

The literature highlights the significant role of blockchain in MiC and inspection tasks. Blockchain-enabled IoT-BIM platforms improve supply chain transparency, traceability, and reliability, while blockchain-based frameworks enhance accountability, automation, and security in diverse inspection scenarios. However, blockchain ensures data integrity and accountability but does not inherently offer intelligent data analysis or adaptive decision-making capabilities.

2.3 Machine learning on blockchain

There are some existing research has explored the integration of blockchain and ML. Liu, et al. [10] surveys the integration of blockchain and ML for communications and networking systems, highlighting benefits, applications, and challenges, while addressing issues like resource management, data processing, scalability, and security for decentralized and intelligent network operations. Chen, et al. [3] reviews the integration of ML and blockchain, highlighting their collaborative potential for secure and efficient data sharing and analysis, and identifies future directions for deeper integration of these transformative technologies. Kayikci and Khoshgoftaar [7] explores the integration of blockchain and ML, highlighting their transformative potential in finance, medicine, supply chain, and security while addressing challenges like scalability, security, and workflow optimization for broader adoption and impactful applications. Acheampong [1] explores the interplay of big data, ML, and blockchain technology, highlighting big data's impact on ML advancements while proposing blockchain integration to enhance data security, reduce centralization challenges, and encourage further research for improved ML models.

In the context of MiC inspection, ensuring robust data security is imperative because the process involves collecting sensitive sensor and imaging data from multiple sources. Equally, computation security is crucial since the analytical algorithms and ML used for processing this data are vulnerable to adversarial attacks that could distort outputs and lead to erroneous decision-making. Addressing both aspects concurrently ensures that not only is the authenticity of the raw inspection data maintained, but also that the computational processes interpreting this data remain reliable and resistant to cyber threats. This dual security requirement is well demonstrated in the MLOB framework proposed by Dong and Lu [4], which is designed to safeguard both data and computation, thereby providing a comprehensive solution to the unique challenges of MiC inspection.

3 Methodology

3.1 The MLOB framework

The MLOB is a novel computational paradigm that integrates ML with blockchain technology to secure engineering computing processes [4]. The primary aim of MLOB is to safeguard both data and computational security by embedding ML processes entirely within blockchain smart contracts. The workflow of MLOB is shown in Figure 1. Methodologically, the MLOB paradigm encompasses four sequential steps: ML Acquisition, ML Conversion, ML Safe Loading, and Consensus-Based Computing Execution. Each stage systematically migrates traditional ML workflows into a secure blockchain-enabled environment, ensuring that all processes remain tamper-proof, transparent, and accountable.

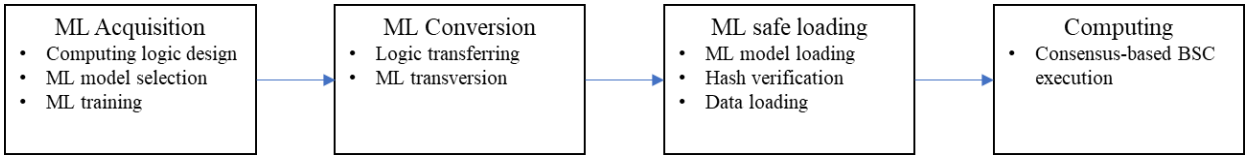


Figure 1. The MLOB framework.

The ML Acquisition phase initiates the methodology by defining and formulating engineering computing logic into ML-compatible tasks, followed by selecting appropriate ML models tailored to specific engineering scenarios and training them with relevant datasets. After acquisition, the trained ML models are converted into blockchain-compatible formats during the ML Conversion phase, ensuring compatibility with blockchain smart contract runtime environments. Subsequently, the ML Safe Loading step securely transfers both the computational models and the required input data onto the blockchain, employing cryptographic hashing to verify integrity and prevent unauthorized modifications during off-chain to on-chain transition.

In the final stage, Consensus-Based Computing Execution, ML computations are executed directly within blockchain smart contracts across multiple decentralized nodes. The transactions—including computational instructions, inputs, and outputs—are collectively validated by consensus algorithms across the blockchain network, significantly enhancing resilience against malicious manipulations. Any inconsistencies or tampering attempts are swiftly identified, preserving computational security. Through these methodological steps, the MLOB paradigm establishes a robust framework for secure ML deployment, effectively addressing both data and computational integrity in engineering applications.

3.2 Case description

The research is set in the empirical context of a MiC

Given the project's advanced and complex nature, it faces challenges such as cross-border transportation, multi-stakeholder collaboration, and stringent quality control. Consequently, the project imposes rigorous demands on inspection processes: accuracy ensures all measurements and installations align precisely with project standards; efficiency focuses on timely and streamlined inspections to prevent delays in prefabrication and on-site assembly; and security safeguards against data tampering and logic corruption. Therefore, adopting MLOB in this research case is essential to enhance the accuracy, efficiency, and security of inspection processes.

Building on insights from previous projects, this research aims to apply the advanced MLOB framework

project in Hong Kong, specifically a student residential development. Selected as one of the Pilot MiC Projects by the Development Bureau of Hong Kong, the project comprises two 17-floor towers, providing 1,224 hostel places, with each room measuring approximately 6.5m². The project involves the assembly of 952 modules across six types. The podium transfer plate and core walls will be constructed using in-situ reinforced concrete, while MiC techniques will be employed for the floors above the transfer plate and the multi-unit bedrooms integrated with core walls. Each room will feature prefabricated polished flooring, anti-mould painted walls, and ceilings, all completed in the factory along with the installation of fixed furniture modules.

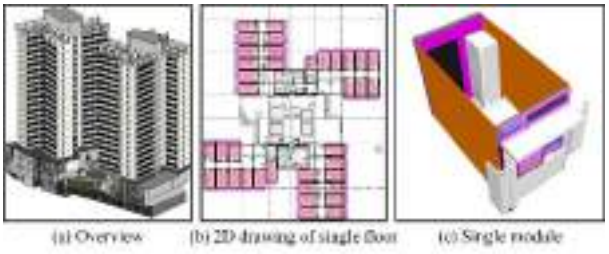


Figure 2. Background information of case study.

Figure 2 highlights the project's design details through both 3D models and 2D drawings. Specifically, Figure 2(a) provides an overview of the 3D model, while Figure 2(b) illustrates a 2D drawing of a single floor. In the 2D drawing, the pink segments indicate the key modules, which serve as focal points for inspection activities (Figure 2(c)).

to enhance construction logistics and supply chain management [11]. A key component of the project management process involves image-based online inspections, which play a crucial role in ensuring quality and compliance. During these inspections, workers are required to upload images of installed structures or facilities, accompanied by a scale bar for dimensional reference. Inspectors then review the submitted images to verify adherence to inspection standards, with particular attention to the presence of the scale bar in the visual documentation to ensure accurate measurements and compliance.

3.3 Evaluation

Qualitative and quantitative methods are used for performance evaluation. The specific evaluation methods for accuracy, efficiency and safety will be introduced in detail below.

The accuracy performance is evaluated using the Accuracy metric. Accuracy is calculated as $Acc = (TP + TN)/(TP + TN + FP + FN)$, where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively, quantifying the proportion of correct predictions among all outcomes.

Efficiency is evaluated by measuring computing latency—that is, the total execution time required for processing engineering tasks. While the security is evaluated by qualitative discussion. Because this research aims to adopt MLOB into a real-world MiC inspection task rather than modify the MLOB architecture, the security performance remains unchanged and unaffected from the original. Therefore, our analysis primarily focuses on quantifying efficiency and accuracy performance, specifically, without altering the inherent security features, especially on safeguarding data and computational security.

4 Prototype implementation

Based on the MLOB framework, we reengineered the inspection workflow, which is shown in Figure 3. The first step (step 1) is to reengineer the current inspection process into the logic that can be loaded into MLOB framework. Then, the worker collects the image (step 2) and uploads it to the blockchain network (step 3) by a transaction. Then the smart contract is automatically invoked (step 4) to process the image data and call the ML model to output the inspection result (step 5). The inspection result is stored in the blockchain as an immutable record (step 6). After that, the blockchain network sends a response to the worker. The worker can check the inspection result (step 7). The worker also has the opportunity to appeal the result and applies for manual review (step 8). The inspector downloads the record from the blockchain and inspects it manually (step 9). If the result is different, the inspection record in the blockchain will be updated (step 10).

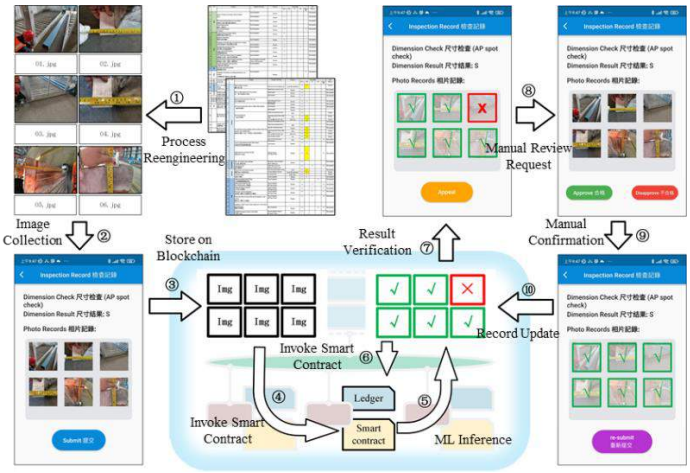


Figure 3. MLOB case study workflow.

4.1 Machine learning acquisition

MobileNet [6] was selected for this ML task, using 312 manually annotated images (qualified/unqualified) for training and 77 for evaluation. Training used sparse categorical cross-entropy loss, RMSprop optimizer, and 1000 epochs on TensorFlow 2.8 (Python 3.8), with accuracy as the metric. The trained model is saved in ".pth" format, storing weights and biases for inference or continued training in off-chain environments.

4.2 Machine learning conversion

The ML model with state dictionary format is used to run on local Python environment and cannot be used directly on blockchain smart contract. Therefore, it need a ML conversion process. As shown in Figure 4, such process needs two steps, firstly the state dictionary is converted to protocol buffer by a customized Python script, then the protocol buffer is converted to layers model by TensorFlow.js converter.



Figure 4. ML conversion workflow.

4.3 Machine learning safe loading

After conversion, an MD5 hash generates a unique

value for the ML model. When uploaded to the blockchain smart contract, this hash is compared against the original to verify integrity. Upon verification, the smart contract loads the necessary vision data for ML-based inspection. This hash comparison effectively detects any tampering or unauthorized modifications, ensuring the model's reliability for inspection tasks.

4.4 Consensus-based computing

After preparation, the ML model executes when inspection images upload to blockchain (Figure 3, Step 4), automatically triggering the smart contract. Hyperledger Fabric manages this process, preventing tampering through its permissioned structure, cryptographic hashing, and consensus mechanisms. The immutable decentralized ledger makes unauthorized changes immediately detectable, maintaining the ML inspection's trustworthiness.

5 Results and evaluation

5.1 Results

We evaluated MobileNet variants' performance and inference times (Figure 5): V1 (blue), V2 (red), and V3 (green), using round markers for local inference and triangular markers for on-chain inference. Newer, larger models like MobileNet V3 (large) and V2 (alpha=1.0) achieved higher accuracy. Notably, models with fewer than 1 million parameters had faster on-chain inference than local inference, demonstrating our ML in BT framework's advantage for small models. This study confirms the framework's feasibility with satisfactory speed and accuracy.

5.2 Evaluation and comparison

Based on our results, we deployed MobileNet V3(large) in the MLOB framework, comparing it against a traditional ML-blockchain isolated solution (baseline 1) [31] and a pure blockchain solution (baseline 2) [11] as shown in Table 1.

Table 1. Comparison with baseline methods

	Accuracy	Efficiency	Data security	Computational security
MLOB	99.52%	106ms	√	√
Baseline1	99.52%	80ms	√	
Baseline2	99.8%	877.6ms		
Baseline2 (HI)	90.2%	1345.6ms	√	

MLOB and Baseline1 both achieve 99.52% accuracy with robust data security, but while Baseline1 is slightly faster (80ms vs 106ms), only MLOB provides computational security. Baseline2's manual inspection achieves 99.8% accuracy but runs slower (877.6ms) and degrades significantly under high interference to 90.2% accuracy and 1345.6ms processing time.

In summary, MLOB delivers the optimal balance of comprehensive security, high accuracy, and competitive efficiency across various operational conditions.

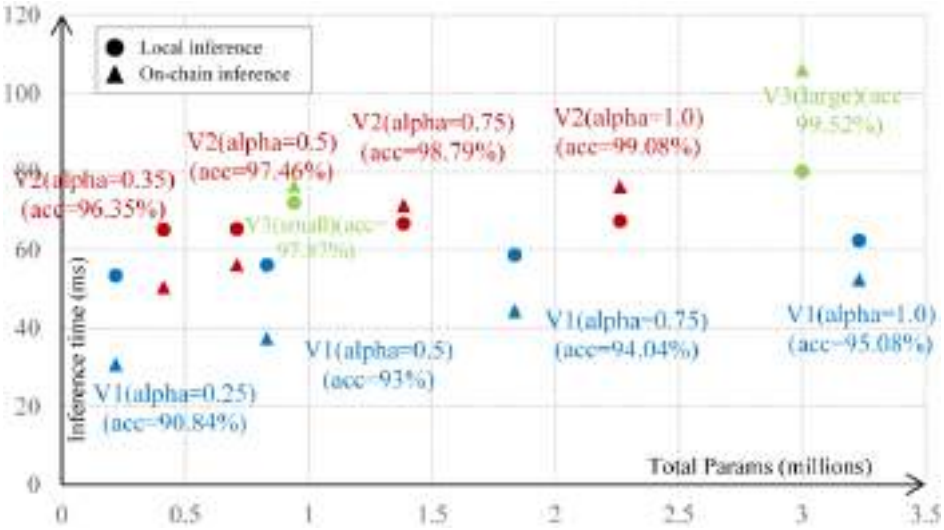


Figure 5. Comparison on different MobileNet variants.

6 Discussion

The contribution of this research lies both on theoretical and practical part. For the practical contribution, this study delivers a working prototype that applies the MLOB framework to reengineer the MiC inspection process. By integrating automated image-based inspections, secure model conversion, and consensus-driven decision-making via smart contracts on a blockchain network, the system achieves enhanced inference speed and robust accuracy while reducing delays and human errors. Its successful implementation in a Hong Kong MiC project demonstrates the approach's feasibility, scalability, and potential to establish new industry benchmarks for quality control and operational efficiency.

For the theorical part, the research offers a contribution by empirically mapping and optimizing the MiC inspection workflow. It identifies critical bottlenecks and validates that the integration of machine learning and blockchain can substantially improve inspection accuracy, efficiency, and security. This systematic re-engineering of traditional processes refines our understanding of how digital innovations can be seamlessly aligned with existing operational practices, setting a new standard for process optimization in construction.

The primary limitation of this research is its narrow focus on specific MiC inspection contexts, with limited real-world testing raising questions about broader applicability. Future research should address this by expanding the MLOB framework to more diverse scenarios and testing it on complex, large-scale projects with varied data characteristics to enhance generalizability. Additionally, collaboration with industry practitioners to integrate MLOB into practical workflows will provide valuable insights, promote real-world adoption, and help address potential implementation feasibility concerns across various construction environments. insights and promote adoption in real-world settings.

7 Conclusion

The MLOB framework integrates ML with blockchain to address accuracy, efficiency, and security challenges in MiC inspections. This approach ensures data integrity while enabling intelligent decision-making. Experimental results demonstrate significant performance improvements, showing potential to transform MiC processes. While proven effective in one real-world scenario, future research should test the framework across diverse construction contexts and optimize its computational resources.

References

[1] F.A. Acheampong, Big data, machine learning and the BlockChain technology: an overview, International Journal of Computer Applications 975 (2018), pp. 8887,

[2] J. Baek, D. Kim, B. Choi, Deep learning-based automated productivity monitoring for on-site module installation in off-site construction, Developments in the Built Environment 18 (2024), pp. 100382,

[3] F. Chen, H. Wan, H. Cai, G. Cheng, Machine learning in/for blockchain: Future and challenges, Canadian Journal of Statistics 49 (4) (2021), pp. 1364-1382,

[4] Z. Dong, W. Lu, Machine Learning on Blockchain (MLOB): A New Paradigm for Computational Security in Engineering, Engineering (2024) https://doi.org/10.1016/j.eng.2024.11.026.

[5] A. Harichandran, B. Raphael, A. Mukherjee, A HIERARCHICAL MACHINE LEARNING FRAMEWORK FOR THE IDENTIFICATION OF AUTOMATED CONSTRUCTION OPERATIONS, Journal of information technology in construction 26 (2021)

[6] A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, Mobilenets: Efficient convolutional neural networks for mobile vision applications, arXiv preprint arXiv:1704.04861 (2017)

[7] S. Kayikci, T.M. Khoshgoftaar, Blockchain meets machine learning: a survey, Journal of Big Data 11 (1) (2024), pp. 9,

[8] X. Li, W. Lu, F. Xue, L. Wu, R. Zhao, J. Lou, J. Xu, Blockchain-enabled IoT-BIM platform for supply chain management in modular construction, Journal of construction engineering and management 148 (2) (2022), pp. 04021195,

[9] C. Liu, S. ME Sepasgozar, S. Shirowzhan, G. Mohammadi, Applications of object detection in modular construction based on a comparative evaluation of deep learning algorithms, Construction Innovation 22 (1) (2022), pp. 141-159,

[10] Y. Liu, F.R. Yu, X. Li, H. Ji, V.C. Leung, Blockchain and machine learning for communications and networking systems, IEEE communications surveys & tutorials 22 (2) (2020), pp. 1392-1431,

[11] W. Lu, L. Wu, J. Xu, J. Lou, Construction E-inspection 2.0 in the COVID-19 pandemic era: A blockchain-based technical solution, Journal of Management in Engineering 38 (4) (2022),

pp. 04022032, https://doi.org/10.1061/(ASCE)ME.1943-5479.0001063.

[12] Z. Ning, H. Chen, X. Wang, S. Wang, L. Guo, Blockchain-enabled electrical fault inspection and secure transmission in 5G smart grids, IEEE Journal of Selected Topics in Signal Processing 16 (1) (2021), pp. 82-96,

[13] T.O. Olawumi, D.W. Chan, S. Ojo, M.C. Yam, Automating the modular construction process: A review of digital technologies and future directions with blockchain technology, Journal of Building Engineering 46 (2022), pp. 103720,

[14] M. Pan, Y. Yang, Z. Zheng, W. Pan, Artificial intelligence and robotics for prefabricated and modular construction: A systematic literature review, Journal of construction engineering and management 148 (9) (2022), pp. 03122004,

[15] X. Pan, B. Zhong, D. Sheng, X. Yuan, Y. Wang, Blockchain and deep learning technologies for construction equipment security information management, Automation in construction 136 (2022), pp. 104186,

[16] N. Papernot, P. McDaniel, A. Sinha, M.P. Wellman, Sok: Security and privacy in machine learning, 2018 IEEE European Symposium on Security and Privacy (EuroS&P), IEEE, 2018, pp. 399-414,

[17] J.-K. Park, B.-K. Kwon, J.-H. Park, D.-J. Kang, Machine learning-based imaging system for surface defect inspection, International Journal of Precision Engineering and Manufacturing-Green Technology 3 (2016), pp. 303-310,

[18] M. Pincheira, M. Antonini, M. Vecchio, Integrating the IoT and blockchain technology for the next generation of mining inspection systems, Sensors 22 (3) (2022), pp. 899,

[19] A. Rachman, R.C. Ratnayake, Machine learning approach for risk-based inspection screening assessment, Reliability Engineering & System Safety 185 (2019), pp. 518-532,

[20] S. Ravikumar, K. Ramachandran, V. Sugumaran, Machine learning approach for automated visual inspection of machine components, Expert Systems with Applications 38 (4) (2011), pp. 3260-3266,

[21] R. Shinde, O. Nilakhe, P. Pondkule, D. Karche, P. Shendage, Enhanced road construction process with machine learning and blockchain technology, 2020 International Conference on Industry 4.0 Technology (I4Tech), IEEE, 2020, pp. 207-210,

[22] Y. Tan, L. Chen, M. Huang, J. Li, G. Zhang, Automated geometric quality inspection for modular boxes using BIM and LiDAR, Automation in construction 164 (2024), pp. 105474,

[23] A.R. Tusnin, A.V. Alekseytsev, O. Tusnina, Using Machine Learning Technologies to Design Modular Buildings, Buildings 14 (7) (2024), pp. 2213,

[24] H. Wu, B. Zhong, H. Li, H.-L. Chi, Y. Wang, On-site safety inspection of tower cranes: A blockchain-enabled conceptual framework, Safety science 153 (2022), pp. 105815,

[25] H. Wu, B. Zhong, H. Li, J. Guo, Y. Wang, On-site construction quality inspection using blockchain and smart contracts, Journal of Management in Engineering 37 (6) (2021), pp. 04021065,

[26] L. Wu, W. Lu, F. Xue, X. Li, R. Zhao, M. Tang, Linking permissioned blockchain to Internet of Things (IoT)-BIM platform for off-site production management in modular construction, Computers in Industry 135 (2022), pp. 103573,

[27] L. Wu, W. Lu, R. Zhao, J. Xu, X. Li, F. Xue, Using blockchain to improve information sharing accuracy in the onsite assembly of modular construction, Journal of Management in Engineering 38 (3) (2022), pp. 04022014,

[28] Z. Yang, W. Lu, Lean Modular Integrated Construction Production Phase Planning under Uncertainties: A Big Data–Driven Optimization Approach, Journal of construction engineering and management 150 (6) (2024), pp. 04024048,

[29] Z. Zhan, Y. Dong, D.M. Doe, Y. Hu, S. Li, S. Cao, W. Li, Z. Han, Deep Learning and Blockchain-Driven Contract Theory: Alleviate Gender Bias in Construction, Journal of construction engineering and management 151 (3) (2025), pp. 04024216,

[30] W. Zhang, G. Sun, L. Xu, Q. Lu, H. Ning, P. Zhang, S. Yang, A trustworthy safety inspection framework using performance-security balanced blockchain, IEEE Internet of Things Journal 9 (11) (2021), pp. 8178-8190,

[31] B. Zhong, X. Hu, X. Pan, X. Chen, Z. Liu, Construction Quality Hazard Management with Deep Learning–Based Multimodal Storage Strategy–Enabled Blockchain, Journal of construction engineering and management 151 (1) (2025), pp. 04024184,

A Critical Review of Supervised and Reduced-Dataset Vision-Based Construction Activity Recognition Methods

Ali Ghelmani[1], Amin Hammad[1]*

[1]Concordia Institute for Information Systems Engineering, Concordia University, Montreal, Canada
ali.ghelmanirashidabad@concordia.ca
amin.hammad@concordia.ca*

Abstract

Tracking and monitoring construction activities (e.g., equipment and workers) is crucial for improving site productivity and safety. Manual monitoring, however, is time-consuming, error-prone, and costly. To address these challenges, computer vision (CV) techniques have emerged for automated detection and classification of construction activities. Traditional supervised CV-based methods often rely on extensive annotated datasets, making it impractical to accommodate the wide range of equipment types and activities. Reduced-dataset methods, such as self-supervised, semi-supervised, and generative approaches, have recently surfaced as promising alternatives. Furthermore, construction digital twins are rapidly gaining traction as an integral component of Construction 4.0. However, their practical implementation requires efficient monitoring systems capable of handling vast data streams and dynamic site conditions using edge devices. Unmanned Aerial Vehicles (UAVs) are an example of such devices that have been increasingly utilized to provide monitoring of dynamic site conditions and consequently, necessitate the development of corresponding activity recognition methods. Previous reviews have primarily focused on supervised methods with stationary cameras, as such, this paper fills a critical gap by examining recent reduced-dataset approaches and also providing a forward-looking review of CV-based activity monitoring on edge devices.

1 Introduction

Monitoring the activities of workers and equipment on construction sites is critical for improving project productivity, safety, and overall efficiency. Traditionally, this monitoring has relied on manual observations by site superintendents, a time-consuming and error-prone approach that struggles to keep pace with the complexity of modern construction projects [1]. As the demand for timely, high-quality project delivery increases, so does the need for automated and accurate continuous monitoring systems [2]. Furthermore, the concept of Construction Digital Twin (CDT) has gained traction as a promising platform for real-time data integration, simulation, and optimization [3]. However, one of the main challenges in CDT implementation, is the availability of accurate, real-time monitoring data, further exacerbating the need for the development of corresponding monitoring methods [3].

Recent prevalence of surveillance cameras in most construction sites has triggered the development of various automated vision-based activity recognition methods for addressing the above challenges [4]. Early vision-based approaches relied heavily on hand-crafted features [4], but the advent of deep learning, particularly Convolutional Neural Networks (CNNs), has shifted the focus towards more robust and scalable solutions [5].

Initial works employed 2D CNNs combined with rule-based logic or recurrent architectures [6, 7] for sequential processing of spatial and temporal information. More recently, 3D CNNs that jointly model spatiotemporal features have emerged to obtain improved performance [8–10]. However, the supervised approach of these methods requires large, annotated datasets for each piece of equipment and activity that is infeasible to collect, hence limiting their scalability [11]. To address these data-related bottlenecks, reduced-dataset methods (e.g., self-supervised [12], semi-supervised [13], and generative techniques [14]) have begun to gain traction. Self-supervised learning involves training models on unlabeled data using labels extracted from the data itself [12] (e.g., predicting rotations or temporal consistency); while semi-supervised learning leverages a small amount of labeled data in conjunction with unlabeled data, typically through techniques such as pseudo-labeling, to enhance learning efficiency [13]. Generative approaches, however, focus on synthesizing additional artificial data that mimic real-world conditions to expand the dataset size and diversity, thereby improving model

generalization [14]. While some studies have explored these approaches for equipment detection [11, 13], few have tackled the more complex problem of activity recognition [12].

Additionally, majority of the state-of-the-art activity recognition methods utilize videos collected from stationary cameras, however, CDT requirement for the availability of real-time data, given the dynamic nature of construction sites, necessitates the development of activity recognition methods employing edge devices such as Unmanned Aerial Vehicles (UAVs). As such, while previous review papers have focused on traditional supervised activity recognition methods (e.g., [4, 15]), there remains a need to thoroughly review these emerging technologies and discuss their potential in construction, particularly for activity recognition. This paper directly addresses these identified gaps by comprehensively reviewing recent advancements in reduced-dataset construction activity recognition methods and exploring the application and potential of Computer Vision (CV)-based monitoring solutions deployed on edge devices. The considered aspects aim to enhance the scalability of current supervised methods and address the challenge of providing dynamic construction site monitoring data essential for the development and implementation of CDTs.

2 Review Method

This study employed a systematic approach to identify and review the existing literature on vision-based construction activity recognition methods. Three main steps were followed: (1) paper collection, (2) initial screening, and (3) in-depth review and synthesis. First, an exhaustive search was conducted using multiple academic databases (e.g., Web of Science, Scopus, Google Scholar) to retrieve publications spanning approximately the last decade. The search keywords were designed to cover three broad categories: (1) terms related to construction (e.g., "construction," "vision-based methods," "automated monitoring," "activity recognition," "productivity analysis,"), (2) terms related to reduced-dataset deep learning methods (e.g., "self-supervised learning," "semi-supervised learning," "few-shot learning", "generative adversarial networks,", "synthetic data"), and (3) terms related to edge devices (e.g., "moving camera activity recognition," "UAV-based monitoring," "aerial construction monitoring").

Following the initial collection, titles and abstracts were screened to remove works focusing solely on non-construction domains or unrelated technologies. However, where gaps existed, particularly for activity recognition using moving edge devices, the search scope was broadened to include highly relevant and applicable methods from the general CV domain. Thus, papers discussing supervised or reduced-dataset approaches for construction activity recognition using stationary cameras, as well as relevant CV methods employing moving edge devices, were selected. Subsequently, a brief content review was then performed to ensure that selected papers substantially contributed to the current review scope. Finally, an in-depth qualitative analysis of the remaining papers was conducted to identify key trends and challenges in the literature.

3 Construction Activity Recognition Using Stationary Cameras

This section reviews supervised and reduced-dataset construction activity recognition methods using stationary cameras. Moreover, Figure 1 presents a categorization of these methods, along with representative publications for each approach.

3.1 Supervised Approaches

In the past decade, many vision-based methods have been proposed for the task of construction activity recognition. With the advent of deep learning, 2D CNNs became central to extracting spatial features from images, while temporal dependencies were increasingly modeled using additional modules such as Hidden Markov Models [16], or Long Short-Term Memory (LSTMs) networks [17]. Such combinations improved performance but often introduced computational overhead by separating spatial and temporal processing into multiple stages [8].

The 3D CNN-based methods alleviated this issue by incorporating the spatiotemporal data extraction into a single architecture. For instance, Chen et al. [1] proposed a three-stage excavator activity recognition method comprising of detection, tracking, and a final activity recognition stage using a 3D CNN network. In another study [5], they used a similar approach to detect excavator and dump truck activities, and to identify reasons for excavator idling by analyzing their interactions. A similar multi-stage approach for worker activity recognition was also proposed by Lou et al. [18]. However, the main limitations of these methods include not being fully optimized and also the propagation of errors from earlier stages to the later ones; thus, degrading the overall performance [19].

To address the abovementioned limitations, Jung et al. [19] proposed a 3D CNN-based single-stage method for simultaneous detection of multiple construction equipment and recognizing their activities. Others have also proposed single-stage methods based on the You Only Watch Once (YOWO) [20] method for joint detection and classification of construction workers' [8] and excavators' [21] activities. Despite improved performance, the main drawback of these methods is the

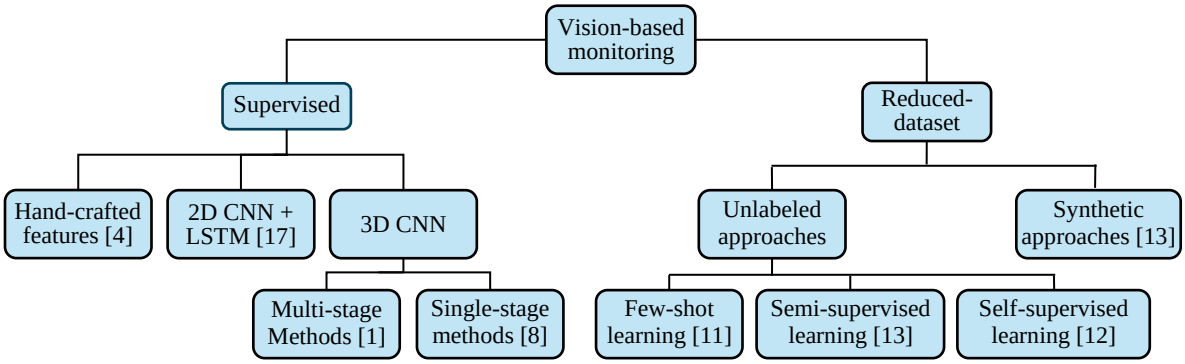


Figure 1. Categories of supervised and reduced-dataset construction activity recognition methods

necessity of post-processing due to their low per-frame activity recognition and localization performance.

Recently, a combination of multiple data modalities for activity recognition has also been suggested by some works. For example, Luo et al. [22] developed a CNN network by integrating RGB, optical flow, and gray-scale images to monitor workers' activities. Yang et al. [23] proposed a two-stream architecture for recognizing workers' unsafe behavior using a transformer model and a 3D CNN network to process RGB and Temporal Gradient (TG) data, respectively. Ghelmani and Hammad [10], also developed a multi-stream method, which leveraged TG data through knowledge distillation to achieve accurate, real-time per-frame activity recognition and localization performance for excavator activities.

Even though the state-of-the-art construction activity recognition methods can achieve high performance, their supervised approach necessitates extensive, annotated datasets for every piece of equipment and each specific activity. Since the types of construction equipment frequently change from one project phase to another [11], generating such comprehensive labeled dataset becomes a significant obstacle for the development of a general activity recognition method.

3.2 Reduced-Dataset Approaches

Researchers in the construction domain have proposed several methods for reducing the time and effort required for dataset collection and annotation. Liu and Golparvar-Fard [24] proposed crowdsourcing the annotation task using Amazon Mechanical Turk. However, such approaches still require human effort for data annotation and for removing the mistakes in the prepared dataset due to crowdsourcing [13]. Generally, reduced-dataset methods can be broadly categorized into methods that utilize a synthetic approach for data generation and methods that utilize unlabeled data for reducing the data annotation cost.

3.2.1 Synthetic Approaches

Calderon et al. [25] employed a virtual and kinematically articulated 3D model of excavators to generate synthetic 2D pose data for excavator activity recognition. Assadzadeh et al. [26] also generated a synthetic dataset for excavator pose estimation using domain randomization. Soltani et al. [27] developed a dataset of synthetically generated images of deformable excavator parts, which were used to estimate the excavator's 3D pose and consequently its kinematic posture. Data augmentation schemes using Generative Adversarial Networks (GANs) were also proposed by some researchers to enhance the variability and size of a collected dataset for construction equipment detection [28], activity recognition [14], and site monitoring [29]. Although synthetic approaches improve model's generalization through generation of challenging scenarios such as viewpoint variation, or partial occlusion, however a major limitation of these methods is the inherent visual differences between synthetic and real data [13], hence limiting their real-world applicability.

3.2.2 Unlabelled-Data Approaches

Few-shot learning. Kim and Chi [11] developed a few-shot learning approach by first training a model on a base dataset with sufficient training images of multiple construction equipment. Afterwards, the trained model was used in a transfer learning scheme to learn to detect new equipment with only a few available labeled images. In another study, Kim et al. [30] proposed a three-stage active learning method in which uncertainty evaluation was used to select an informative subset from the unlabeled data. Then, the selected small subset was labeled manually by the user and used for model training. Recently, Chen et al. [31] developed a multi-stage approach for recognizing excavator activities in a zero-shot learning setting in which the detected non-idling frames were input into a pre-trained CLIP (Contrastive Language-Image Pre-training) model [32] for activity recognition. A similar CLIP-based method for construction safety inspection was also proposed by Tsai et al. [33]. While few-shot and zero-shot learning techniques aim to minimize reliance on large, labeled

datasets, they remain restricted to learning only from the limited number of provided annotated images. Consequently, their final performance can fluctuate greatly depending on the quality and characteristics of the provided data [11].

Semi-supervised learning. Xiao et al. [13] introduced a teacher-student architecture in which a teacher network was first trained on a small, labeled dataset using weak data augmentation (e.g., horizontal translation). The trained teacher then generated pseudo-labels for unlabeled data that underwent stronger augmentations (e.g., cropping) to train the student network (the final output of this method). Such strong augmentations can simulate challenging scenarios, such as occlusions and significant viewpoint variations, potentially enhancing the student network's robustness to these real-world challenges. However, the main drawback of this method is the reliability of the generated pseudo-labels. In other words, Since the teacher network itself is trained in a supervised manner, it requires a sufficiently large, labeled dataset to produce accurate pseudo-labels, which greatly limits the method's applicability for reducing the annotation requirements. Furthermore, pseudo-labeling inherently risks introducing biases into the training data, as incorrect predictions made by the teacher model might be confidently assigned as accurate labels. Such biases could subsequently skew the model toward frequently occurring or easily identifiable examples, potentially diminishing performance on less common, complex, or ambiguous scenarios.

Self-supervised learning. The previous methods leveraged unlabeled data solely for construction equipment detection. However, the collection of an annotated dataset for activity recognition is even more challenging, as they often require frame-level labeling. Thereby increasing the necessity of developing methods that exploit unlabeled data for activity recognition. Nevertheless, only one study has addressed this need: Ghelmani and Hammad [12] proposed a self-supervised, contrastive learning approach for pretraining a 3D CNN network using unlabeled data, followed by fine-tuning using a small amount labeled data for excavator activity recognition. Although this method reduced annotation demands, it only applied self-supervised learning to activity recognition and not localization. No research to date has employed unlabeled data to facilitate the localization of construction equipment, even though this step is among the most annotation-intensive processes. Advancing research in this area could thus pave the way for more general and scalable construction activity recognition methods.

4 CV-based Activity Recognition on Edge Devices

As mentioned in Section 1, an indispensable component for CDT implementation is real-time monitoring using edge devices which include stationary and moving cameras (e.g., on UAVs). Methods developed for stationary cameras were reviewed in Section 3 while this section reviews methods developed for activity recognition using moving edge devices. Moreover, given the lack of activity recognition methods in the construction domain using edge devices, this section also presents a comprehensive review of the corresponding methods developed in the CV domain.

Activity recognition methods using moving edge devices can generally be divided into ground-based and aerial approaches.

4.1 Ground-based Methods

Ground-based methods in the CV domain are mainly concerned with applications related to self-driving cars. For instance, Yu et al. [34] proposed a method for real-time activity recognition in unconstrained video streams by utilizing frame-level foreground segmentation to obtain spatiotemporal proposals for activity recognition. Zhuang et al. [35] proposed a multi-stage road active agent detection and activity classification by initially passing the optical flow and RGB frames through corresponding backbones for object detection, followed by tracking and activity recognition stages. However, the main limitations of ground-based methods include: (1) the need for automated or manual navigations systems, which can be particularly challenging in a construction site; and (2) the limited field of view and ground-level camera angle, which can greatly increase the chance for occlusion and hence, reduce the model performance.

4.2 Aerial Methods

Compared with ground-based methods, a greater volume of research has been conducted on activity recognition using aerial imagery [36, 37]. Unlike traditional stationary cameras, UAVs capture videos from higher altitudes, allowing for a broader field of view. Additionally, UAVs can cover an large surveillance area due to their significant range of movement [36]. However, these factors introduce complex challenges related to detection, tracking, and activity recognition [38]. These challenges include the small resolution of objects of interest due to the high camera altitude. As a result, most of the pixels in a video frame are occupied by the background, with only a small fraction representing the targeted actions. Furthermore, the size and scale of the video can vary dramatically depending on the UAV's flying altitude, making the

extraction of spatiotemporally representative features quite challenging with the possible necessity of a preprocessing stage for scene stabilization [38]. Current activity recognition techniques are computationally demanding and memory-intensive, making it challenging to respect the time constraints in real-world applications. This is particularly true for edge devices with limited computational resources, such as UAVs [39]. To mitigate these constraints, existing methods typically adopt lightweight neural network architectures or hybrid inference schemes, where initial lightweight processing occurs onboard, and computationally intensive processing is delegated to remote servers or cloud-based resources (e.g., [38]). Nonetheless, effectively balancing computational efficiency with high model performance remains an active area of open research.

Dataset collection. The first step in addressing the above challenges using deep learning-based methods is the availability of extensive datasets. To this end, Barekatain et al. [40] proposed one of the first aerial video datasets for activity recognition comprising of 13 different types of human activities. Perera et al. [41] also developed a drone video dataset for activity recognition and pose estimation including 240 video clips recorded at low altitudes and low drone speed to provide high resolution human pose details.

Supervised approaches. Mliki et al. [42] proposed one of the first UAV activity recognition methods comprising of an offline training stage and an inference stage. Their approach utilized a scene stabilization technique for preprocessing followed by a human and non-human detector to filter out moving non-human entities. Finally, the detecting human agents were used for activity recognition. To address the problem of small-scale human targets in high altitude drone videos, Wang et al. [38] proposed the auto zoom and temporal reasoning method, which initially scales and crops human detections. Subsequently, a hybrid approach combining a lightweight 3D CNN network on edge devices and a more powerful 3D CNN network on local devices were employed for activity recognition.

Recently, many approaches have suggested a combination of feature-based and deep learning methods for aerial video activity recognition. For instance, Gundu and Syed [43] utilized Histogram of Oriented Gradients features for pattern extraction followed by a combination of CNN and LSTM networks for human activity recognition. Peng and Razi [44] used speeded up robust features (SURF) and Lucas-Kanade method for motion stabilization of the UAV captured videos, followed by detection and activity recognition stages. Kothandaraman et al. [45] employed Fourier analysis to separate out the human agent from the background in the frequency domain. Afterwards, a novel Fourier space-time attention module was utilized for activity recognition. Despite their improved performance, these methods still depend on manually engineered feature extractors, which introduce biases and constrain the representational capacity of the subsequent deep learning networks. The prevalence of such hybrid approaches can be attributed, in part, to the scarcity of large, high-quality aerial video datasets as reviewed earlier. As such, further development in this domain warrants the availability of large, annotated datasets.

Reduced-dataset approaches. As stated in Section 3.2, another possible approach for addressing the abovementioned limitations can be to utilize reduced-dataset approaches. To this end, Sultani and Shah [46] introduced a multi-task learning approach for activity recognition using aerial images, which leveraged a GAN network to synthetically transform ground activity video features into their aerial counterparts. To further improve performance, a small number of real aerial activity videos were also incorporated alongside the synthetic data. Choi et al. [47] proposed unsupervised and semi-supervised domain adaptation approaches to leverage the existing action recognition datasets in combination with unannotated or only few annotated drone videos for activity recognition. Although these approaches lessen the need for annotated aerial data, they mostly focus on domain adaptation from ground videos to aerial counterparts, a process that inevitably introduces errors and restricts the final model's accuracy. As such, there is a need in this area for new methods that can directly exploit unlabeled aerial videos to facilitate both activity recognition and localization, without relying solely on cross-domain transformations.

Aerial approaches in construction. Several studies have addressed the task of detecting construction entities from aerial images. For example, D. Kim et al. [48] proposed a UAV-assisted monitoring method as a proactive safety measure against struck-by hazards by utilizing a YOLOv3 [49] network for object localization and an image rectification method to measure actual distance from collected images. Guo et al. [50] proposed a method for detecting construction equipment from UAV-captured images. The proposed method combined orientation-aware feature fusion with a 2D CNN network for detecting equipment with arbitrary orientations. Comparison between different object detection methods were also carried out by Mutis et al. [51] for construction resource detection and by Pi et al. [52] to detect various construction site entities using aerial top-view images. A few studies have also developed UAV-based methods to improve construction site safety by detecting small and medium objects to avoid struck-by accidents [53–55] and for monitoring worker safety compliance [56].

To address dataset requirements for training deep learning methods Ersoz et al. [57] developed an aerial image dataset for construction machinery detection.

Others have developed methods to reduce annotated data requirements for aerial construction entity detection. For example, Bang et al. [58] proposed a variety of data augmentation techniques including moving-and-inpainting, cut-and-paste, and image-variation to improve the performance of a Faster R-CNN [59] network for detecting construction resources from UAV-captured images. Wang et al. [60] employed a 3D reconstruction technique using a UAV and photogrammetry to synthetically generate an image dataset for the analysis of crane productivity. Specifically, they reconstructed a crane bucket and used this synthetic data to conduct their study. However, as discussed in Section 4, all current aerial approaches in construction have focused solely on equipment or resource detection. To enable accurate, real-time monitoring that supports CDT implementation, methods must also address activity recognition, an area that remains unexplored for aerial-based methods.

5 Summary and Contributions

This paper provided a critical review of the state-of-the-art in vision-based construction activity recognition, examining both traditional supervised methods and the emerging reduced-dataset methods that leverage self-supervised, semi-supervised, and generative approaches. The data annotation barriers in current supervised methods were identified, followed by highlighting promising reduced-dataset strategies that mitigate these challenges. Moreover, given the growing demand of emerging CDT systems for the availability of accurate, real-time monitoring data, this review addressed a previously underexplored area: activity recognition using moving cameras (e.g., on UAVs) on edge devices, which is essential for capturing dynamic site conditions. Finally, the existing research gaps were identified, especially in reduced-dataset activity localization and aerial-based applications, hence proposing future directions for development.

References

[1] C. Chen, Z. Zhu, and A. Hammad, "Automated excavators activity recognition and productivity analysis from construction site surveillance videos," *Automation in Construction*, vol. 110, Feb. 2020.

[2] J. Kim, S. Chi, and J. Seo, "Interaction analysis for vision-based activity identification of earthmoving excavators and dump trucks," *Automation in Construction*, vol. 87, pp. 297–308, Mar. 2018.

[3] D.-G. J. Opoku, S. Perera, R. Osei-Kyei, and M. Rashidi, "Digital twin application in the construction industry: A literature review," *Journal of Building Engineering*, vol. 40, Aug. 2021.

[4] C. Chen, Z. Zhu, and A. Hammad, "Critical Review and Road Map of Automated Methods for Earthmoving Equipment Productivity Monitoring," *Journal of Computing in Civil Engineering*, vol. 36, no. 3, May 2022.

[5] C. Chen, Z. Zhu, A. Hammad, and M. Akbarzadeh, "Automatic Identification of Idling Reasons in Excavation Operations Based on Excavator–Truck Relationships," *Journal of Computing in Civil Engineering*, vol. 35, no. 5, Sep. 2021.

[6] J. Zhang, L. Zi, Y. Hou, M. Wang, W. Jiang, and D. Deng, "A Deep Learning-Based Approach to Enable Action Recognition for Construction Equipment," *Advances in Civil Engineering*, vol. 2020, Nov. 2020.

[7] M. Younesi Heravi, Y. Jang, I. Jeong, and S. Sarkar, "Deep learning-based activity-aware 3D human motion trajectory prediction in construction," *Expert Systems with Applications*, vol. 239, Apr. 2024.

[8] G. Torabi, A. Hammad, and N. Bouguila, "Two-Dimensional and Three-Dimensional CNN-Based Simultaneous Detection and Activity Classification of Construction Workers," *Journal of Computing in Civil Engineering*, vol. 36, no. 4, Jul. 2022.

[9] S. Jung, J. Jeoung, D. Lee, H. Jang, and T. Hong, "Visual–auditory learning network for construction equipment action detection," *Computer-Aided Civil and Infrastructure Engineering*, vol. 38, no. 14, pp. 1916–1934, Feb. 2023.

[10] A. Ghelmani and A. Hammad, "Improving single-stage activity recognition of excavators using knowledge distillation of temporal gradient data," *Computer-Aided Civil and Infrastructure Engineering*, vol. 39, no. 13, Jul. 2024.

[11] J. Kim and S. Chi, "A few-shot learning approach for database-free vision-based monitoring on construction sites," *Automation in Construction*, vol. 124, Apr. 2021.

[12] A. Ghelmani and A. Hammad, "Self-supervised contrastive video representation learning for construction equipment activity recognition on limited dataset," *Automation in Construction*, vol. 154, Oct. 2023.

[13] B. Xiao, Y. Zhang, Y. Chen, and X. Yin, "A semi-supervised learning detection method for vision-based monitoring of construction sites by integrating teacher-student networks and data augmentation," *Advanced Engineering Informatics*, vol. 50, Oct. 2021.

[14] Y. Shen, J. Wang, S. Mo, and X. Gu, "Data augmentation aided excavator activity recognition using deep convolutional conditional generative

adversarial networks," *Advanced Engineering Informatics*, vol. 62, Oct. 2024.

[15] B. Sherafat, C. R. Ahn, R. Akhavian, A. H. Behzadan, M. Golparvar-Fard, H. Kim, *et al.*, "Automated Methods for Activity Recognition of Construction Workers and Equipment: State-of-the-Art Review," *Journal of Construction Engineering and Management*, vol. 146, no. 6, Jun. 2020.

[16] X. Luo, H. Li, D. Cao, F. Dai, J. Seo, and S. Lee, "Recognizing Diverse Construction Activities in Site Images via Relevance Networks of Construction-Related Objects Detected by Convolutional Neural Networks," *Journal of Computing in Civil Engineering*, vol. 32, May 2018.

[17] J. Kim and S. Chi, "Action recognition of earthmoving excavators based on sequential pattern analysis of visual features and operation cycles," *Automation in Construction*, vol. 104, Aug. 2019.

[18] X. Luo, H. Li, Y. Yu, C. Zhou, and D. Cao, "Combining deep features and activity context to improve recognition of activities of workers in groups," *Computer-Aided Civil and Infrastructure Engineering*, vol. 35, no. 9, pp. 965–978, 2020.

[19] S. Jung, J. Jeoung, H. Kang, and T. Hong, "3D convolutional neural network-based one-stage model for real-time action detection in video of construction equipment," *Computer-Aided Civil and Infrastructure Engineering*, vol. 37, no. 1, pp. 126–142, Jan. 2022.

[20] O. Köpüklü, X. Wei, and G. Rigoll, "You Only Watch Once: A Unified CNN Architecture for Real-Time Spatiotemporal Action Localization." arXiv, Oct. 18, 2021.

[21] M.-Y. Cheng, M.-T. Cao, and C. K. Nuralim, "Computer vision-based deep learning for supervising excavator operations and measuring real-time earthwork productivity," *The Journal of Supercomputing*, vol. 79, no. 4, Mar. 2023.

[22] H. Luo, C. Xiong, W. Fang, P. E. D. Love, B. Zhang, and X. Ouyang, "Convolutional neural networks: Computer vision-based workforce activity assessment in construction," *Automation in Construction*, vol. 94, pp. 282–289, Oct. 2018.

[23] M. Yang, C. Wu, Y. Guo, R. Jiang, F. Zhou, J. Zhang, *et al.*, "Transformer-based deep learning model and video dataset for unsafe action identification in construction projects," *Automation in Construction*, vol. 146, Feb. 2023.

[24] K. Liu and M. Golparvar-Fard, "Crowdsourcing Construction Activity Analysis from Jobsite Video Streams," *Journal of Construction Engineering and Management*, vol. 141, no. 11, Nov. 2015.

[25] W. T. Calderon, D. Roberts, and M. Golparvar-Fard, "Synthesizing Pose Sequences from 3D Assets for Vision-Based Activity Analysis," *Journal of Computing in Civil Engineering*, vol. 35, no. 1, 2021.

[26] A. Assadzadeh, M. Arashpour, I. Brilakis, T. Ngo, and E. Konstantinou, "Vision-based excavator pose estimation using synthetically generated datasets with domain randomization," *Automation in Construction*, vol. 134, 2022.

[27] M. M. Soltani, Z. Zhu, and A. Hammad, "Skeleton estimation of excavator by detecting its parts," *Automation in Construction*, vol. 82, Oct. 2017.

[28] F. Baek, S. Park, and H. Kim, "Data Augmentation Using Adversarial Training for Construction-Equipment Classification." Nov. 26, 2019.

[29] H.-S. Kim, J. Seong, and H.-J. Jung, "Optimal domain adaptive object detection with self-training and adversarial-based approach for construction site monitoring," *Automation in Construction*, vol. 158, Feb. 2024.

[30] J. Kim, J. Hwang, S. Chi, and J. Seo, "Towards database-free vision-based monitoring on construction sites: A deep active learning approach," *Automation in Construction*, vol. 120, Dec. 2020.

[31] C. Chen, B. Xiao, Y. Zhang, and Z. Zhu, "Automatic vision-based calculation of excavator earthmoving productivity using zero-shot learning activity recognition," *Automation in Construction*, vol. 146, Feb. 2023,.

[32] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, *et al.*, "Learning Transferable Visual Models from Natural Language Supervision," In *Proceedings of the 38th International Conference on Machine Learning*, Jul. 2021, vol. 139, pp. 8748–8763.

[33] W.-L. Tsai, P.-L. Le, W.-F. Ho, N.-W. Chi, J. J. Lin, S. Tang, *et al.*, "Construction safety inspection with contrastive language-image pre-training (CLIP) image captioning and attention," *Automation in Construction*, vol. 169, p. 105863, Jan. 2025.

[34] L. Yu, Y. Qian, W. Liu, and A. G. Hauptmann, "Argus++: Robust Real-time Activity Detection for Unconstrained Video Streams with Overlapping Cube Proposals," In *2022 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW)*, Jan. 2022, pp. 112–121.

[35] A. Zhuang, E. Zhou, Q. Li, R. Dempster, A. Budhwani, M. Al-Sharman, *et al.*, "RADACS: Towards Higher-Order Reasoning using Action Recognition in Autonomous Vehicles." arXiv, Sep. 28, 2022.

[36] K. Nguyen, C. Fookes, S. Sridharan, Y. Tian, F. Liu, X. Liu, *et al.*, "The State of Aerial Surveillance: A Survey." arXiv, Jan. 12, 2022.

[37] G. U. Sai Theja, M. S. Murari, M. F. Singha, R. Patgiri, and A. Choudhury, "A Survey on

Surveillance Using Drones," In *Proceedings of the 2022 Fourteenth International Conference on Contemporary Computing*, New York, USA, 2022.

[38] X. Wang, R. Xian, T. Guan, C. M. de Melo, S. M. Nogar, A. Bera, *et al.*, "AZTR: Aerial Video Action Recognition with Auto Zoom and Temporal Reasoning." arXiv, Mar. 02, 2023.

[39] M. Ding, N. Li, Z. Song, R. Zhang, X. Zhang, and H. Zhou, "A Lightweight Action Recognition Method for Unmanned-Aerial-Vehicle Video," In *2020 IEEE 3rd International Conference on Electronics and Communication Engineering (ICECE)*, Dec. 2020, pp. 181–185.

[40] M. Barekatain, M. Martí, H.-F. Shih, S. Murray, K. Nakayama, Y. Matsuo, *et al.*, "Okutama-Action: An Aerial View Video Dataset for Concurrent Human Action Detection," In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jul. 2017, pp. 2153–2160.

[41] A. G. Perera, Y. W. Law, and J. Chahl, "Drone-Action: An Outdoor Recorded Drone Video Dataset for Action Recognition," *Drones*, vol. 3, 2019.

[42] H. Mliki, F. Bouhlel, and M. Hammami, "Human activity recognition from UAV-captured video sequences," *Pattern Recognition*, vol. 100, p. 107140, Apr. 2020.

[43] S. Gundu and H. Syed, "Vision-Based HAR in UAV Videos Using Histograms and Deep Learning Techniques," *Sensors*, vol. 23, no. 5, 2023.

[44] H. Peng and A. Razi, "Fully Autonomous UAV-Based Action Recognition System Using Aerial Imagery," In *Advances in Visual Computing*, Cham, 2020, pp. 276–290.

[45] D. Kothandaraman, T. Guan, X. Wang, S. Hu, M. Lin, and D. Manocha, "FAR: Fourier Aerial Video Recognition," In *Computer Vision – ECCV 2022*, Cham, 2022, pp. 657–676.

[46] W. Sultani and M. Shah, "Human action recognition in drone videos using a few aerial training examples," *Computer Vision and Image Understanding*, vol. 206, p. 103186, May 2021.

[47] J. Choi, G. Sharma, M. Chandraker, and J.-B. Huang, "Unsupervised and Semi-Supervised Domain Adaptation for Action Recognition from Drones," In *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Mar. 2020, pp. 1706–1715.

[48] D. Kim, M. Liu, S. Lee, and V. R. Kamat, "Remote proximity monitoring between mobile construction resources using camera-mounted UAVs," *Automation in Construction*, vol. 99, pp. 168–182, Mar. 2019.

[49] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[50] Y. Guo, Y. Xu, and S. Li, "Dense construction vehicle detection based on orientation-aware feature fusion convolutional neural network," *Automation in Construction*, vol. 112, Apr. 2020.

[51] I. Mutis, V. A. Joshi, and A. Singh, "Object Detectors for Construction Resources Using Unmanned Aerial Vehicles," *Practice Periodical on Structural Design and Construction*, vol. 26, no. 4, Nov. 2021.

[52] Y. Pi, N. D. Nath, and A. H. Behzadan, "Convolutional neural networks for object detection in aerial imagery for disaster response and recovery," *Advanced Engineering Informatics*, vol. 43, p. 101009, Jan. 2020.

[53] Y. Lu, W. Qin, C. Zhou, and Z. Liu, "Automated detection of dangerous work zone for crawler crane guided by UAV images via Swin Transformer," *Automation in Construction*, vol. 147, Mar. 2023.

[54] H. Kim, J. Seong, and H.-J. Jung, "Real-Time Struck-By Hazards Detection System for Small- and Medium-Sized Construction Sites Based on Computer Vision Using Far-Field Surveillance Videos," *Journal of Computing in Civil Engineering*, vol. 37, no. 6, Nov. 2023.

[55] S. Kim, S. H. Hong, H. Kim, M. Lee, and S. Hwang, "Small object detection (SOD) system for comprehensive construction site safety monitoring," *Automation in Construction*, vol. 156, p. 105103, Dec. 2023.

[56] A. Akinsemoyin, I. Awolusi, D. Chakraborty, A. J. Al-Bayati, and A. Akanmu, "Unmanned Aerial Systems and Deep Learning for Safety and Health Activity Monitoring on Construction Sites," *Sensors*, vol. 23, no. 15, 2023.

[57] A. B. Ersoz, O. Pekcan, and E. Akbas, "AIDCON: An Aerial Image Dataset and Benchmark for Construction Machinery," *Remote Sensing*, vol. 16, no. 17, 2024.

[58] S. Bang, F. Baek, S. Park, W. Kim, and H. Kim, "Image augmentation to improve construction resource detection using generative adversarial networks, cut-and-paste, and image transformation techniques," *Automation in Construction*, vol. 115, p. 103198, Jul. 2020.

[59] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[60] D. Wang, X. Wang, B. Ren, J. Wang, Z. Zeng, D. Kang, *et al.*, "Vision-Based Productivity Analysis of Cable Crane Transportation Using Augmented Reality-Based Synthetic Image," *Journal of Computing in Civil Engineering*, vol. 36, no. 1, Jan. 2022.

Automating Weekly Construction Activity Progress Reporting: Leveraging AI-Driven Workflows

Eyob Mengiste, Muammer Semih Sonkor, Zihao Zheng, Samuel A. Prieto, and Borja García de Soto

S.M.A.R.T. Construction Research Group, Division of Engineering, New York University Abu Dhabi (NYUAD), Experimental Research Building, Saadiyat Island, P.O. Box 129188, Abu Dhabi, United Arab Emirates
eyob.mengiste@nyu.edu, semih.sonkor@nyu.edu, zihao.zheng@nyu.edu, samuel.prieto@nyu.edu, garcia.de.soto@nyu.edu

Abstract –

This paper presents an automated workflow for weekly construction progress reporting, streamlining data integration and analysis. The proposed approach focuses on three key areas: planned activities, weekly performance, and projected progress. Inputs include an updated baseline schedule and weekly inspection data, such as images. Outputs, generated autonomously, provide project status and performance metrics, including Earned Value and Planned Value. Leveraging multimodal large language models (MLLMs), the system processes text and images, enabling seamless data integration. Key contributions include a simplified, reliable reporting process that reflects actual construction execution and planning while reducing time and resource demands. The paper also addresses implementation challenges, AI-driven solutions, and scalability for broader construction reporting applications.

Keywords –

Construction Progress Reporting, Automation Workflow, Multimodal Large Language Models, Project Management, EVM

1 Introduction

Progress reporting is a cornerstone of effective construction management, offering stakeholders insights into the current status, challenges, and future directions of a project. Weekly reports serve as a critical tool for tracking planned versus actual progress, identifying delays, and updating schedules. However, generating these reports remains a labor-intensive process, often requiring manual consolidation of diverse data sources, including project plans, contracts, on-site observations, and multimedia documentation. Not to mention that such reports are very subjective and based on multiple human-relying factors, such as the quality and completeness of the collected data, the analysis made on said data, and the completion of the actual report.

This paper addresses the challenge of automating construction site progress reporting by designing and implementing a streamlined semi-automated workflow. The proposed system focuses on three core sections of the weekly report:

- Automatically identifying what has been planned for the reporting week;
- Evaluate and determine the activity performance of the reporting week using sight inspection data inputs.
- Update and generate feasible activity plans for the following week.

These sections were selected due to their recurring importance for tracking progress and generating reliable weekly construction status reports.

2 Background

In the construction industry, weekly reports are essential tools for monitoring project progress and ensuring effective communication among stakeholders [1]. A critical component of these reports is the assessment of progress made by construction activity. This evaluation typically involves referencing data from previous reports or updated schedules, while the actual progress for the current reporting week is manually estimated or measured on-site and manually entered into the report. This manual process demands significant time from the engineering team for data synthesis and evaluation, potentially diverting attention from other vital tasks [2].

The quality and accuracy of weekly progress reports heavily depend on the competencies of the individuals preparing them, which can lead to inconsistencies, subjectivity, and biases [3]. The quality of weekly reports is directly linked to the successful coordination between stakeholders and the completion of construction projects. High-quality reports provide early warnings of potential

risks and constraints, enabling timely corrective actions.

Automating the generation of construction reports could enhance their quality and accuracy, minimizing subjectivity. By reducing time spent on less productive activities, such as inspection of the collected data and writing the report, automation aligns with lean construction principles, which emphasize waste reduction and value maximization [4]. Within the Last Planner® System (LPS), accurate weekly reports are vital for improving communication and collaboration among all stakeholders involved in the construction process [5]. Weekly plans and reports play a crucial role in identifying and managing construction constraints. They structure constraint information and facilitate the development of data-informed lookahead plans, detailing upcoming work and potential obstacles. This process incorporates actual construction progress into the master schedule, ensuring preparedness for potential issues and adherence to plans [6].

Incorporating automation into report generation not only improves efficiency but also supports the continuous improvement ethos central to lean construction. By streamlining information flow and reducing manual errors, automated reporting systems contribute to more predictable and successful project delivery [7].

2.1 Automated Construction Reporting

As stated above, automated construction reporting have multiple benefits of handling overwhelming data influx from multiple sources and improving productivities. Various studies have proposed workflows aimed at automating the whole construction progress monitoring and reporting system, which consists of four main processes: data collection, data analysis, progress estimation, and visualization [8]. El-Omari and Moselhi [9] implemented automated data acquisition technologies within an IT platform for construction project reporting. Shrestha et al. [10] developed computational algorithms to generate as-built schedules, both during and post-construction, updating progress dynamically. Jafari et al. [11] leveraged Natural Language Processing (NLP) and Machine Learning (ML) for automated reporting extraction, coupled with stochastic simulations to predict time and cost. Xiao et al. [12] integrated computer vision-based data acquisition with ChatGPT to automatically generate construction reports. Shamsollahi et al. [13] created a complete workflow for automated progress monitoring, incorporating object recognition, objectives tracking, and automated report generation. Ekanayake [14] explored computer vision-based progress monitoring, focusing on automated data collection, analysis, and progress report generation.

These advancements underscore the potential of automation in improving the efficiency and accuracy of construction reporting. However, many of them focus on only one or two sub-processes while overlooking manual efforts and the challenges of developing automation tools, which hinders the implementation of a fully integrated workflow.

2.2 Vision-Based Construction Progress Recognition

One of the challenging parts of automated construction reporting is the automated construction progress recognition. The use of images in construction progress monitoring has grown significantly, facilitated by advances in ML [15]. Martinez et al. [16] used an R-CNN model and CCTV footage to track the progress of floor paneling workstations. Wei et al. [17] employed a Mask R-CNN model to track construction activities such as wall construction progress on a single floor. They also used an improved PointRend model to monitor the progress of prefabricated components across an entire building [18]. While these models have demonstrated high accuracy in task-specific scenarios, their effectiveness hinges on high-quality annotated datasets and specialized training. Extensive validation is necessary to ensure these models perform reliably across diverse construction contexts.

In recent days, MLLMs and Vision Language Models (VLMs) have shown tremendous ability to handle text and image information simultaneously in zero-shot tasks transferred to many real-world scenarios [19]. A VLM developed by Tsai [20] for construction safety reporting demonstrated its ability to generate textual descriptions for safety inspectors. Similarly, GPT models, which support few-shot and zero-shot learning can quickly analyze data, adapt to tasks with minimal training, and improve reporting, progress monitoring, and communication, have shown great potential for the construction industry [21]. Pu et al. [22] explored the capabilities of MiniGPT4-7B and GPT-4o in generating automated construction reports, combining automated data collection with LLMs. Despite the aforementioned efforts, the integration of the advanced construction progress recognition methods with the whole automated construction reporting remains underexplored.

In summary, automated construction reporting involves multiple processes, including data acquisition, data extraction, data analysis (e.g., progress description, delay identification, and forecasting), and visualization. While existing research has showcased advancements in automating multi-source data integration and generating construction reports to support decision-making, these efforts often rely on code-intensive workflows and substantial manual input in each sub-process. Vision-based data analysis remains heavily dependent on task-specific datasets and exclusive access to long-term computational resources.

The proposed approach in this paper uses an open-source and expandable workflow automation platform, focusing on seamless and easy integration in current practices and processes. By utilizing a low-code environment and integrating various data processing algorithms, the proposed approach aims to:

1. Simplify and automate progress reporting processes, reducing potential time waste.
2. Eliminate the need for specialized expertise in data collection, processing, and interpretation, by leveraging existing widely used tools for data generation and integration.
3. Minimize bias, reduce manual operations, and enhance workflow efficiency through automation.

3 Methodology

The methodology employed in this study is explained through process flowcharts using the BPMN notation [23]. The overall process is divided into two main sections: (1) manual process (Figure 1) and (2) automated process (Figure 2). BPMN uses "pools" to show the processes of different entities and "lanes" within these pools to present different sub-entities. In this paper, automated and manual processes are presented in separate pools, and different roles involved in the manual process flow are shown in different lanes.

The overall process involves using a developed chatbot to upload site images, and further request weekly progress reports from the chatbot conversation. Telegram was chosen to create the chatbot due to its seamless integration with automation tools. As shown in Figure 1 and Figure 2, a Common Data Environment (CDE) is employed for storing images and schedule files, which serve as the process inputs. While the process reduces manual effort by automating progress reporting, the actual schedule updating remains a separate task and is outside the scope of this paper. This study used Google Drive as the CDE because of its convenience. Overviews of manual and automated parts of the process are presented in the following subsections.

3.1 Manual Process

The manual process involves three roles: (1) site employee, (2) project scheduler, and (3) management/client (Figure 1). The details of each role and related tasks are as follows:

1. **Site employee:** This role is responsible for taking site inspection photos to capture and document the progress. It can be anyone who works on-site and has access to the developed chatbot. The only tasks related to this role are taking site photos and uploading them to the chatbot.
2. **Project scheduler:** This role is responsible for converting the most recently updated MS Project schedule file into the comma-separated values (CSV) format and uploading it to the CDE. CSV was chosen in this study for its structured format, which facilitates data extraction. These tasks and the site employee's tasks are performed in parallel since they are not dependent on each other (Figure 2).
3. **Management/client:** This role represents the stakeholders requesting the report from the automated process flow. Some examples of this role are the site manager, the project manager, and the client. Due to the unbiased nature of the workflow, the generated report provides transparency regarding the project progress and improves the information flow among the stakeholders. The "request report through chatbot" task is dependent on the previous tasks related to photo and schedule uploads performed by the site employee and scheduler.

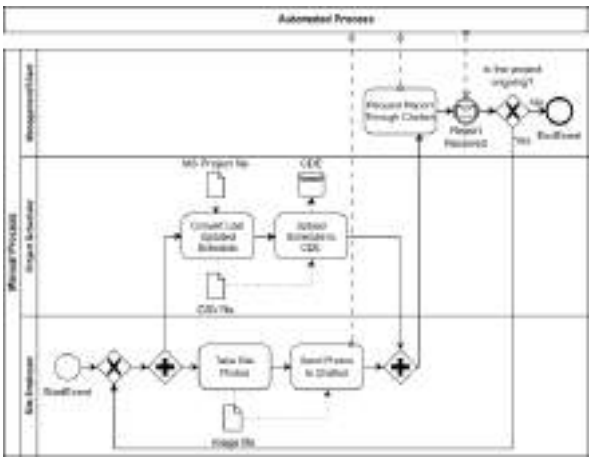


Figure 1. Manual part of the overall process flow

3.2 Automated Process

The automated part of the overall process flow was developed on the low-code workflow automation platform, n8n [24]. This platform was used due to its advanced graphical user interface, which allows visual programming without requiring extensive coding knowledge. However, other low-code development platforms or traditional programming can serve the same purpose.

The automated part of the process flow has two message start events, "photo received" and "report request received", triggered by the tasks from the manual process. Once the site inspection photos are uploaded to the chatbot in the manual process, the "photo received" event is triggered, and the photos are automatically uploaded to the CDE. The connection between the chatbot and the automation process was made using the Telegram node in n8n.

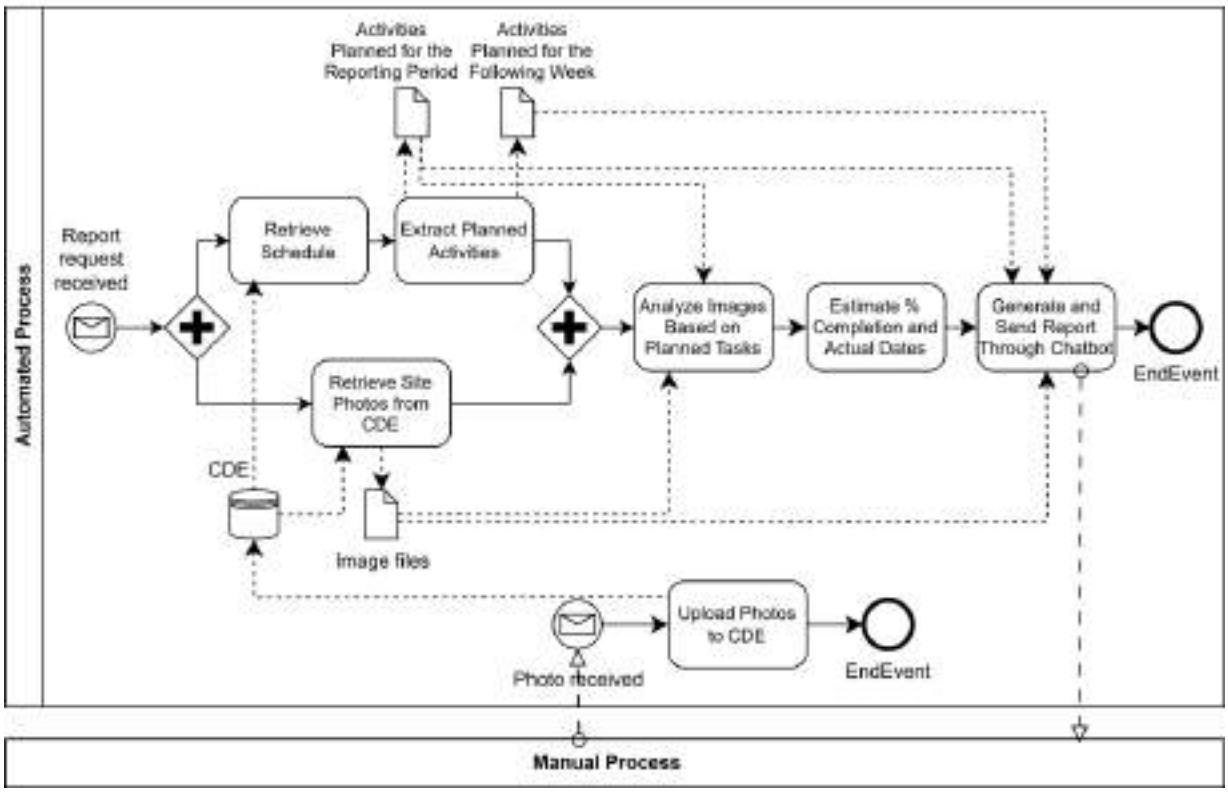


Figure 2. Automated part of the overall process flow

Once the report request is received from the project management or the client, the retrieval of the schedule and photos is performed. These are facilitated through the Google Drive integration of the n8n platform in this study. The activities planned for the reporting period and the upcoming week are extracted from the schedule file using custom JavaScript code and integrated into the n8n workflow via the Code node. The images retrieved from the CDE are analyzed based on the planned tasks using a MLLM. GPT-4o was used in this study due to its powerful capabilities and the built-in OpenAI nodes in n8n. However, an open-source model could be utilized for the same purpose if data privacy or cost were to be prioritized. GPT-4o was used to generate the prompts for its own tasks. This approach ensures structured and efficient prompts for each task.

The MLLM is asked to analyze the construction site images and determine the percent completion of each planned activity, having the rest of the provided information as context for said analysis. The result from this analysis is passed into another LLM instance to compare it with the percent complete values from the last updated schedule, and then output the highest percentages based on this comparison. These percent complete values are then used to calculate the estimated actual start dates for the activities started during the reporting period and the estimated actual finish dates for the activities completed during the reporting period.

Finally, the planned tasks for the reporting period, their percent complete values, the activities planned for the following week, and the site inspection photos are included in the weekly progress report generated as the last step of the automated workflow. The generated report is sent to the user who requested the report through the chatbot.

4 Case-study

This case study presents the automated generation of a weekly construction progress report for a fit-out construction project conducted on a university campus. The project involves the conversion of a meeting venue into an auditorium, incorporating several specialized upgrades to meet the functional and aesthetic requirements of the new space. These upgrades include structural modifications for tiered seating arrangements, enhanced audio-visual systems tailored for auditorium use, and upgrading the heating, ventilation, and air conditioning (HVAC) systems to accommodate the revised space usage. Additionally, the project involves aesthetic and functional enhancements such as floor carpeting and the integration of electrical systems,

particularly beneath the stepped seating, to support the operational needs of the auditorium.

The specific week in the timeline used for this case study was selected due to the significant activity observed on-site during that week and the availability of comprehensive data, making it an ideal timeframe for detailed analysis. The focus on this time period captures substantial progress while providing a reliable snapshot of the construction activities. For consistency and accuracy in assessment, the schedule is assumed to be up to date as of the last day of the previous reporting date, reflecting all completed activities up to that point. This assumption ensures a clear baseline for evaluating the progress during this time.

4.1 Data inputs

One of the data inputs necessary for the process was the updated schedule. As of the last day of the previous reporting week, the overall project progress was 54%, with significant milestones achieved. Partition work was 83% complete, with closure and wall finishing still pending. The ceiling installation had reached 37%, with ceiling panel installation not fully completed. Floor preparation tasks were fully completed at 100%. MEP work stands at 38%, with ductwork and data conduit installations finished, wiring and cable installation at 60%, and initial stages of HVAC air diffuser installation underway. The seating area framework and concreting were 63% complete, with framework installation fully finished and rebar installation nearing completion at 90%. The updated list of tasks with their percentage completion is shown in Figure 3. It highlights the project's progress, with several activities scheduled for completion in the target week. The second critical input to the process was raw data from the site inspection. In this case, the site inspection was conducted on the 5th day of the reporting week, and 360 images, such as the one shown in Figure 4, were collected.

4.2 Report generation process

The report generation process is initiated by sending the image taken from the construction site during the inspection directly to the Telegram bot. The Telegram bot initiates the image renaming and uploading process to the cloud data storage space, where the image is renamed to carry the data collection date and stored in the Google Drive folder that is prepared for this purpose.

Custom-made plugins in the scheduling software, Microsoft Projects in this particular case study, were employed to convert the existing schedule to a structured, easily readable data format such as CSV and uploaded to the cloud storage that the workflow can access. As mentioned previously, this case-study experiment employs n8n as the low-code workflow automation platform to connect all the different processes. Hence, a command "/report" in the Telegram bot initiates the relevant information generation process, using the available data inputs that are pre-uploaded in the cloud storage, following the process shown in Figure 1 and Figure 2.

Task Name	% Complete
A6 Room 2 (3B) construction	**54%**
Partition Work	**83%**
Gypsum Board Installation	100%
Closure, finishing wall surfaces	0%
Ceiling Installation	**37%**
Ceiling Frame Installation	100%
Ceiling Panel Installation	10%
Floor preparation	**100%**
Layout and mark the floor for precise positioning of partitions, conduits, seating framework, and other installations.	100%
MEP work	**38%**
Installation of MEP Ductwork and fixture installation	100%
Installation of MEP and Data Conduits	100%
Wiring and Cable Installation	60%
Installation of Wall Fixtures (MEP and Data)	0%
Installation of Floor Boxes (MEP and Data)	0%
Installation of HVAC Air Diffusers and Grilles	7%
Conduct final checks and installation of any remaining electrical components	0%
Seating area framework and concreting	**63%**
Seating Steps Framework Installation	100%
Seating Area Floor Rebar Installation	90%
Concrete casting and covering it with wet curing blanket for seating area	0%

Figure 3. Extract of the Baseline tasks list and percentage completion updated as of the last day of the previous reporting week.



Figure 4. 360 image taken during the site inspection on the 5th day of the reporting week

5 Results

The generated report is composed of five sections, including (1) basic project information, (2) a high-level description of the current weekly site work progress, (3) a summary of weekly project status, (4) a list of tasks planned for next week, and finally (5) a list and descriptions of the inspection photos.

The basic project information section carries the static project information such as Project ID, and designation of the specific zone in the project where the current report is targeted to. This section also provides information regarding the specific reporting week.

The weekly report is meant to be delivered to

nontechnical stakeholders, such as clients with limited understanding of the detailed project status metrics. Therefore, the second section provides the list of tasks planned for the current week, dividing these tasks into three different categories: (1) *complete*, mainly for the set of tasks that are finalized during the week; (2) *in progress*, for tasks that are still active but not complete; and the last category (3) *planned but not started yet*, mainly for those tasks that the schedule indicates that they need to be started during the week but have not actually been started in the construction site yet.

For the project used in this case study, even though all of the tasks except ceiling panel installation were planned to be completed by the end of the studied week, the process generated a report showing that none of the tasks were actually completed. Moreover, three of the six planned tasks (i.e., installation of wall fixtures, installation of floor boxes, and concrete work for the seating area) are classified as planned but not started yet in the generated report.

This report offers (Table 1) even more expanded information on the planned vs actual performance of the project showing the Work Breakdown Structure (WBS) ID, Task Name, Status, % Completion to Date (EV), % Planned Completion (PV), and Actual Completion of the work at the end of the reporting week (AC).

Table 1: Output summary of the project status

WBS	Task Name	Status	EV*	PV	AC
1.2.2	Ceiling Panel Installation**	In progress	60%	85%	21%
1.4.3	Wiring and Cable Installation**	In progress	70%	100%	80%
1.4.4	Installation of Wall Fixtures (MEP and Data)**	Not started	0%	100%	0%
1.4.5	Installation of Floor Boxes (MEP and Data)**	Not started	0%	100%	0%
1.4.6	Installation of HVAC Air Diffusers and Grilles**	In progress	50%	100%	90%
1.5.2	Concrete Work for Seating Area**	Not started	0%	100%	100%

* The information was generated from site inspection photos using VLM; hence, verification from the site crew would be required.

** Shows tasks that are potentially delayed according to their SPI value.

Based on the performance of the current week, the automated workflow also presents a section in the report suggesting the list of planned tasks for the next week. This computation considers the SPI of the EVM. Given this case study does not involve the cost calculation, the SPI calculation is summarized as EV/PV, where EV is taken as equivalent to the percentage completion and PV is computed as the ratio of the number of actual activity days related to a planned task with the number of planned days of the task-related activity.

Based on the SPI, the baseline finish date and the revised task duration, the new list of tasks planned for the following week is reported. The last segment of the report presents the raw image data involved in the generation of the report with systematically defined captions that relate the site status and active tasks with the actual image collected. Readers interested in the specific details of the report generated and the data used for the generation of the report are encouraged to contact the authors to request a copy.

6 Discussion

This study successfully demonstrates the potential of automated processes for generating construction weekly progress reports, effectively addressing the three key objectives stated at the beginning of the paper: reducing potential time waste, eliminating the need for specialized expertise in data collection, processing, and interpretation, and lastly minimizing bias in the report results.

The automation demonstrated in this case study reduces potential time waste by streamlining the data collection and reporting processes. Traditionally, generating progress reports involves multiple manual steps, including data collection, image labeling, information synthesizing and processing into a coherent report format. The workflow presented in this study minimizes these repetitive tasks by automating image renaming, cloud storage, data integration, structuring the information into the report, and finally, updating the schedule as per the actual site findings. As a result, site personnel can allocate more time to critical construction activities rather than administrative tasks. This efficiency is particularly valuable in time-sensitive projects where delays in reporting can lead to cascading schedule disruptions.

A significant strength of the approach is its ability to eliminate the dependency on specialized expertise for data collection, processing, and interpretation. By leveraging widely used tools, such as Microsoft Project and Google Drive, the system simplifies complex workflows. The integration of a Telegram bot for streamlined image data uploading and a low-code workflow automation platform like n8n ensures accessibility for non-specialist stakeholders. This supports broader adoption in the construction industry, aligning with recent findings emphasizing the role of digital technologies in minimizing manual operations and enhancing workflow efficiency.

The automated generation of reports succeeded in capturing nuanced aspects of project progress. For instance, the system provided detailed insights into the planned versus actual status of tasks, identifying discrepancies such as the non-initiation of key activities (e.g., wall fixture installations and HVAC diffusers).

Such insights are critical for stakeholders, enabling proactive decision-making and improving a deeper understanding of site progress without requiring technical expertise. This aligns with previous studies advocating for automated reporting systems to enhance decision-making efficiency.

The case study highlights the ability of the proposed automation system to reduce bias, a major challenge in manual reporting processes. By relying on data inputs directly from site inspections and schedule updates, the developed workflow fully eliminates potential subjective interpretations often introduced by human intermediaries. For example, the categorization of tasks as "not started" or "in progress" is derived from structured data rather than personal judgment, ensuring consistency in reporting. This objectivity is further enhanced by visual evidence from 360-degree images, which provide a verifiable basis for evaluating site conditions. Such an approach aligns with findings that emphasize the role of the proposed automation workflow in enhancing transparency and objectivity in construction management.

7 Limitations and Future Work

Despite these strengths, the study identifies some limitations. First, the reliance on scheduled updates and inspection photos as primary data inputs introduces vulnerabilities to inaccuracies. For instance, the discrepancy between planned and actual progress, particularly the underreporting of completed tasks (i.e., in the case study, task “Concrete work for seating area”), underscores the need for site verification. While automation streamlines data integration, human oversight remains essential to validate outputs.

Second, the scope of the study focuses on construction progress reporting without considering cost metrics. While the SPI computation in the study provides a functional basis for assessing schedule adherence, the lack of cost-related indices limits its applicability in budget-critical projects. Future work should integrate cost parameters to provide a more comprehensive project performance evaluation. Additionally, the use of static image data and dependency on a basic interpretation of the image content introduces challenges in capturing dynamic site conditions, such as real-time adjustments to workflows or unforeseen disruptions. Moreover, without performing evaluations of the work volume, and purely relying on the evaluations performed by the MLLM, understanding the accurate extent of work from the existing information could be challenging. To address these limitations, future work should incorporate advancements in real-time data acquisition methods (e.g., automated LiDAR data acquisition), considering more information as data input (e.g., an audio voice note from the worker collecting the images clarifying details about the extent of completed work) as well as automated work volume measurement mechanisms in the workflow.

8 Conclusions

This study introduces an innovative and practical approach to automating construction progress reporting, demonstrating benefits in minimizing manual effort, reducing reporting bias, and improving workflow efficiency. A case study conducted at a university campus highlights the system’s ability to generate detailed, accurate, and accessible reports that cater to both technical and non-technical stakeholders. By integrating schedule data with image-based site inspections and leveraging MLLMs for data processing and analysis, the proposed approach effectively bridges the gap between traditional reporting methods and the evolving demands of modern construction management. This integration ensures seamless data processing, enhances reporting accuracy, and supports more informed decision-making throughout the project lifecycle.

Acknowledgments

This research was partially supported by different Centers at NYUAD. In particular, the Center for Sand Hazards and Opportunities for Resilience, Energy, and Sustainability (SHORES) funded by Tamkeen under the NYUAD Research Institute Award CG013, the Center for Interacting Urban Networks (CITIES), funded by Tamkeen under the NYUAD Research Institute Award CG001, and the Center for Artificial Intelligence and Robotics (CAIR), funded by Tamkeen under the NYUAD Research Institute Award CG010. The authors would like to thank William Fulton and Mohamed Rasheed from the Campus Planning and Projects Office at NYUAD for their support and for allowing access to ongoing projects on campus to be used for our research.

References

[1] M. ElGindi, “User friendly progress reporting system for construction projects,” *Theses Diss.*, Jun. 2017, [Online]. Available: https://fount.aucegypt.edu/etds/383

[2] W. N. L. Lamptey and A. R. Fayek, “Developing a Project Status Dashboard for Construction Project Progress Reporting,” *Int. J. Archit. Eng. Constr.*, pp. 112–120, 2012, doi: 10.7492/IJAEC.2012.013.

[3] S. El-Omari and O. Moselhi, “Integrating automated data acquisition technologies for progress reporting of construction projects,” *Autom. Constr.*, vol. 20, no. 6, pp. 699–705, Oct. 2011, doi: 10.1016/j.autcon.2010.12.001.

[4] L. Koskela, G. Howell, G. Ballard, and I. Tommelein, "The foundations of lean construction," in *Design and Construction*, R. Best and G. de Valence, Eds., Abingdon & New York: Routledge, 2002, pp. 211–226. Accessed: Dec. 22, 2024. [Online]. Available: http://www.scopus.com/inward/record.url?scp=85056052718&partnerID=8YFLogxK

[5] G. Ballard, F. Hamzeh, and I. Tommelein, "The Last Planner Production Workbook-Improving Reliability in Planning and Workflow," *Lean Constr. Inst. San Franc. Calif. USA 81pp*, Jan. 2007.

[6] F. Hamzeh, *The Lean Journey: Implementing The Last Planner ® System in Construction*. 2011. doi: 10.13140/RG.2.1.3648.7522.

[7] O. Kassab, B. Young, and O. Lædre, *Implementation of Last Planner® System in an Infrastructure Project*. 2020, p. 528. doi: 10.24928/2020/0089.

[8] D. Shamsollahi, O. Moselhi, and K. Khorasani, "Construction Progress Monitoring and Reporting using Computer Vision Techniques – A Review," in *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, Waterloo, Canada: IAARC Publications, 2022, pp. 467–474. Accessed: Dec. 16, 2024. [Online]. Available: https://www.proquest.com/docview/2711558137/abstract/50A6AD6A28624C17PQ/1

[9] S. El-Omari and O. Moselhi, "Integrating automated data acquisition technologies for progress reporting of construction projects," *Autom. Constr.*, vol. 20, no. 6, pp. 699–705, Oct. 2011, doi: 10.1016/j.autcon.2010.12.001.

[10] K. J. Shrestha and H. D. Jeong, "Computational algorithm to automate as-built schedule development using digital daily work reports," *Autom. Constr.*, vol. 84, pp. 315–322, Dec. 2017, doi: 10.1016/j.autcon.2017.09.008.

[11] P. Jafari, M. Al Hattab, E. Mohamed, and S. AbouRizk, "Automated Extraction and Time-Cost Prediction of Contractual Reporting Requirements in Construction Using Natural Language Processing and Simulation," *Appl. Sci.*, vol. 11, no. 13, Art. no. 13, Jan. 2021, doi: 10.3390/app11136188.

[12] B. Xiao, Y. Wang, Y. Zhang, C. Chen, and A. Darko, "Automated daily report generation from construction videos using ChatGPT and computer vision," *Autom. Constr.*, vol. 168, p. 105874, Dec. 2024, doi: 10.1016/j.autcon.2024.105874.

[13] D. Shamsollahi, O. Moselhi, and K. Khorasani, "Data integration using deep learning and real-time locating system (RTLS) for automated construction progress monitoring and reporting," *Autom. Constr.*, vol. 168, p. 105778, Dec. 2024, doi: 10.1016/j.autcon.2024.105778.

[14] B. Ekanayake, J. K. W. Wong, A. A. F. Fini, P. Smith, and V. Thengane, "Deep learning-based computer vision in project management: Automating indoor construction progress monitoring," *Proj. Leadersh. Soc.*, vol. 5, p. 100149, Dec. 2024, doi: 10.1016/j.plas.2024.100149.

[15] K. Mostafa and T. Hegazy, "Review of image-based analysis and applications in construction," *Autom. Constr.*, vol. 122, p. 103516, Feb. 2021, doi: 10.1016/j.autcon.2020.103516.

[16] P. Martinez, B. Barkokebas, F. Hamzeh, M. Al-Hussein, and R. Ahmad, "A vision-based approach for automatic progress tracking of floor paneling in offsite construction facilities," *Autom. Constr.*, vol. 125, p. 103620, May 2021, doi: 10.1016/j.autcon.2021.103620.

[17] W. Wei, Y. Lu, T. Zhong, P. Li, and B. Liu, "Integrated vision-based automated progress monitoring of indoor construction using mask region-based convolutional neural networks and BIM," *Autom. Constr.*, vol. 140, p. 104327, 2022.

[18] W. Wei, Y. Lu, X. Zhang, B. Wang, and Y. Lin, "Fine-grained progress tracking of prefabricated construction based on component segmentation," *Autom. Constr.*, vol. 160, p. 105329, 2024.

[19] F. Bordes *et al.*, "An Introduction to Vision-Language Modeling," May 27, 2024, *arXiv*: arXiv:2405.17247. doi: 10.48550/arXiv.2405.17247.

[20] W. L. Tsai, J. J. Lin, and S.-H. Hsieh, "Generating Construction Safety Observations via CLIP-Based Image-Language Embedding," in *Computer Vision – ECCV 2022 Workshops*, L. Karlinsky, T. Michaeli, and K. Nishino, Eds., Cham: Springer Nature Switzerland, 2023, pp. 366–381. doi: 10.1007/978-3-031-25082-8_24.

[21] S. A. Prieto, E. T. Mengiste, and B. García de Soto, "Investigating the Use of ChatGPT for the Scheduling of Construction Projects," *Buildings*, vol. 13, no. 4, p. 857, Mar. 2023, doi: 10.3390/buildings13040857.

[22] H. Pu, X. Yang, J. Li, and R. Guo, "AutoRepo: A general framework for multimodal LLM-based automated construction reporting," *Expert Syst. Appl.*, vol. 255, p. 124601, Dec. 2024, doi: 10.1016/j.eswa.2024.124601.

[23] OMG, "About the Business Process Model and Notation Specification Version 2.0.2." Accessed: Aug. 27, 2024. [Online]. Available: https://www.omg.org/spec/BPMN

[24] n8n, "Powerful Workflow Automation Software & Tools - n8n." Accessed: Dec. 23, 2024. [Online]. Available: https://n8n.io/

How Reliable Are Large Language Models? Zero-Shot Detection of Construction Hazards

Nishi Chaudhary[1], S M Jamil Uddin[1], Mahzabin Tamanna[2], Alex Albert[3], Abdur Rahman Bin Shahid[4]

[1]Department of Construction Management, Colorado State University, USA
[2]Department of Computer Science, North Carolina State University, USA
[3]Department of Civil, Construction, and Environmental Engineering, North Carolina State University, USA
[4]School of Computing, Southern Illinois University, USA
nishi.chaudhary@colostate.edu, smj.uddin@colostate.edu, mtamann@ncsu.edu, alex_albert@ncsu.edu, shahid@cs.siu.edu

Abstract – The construction industry persistently underperforms in hazard recognition, often leading to severe workplace injuries due to unrecognized hazards. With the recent advancements in Artificial Intelligence (AI) and the emergence of Large Language Models (LLM), the construction sector has begun exploring these technologies for various applications. However, a systematic comparison of popular LLMs to evaluate their effectiveness in identifying construction hazards remains unexplored. Additionally, previous studies have primarily focused on assessing LLMs using textual input and output, leaving their performance with visual inputs underexplored. This study addresses this gap by systematically assessing and comparing the hazard recognition performance of five widely used LLMs using construction case images. The findings establish a baseline standard for LLMs in construction hazard identification through zero-shot learning and reveal that LLMs do not perform significantly well in this context. Additionally, the study provides valuable insights into the reliability and potential applications of LLMs for enhancing hazard recognition in the construction industry.

Keywords – LLM, Hazard Recognition, Construction Safety, Image Analysis

1 Introduction

The construction industry plays a pivotal role in developing and maintaining the infrastructure, thereby shaping broader society [1]. Despite its importance, the construction industry reports a disproportionate number of safety incidents around the world [2]. These incidents often result in fatal and non-fatal injuries to the workers. In the United States alone, approximately 1,000 fatal injuries take place every year and the number go well beyond 200,000 for non-fatal injuries [3]. In addition to the fatal and non-fatal injuries, these incidents also cause a significant amount of economic damage [4].

Over the years, researchers and industry professionals have put much effort into investigating the reasons behind these safety incidents. Evidence suggests that one of the main reasons for these incidents is the failure to identify construction hazards [5, 6]. Studies show that the construction industry performs poorly in recognizing workplace hazards, with over 70% of work-related incidents attributed to poor hazard recognition [7].

Several tools have been developed over the years to improve the hazards recognition efforts in the construction industry. For example, job hazard analysis (JHA) tool is used to catalog hazards that are associated with specific construction tasks. However, this method assumes that workers possess a level of proficiency in identifying hazards, which in reality is not the case [8, 9]. On the other hand, tools such as safety checklists fail to contribute due to their limitation of identifying a limited number of hazards [10]. A number of studies have demonstrated that, despite the usage of such safety tools, over 40% of construction hazards remain unrecognized in the construction workplaces and they in turn cause these fatal and non-fatal incidents [11, 12].

On the technological front, several other tools have been developed to aid hazard recognition over the years. For example, usage of Building Information Modeling (BIM) has been popular among researchers and industry professionals [13]. BIM enables collaboration, visualization, and enhances the hazard identification process. Other efforts have focused on technologies such as eye tracking, augmented reality (AR), and virtual reality (VR) environments [14, 15]. While these technological solutions aid the hazard recognition process, they have their own limitations. For instance, these technology-driven tools require significant expertise, development time, resources, and financial investments. These barriers often hinder widespread

adoption of these technological solutions.

In recent years with the development and uprising of Artificial Intelligence (AI) and Large Language Models (LLM), the construction industry has also leaned towards adopting these solutions to improve the safety situation of construction workplaces. Different studies have demonstrated how some popular LLMs, such as ChatGPT, can be beneficial in improving construction safety [16-19]. In light of recent advancements, this study focuses on evaluating the capabilities of five different LLMs in identifying construction hazards using zero-shot learning. This evaluation aims to establish a baseline standard for LLM performance with no prior training, additional knowledge base, or guidelines, and offer insights on the reliability of LLMs in recognizing construction hazards.

2 Use of LLM in Construction Safety

In recent years, Large Language Models (LLM) have gained significant popularity across different industries and domains. Construction industry is no different from others. Although the industry is in its very early stages of adopting LLMs on a full scale for diverse applications, several studies have explored the opportunities these LLMs present. For example, a number of studies explored the possibility of integrating LLM and BIM to support information retrieval from the building models [19-21]. Other studies have focused on leveraging different LLMs for project management tasks such as automated sequence planning [22], generating construction schedules [23], automated classification of contractual risk clauses [24], automatic matching of look ahead planning tasks [25] etc. Additionally, some studies focused on how to effectively use LLMs to improve the construction education outcomes [16, 26, 27].

On the safety front, several studies have examined the usability of LLMs to improve the health and safety condition of construction workplaces. For example, Uddin et al. [18] conducted a controlled experiment with 42 construction engineering students. Their effort demonstrated that LLMs can be particularly effective in aiding hazard recognition efforts. They also suggest that LLMs can be integrated as part of safety education for construction students albeit with caution. Another study by Uddin et al. [17] explored the usability of LLM in aiding construction hazard prevention through design efforts. The study demonstrated that LLM can improve the hazard recognition efforts during the design phase by approximately 40%. Wang et al. [28] evaluated LLM's ability to extract causal factors from construction accident reports. The study found that LLM can perform well as an assisting tool, offering clear and reliable insights, but it still requires further development for professional applications like crane safety. The research highlights the potential of LLMs in construction while emphasizing the need for refinement to enhance their practical utility.

Smetana et al. [29] leveraged an LLM model to analyze textual data from OSHA's Severe Injury Reports (SIR) for highway construction accidents. Using advanced NLP techniques, clustering, and LLM prompting, they identified major accident types, including heat-related and struck-by injuries, while uncovering commonalities between incidents. The findings demonstrate the potential of AI and LLMs to enhance data-driven safety analysis and support the development of more effective prevention strategies in the highway construction industry. Hussain et al. [30] developed a virtual reality-based safety training system incorporating LLM as a live AI instructor to address communication barriers and trainer limitations, particularly for migrant workers. Testing across five countries showed a 23% improvement in knowledge scores, demonstrating the system's effectiveness. The research highlights the system's potential to improve safety training globally, reduce construction site accidents, and advance immersive and AI-driven training methodologies.

While the existing studies have demonstrated the potential of LLMs in various aspects of construction including workers' health and safety, they predominantly focused on evaluating a single LLM and relied exclusively on textual inputs and outputs. None of these efforts explored a comparative analysis of multiple LLMs to assess their relative effectiveness in achieving safety outcomes. Additionally, the studies did not investigate the use of image-based inputs for hazard recognition, despite the growing capabilities of modern LLMs to process and interpret visual data.

With recent advancements in LLMs enabling them to analyze multimodal inputs, including images, it becomes crucial to evaluate whether these models can effectively extract safety-critical information from construction images and identify hazards. Such an investigation would address a significant gap in the literature and provide insights into the broader applicability of LLMs for improving construction safety practices.

Hence, this study focuses on assessing five different LLMs' capability of identifying potential hazards from construction case images and then conducting a comparative analysis to demonstrate if all the LLMs are reliable and if they perform equally in identifying potential hazards from construction case images.

3 Methodology

To achieve the goals of this study, it was necessary to establish a ground truth for evaluating the hazard recognition capabilities of the LLMs. This was

accomplished by using a dataset of 16 different images of construction activities captured as part of a previous study [31]. These images depict a wide variety of construction activities such as welding, cutting, drilling, crane rigging among others.

Once gathered, these images were then analyzed by a panel of 17 construction industry safety experts. The collective experience of this panel was over 300 years at the time of this assessment. The expert panel was tasked with analyzing the images and identifying all the potential safety hazards for each image. One example image with annotated safety hazards is shown in Figure 1. The expert panel identified a total of 120 construction hazards from these 16 case images. These pre-identified hazards are going to serve as ground truth for our LLM assessment experiment.



Figure 1. Example image of construction activity with pre-identified hazards.

Next, we chose five LLMS, i.e., ChatGPT 4, GPT 4o, GPT o1, Gemini 2.0, and Claude for this experiment. The selection of these five LLMs was based on their industry adoption [17, 32, 33], multimodal capabilities [34, 35], architectural diversity [36], and expected performance variation. These models are widely used and include multimodal processing which is crucial for hazard recognition from construction case images. The study specifically leveraged the zero-shot learning capabilities of these models, allowing them to identify hazards without requiring domain-specific training or environment customization. This approach was deemed particularly relevant to assess the baseline performance of these LLMs without any training or additional knowledge input.

To ensure consistency and standardization, a systematic approach was developed to input all the images into these LLMs for hazard identification. First, the following prompt was provided:

"You're a construction safety expert. You will be provided with 16 different construction images. You will have to analyze these images and identify the potential safety hazards for each image."

Then each image was fed into to LLM chat prompt along with the following instruction:

"Identify all potential safety hazards for the construction activity in this image."

No additional contextual or domain-specific prompts were provided, ensuring the zero-shot learning approach remained intact. This standardized methodology facilitated comparability of outputs across all the models.

Once the LLMs returned their results for each image, the hazards were then recorded carefully by the research team ensuring traceability and consistency for subsequent analysis. The process was conducted under controlled experimental conditions, ensuring that all models were evaluated using identical inputs and queries.

The output generated by the LLMs were then compared against the ground truth dataset to evaluate their hazard recognition performance. Three different metrics were used to evaluate the performance of these LLMs, i.e., Precision, Recall, and F1-score [37, 38]. Precision score denotes the proportion of correctly identified hazards relative to the total hazards identified by the model. Recall is the proportion of correctly identified hazards relative to the total hazards present in the ground truth dataset. And F1-score is the harmonic mean of precision and recall, providing a balanced measure of accuracy and completeness of the models.

4 Data Analysis and Results

In order to measure the Precision, Recall, and F1-Score of each LLM, we used equations 1, 2, and 3 respectively for each image.

$$Precision_i = \frac{TP_i}{TP_i + FP_i} \quad (1)$$

$$Recall_i = \frac{TP_i}{TP_i + FN_i} \quad (2)$$

$$F1 - Score_i = 2 \times \frac{Precision_i \times Recall_i}{Precision_i + Recall_i} \quad (3)$$

Where:

TP_i = Hazards correctly identified by each LLM
FP_i = Hazards incorrectly identified by each LLM
FN_i = Hazards that exist in the ground truth but were missed by each LLM

Once the Precision, Recall, and F1-Score for all 16 images were gathered, then they were averaged to get the mean Precision, Recall, and F1-Score using equations 4, 5, and 6 where n is equal to 16.

$$Avg.Precision = \frac{1}{n}\sum_{i=1}^{n} Precision_i \quad (4)$$

$$Avg.Recall = \frac{1}{n}\sum_{i=1}^{n} Recall_i \quad (5)$$

$$Avg.F1-Score = \frac{1}{n}\sum_{i=1}^{n} F1-Score_i \quad (6)$$

Table 1 below shows the summarized results for the five LLM models that were assessed.

Table 1 Summary of LLM Performance

Models	Precision	Recall	F1-Score
ChatGPT-4	0.30	0.29	0.30
GPT4o	0.25	0.30	0.27
GPTo1	0.27	0.30	0.27
Gemini 2.0	0.31	0.33	0.31
Claude	0.42	0.29	0.34
Overall	0.31	0.30	0.30

4.1 ChatGPT-4

As can be seen in Table 1, ChatGPT-4 achieved 0.30, 0.29, and 0.30 scores in Precision, Recall, and F1-Score metrics respectively. A precision score of 0.30 indicates that only 30% of the hazards identified by the model were accurate, which also highlights a significant proportion of false positives. This indicates that ChatGPT-4's level of precision in recognizing construction hazards as a standalone tool may not be very reliable. Similarly, the recall score of 0.29 demonstrates the model's inability to detect the majority of the hazards present in the actual scenarios. The F1-score of 0.30 shows that the model struggles to achieve both accuracy and completeness in construction hazard identification from images.

4.2 GPT4o

GPT4o's performance in construction hazard recognition from the images is similar to GPT4's. GPT4o's precision score of 0.25 indicates that only 25% of the identified hazards were accurate, reflecting a higher rate of false positives. Additionally, the recall score of 0.30 shows that the model was able to identify only 30% of the actual hazards in the test scenario. The F1-score of 0.27 indicates that this model also struggles to identify construction hazards on its own from the construction images.

4.3 GPTo1

GPTo1 model performed rather poorly in construction hazard identification in this study. GPTo1 achieved 0.27 in precision which demonstrates a poor performance in recognizing hazards accurately, with a higher level of false positives. The recall score of 0.30, which is consistent with GPT4 and GPT4o, shows that the model was able to identify 30% of the actual hazards present in the test scenarios. This underscores that a significant proportion of hazards went unrecognized. The F1-score of 0.27, which is similar to GPT4o, highlights the model's limitations in recognizing construction hazards successfully.

4.4 Gemini 2.0

Gemini 2.0 with precision score of 0.31, recall score of 0.33, and F1-score of 0.31 demonstrated a slight improvement over the other three LLMs. The precision score indicates that 31% of the hazards were identified accurately. While the recall score of 0.33 demonstrates a higher percentage of correctly identified hazards from the hazards present in the ground truth dataset. The F1-score of 0.33 shows the balanced performance between precision and recall for this model.

4.5 Claude

Claude's performance in construction hazard recognition with a precision score of 0.42 demonstrates its ability to identify hazards with relatively higher accuracy, which is the highest among other models tested in this study. This score also highlights a significantly lower rate of false positives. However, a lower recall score of 0.29 indicates that Claude identified only 29% of the actual hazards from the ground truth dataset. While its precision ensures that the hazards identified are likely correct, the lower recall suggests that many actual hazards remain unrecognized, which could be a concern in environments requiring exhaustive risk identification.

4.6 Trend of Hazard Identification

Along with assessing different LLMs' hazard recognition performance, it is also important to examine if there's a common trend demonstrated in construction hazard recognition. To achieve this, we aggregated all the hazards that were correctly identified by the LLMs, referred to as True Positives (TP), to gain insight into which hazards were most successfully recognized. We also summed up all the False Negatives (FN) to identify the hazards that were least frequently recognized by the LLMs. Figure 2 and figure 3 below show the top five and bottom five hazards identified by the LLMs.

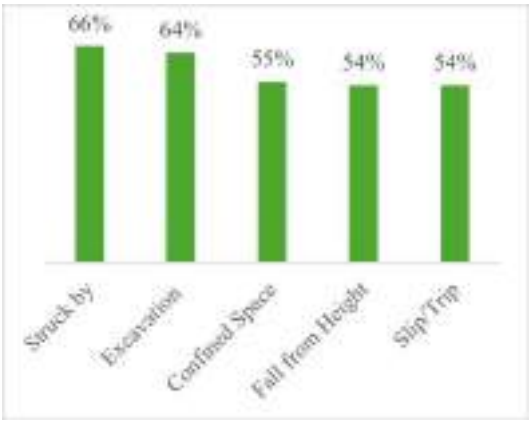


Figure 2. Top five hazards identified by the LLMs.

As can be seen in figure 2, the LLMs overall performed better in identifying more common hazards such as struck by, excavation, confined space, and fall hazards.

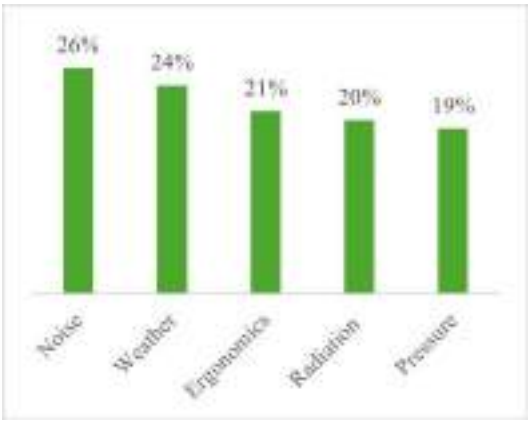


Figure 3. Least five hazards identified by LLMs.

On the other hand, figure 3 demonstrates that the LLMs performed rather poorly in identifying hazards such as noise, weather, ergonomics, radiation, and pressure. This finding is consistent with previous research efforts where the studies evaluated the hazard recognition trend of the construction workforce [12, 39].

5 Discussion

The performance of the evaluated LLMs in construction hazard recognition from images under zero-shot learning conditions reveals a varied outcome across all the measured metrics. Among the five tested models, Claude showed the highest precision level in identifying construction hazards. This suggests that Claude minimizes false positives better than the other models, making it particularly suitable for applications where avoiding unnecessary interventions is critical. However, the lower recall value of Claude indicates the failure to identify all the hazards present in the ground truth dataset. On the other hand, Gemini 2.0 achieved the highest recall score and maintained a balanced F1-score. Compared to Claude, Gemini 2.0 sacrifices some precision but compensates with better recall, making it a more reliable choice for scenarios requiring broader hazard identification.

The superior performance of Claude in precision may be attributed to its underlying architecture, which possibly favors conservative predictions by avoiding overgeneralizations. Meanwhile, Gemini 2.0's balanced performance could stem from a more robust generalization capability, enabling it to identify a larger proportion of hazards without excessively sacrificing precision.

The remaining three models, ChatGPT-4, GPT4o, and GPT-o1 exhibited comparable performance, with precision ranging from 0.25 to 0.30 and recall clustering around 0.30. These models struggled to achieve a balance between precision and recall, with their F1 scores hovering at 0.27 to 0.30. Notably, ChatGPT-4 demonstrated the highest F1 score among this group at 0.30, indicating a slightly better overall performance in balancing accuracy and coverage. The relatively lower precision of GPT4o (0.25) suggests a greater tendency toward false positives, while GPT-o1 showed a slight improvement in reducing these errors.

The overall average metrics across all models, precision of 0.31, recall of 0.30, and F1 score of 0.30, highlight a consistent challenge for LLMs in performing hazard recognition tasks as standalone tools under zero-shot learning conditions. However, it is worth noting that the hazard recognition performance of LLMs using zero-shot learning is comparable to that of human participants with no prior intervention or training, as observed in previous studies [12, 31, 40].

The variation in precision, recall, and F1-score across LLMs can be attributed to differences in model architecture, training data, and inference strategies. For example, Claude exhibits high precision but low recall as it applies conservative thresholds, flagging only high-confidence hazards to minimize false positives. In contrast, Gemini 2.0, designed as a multimodal AI, prioritizes higher recall over precision, leading to more hazard detections but also increased false positives. The GPT models from OpenAI are balanced models that achieve moderate precision and recall, making them consistent but not specialized in hazard detection.

The results obtained from this experiment indicate that LLMs face significant challenges in construction hazard recognition from images when they are used without domain specific knowledge, fine tuning, or specific guidance. The findings contrast with previous studies that showed LLMs perform exceedingly well when they are integrated into safety training or education contexts for construction professionals. Previous studies

used LLMs to support humans by generating insights or suggesting hazards, leveraging human reasoning to fill gaps in understanding [17, 18]. While on the other hand, the current zero-shot learning setup relied entirely on the models' ability to independently interpret the construction images and identify hazards. This exposes the limitations of the LLMs' specific knowledge base.

These findings of this study demonstrate that, while LLMs perform remarkably well with structured prompts, and additional knowledge base, without these interventions and without human assistance, LLM's can perform rather poorly in construction hazard recognition.

6 Conclusion

Construction hazard recognition is one of the pivotal steps in order to ensure a safer construction environment. Previous studies have demonstrated that construction hazards often remain unrecognized in the workplace which translates to both fatal and non-fatal accidents. In order to tackle this issue, numerous studies have focused on developing and adopting various tools to help with hazard recognition.

With the advancement of modern technology, Artificial Intelligence (AI), and Large Language Models (LLM), researchers and practitioners have started to utilize these technological solutions to improve construction hazard recognition. However, since there are multiple different LLM platforms available, it was necessary to conduct a comparative analysis to see if they are reliable and which model performs better than others. Additionally, it was also important to assess these models' performance with zero-shot learning, in other words, with no enforced knowledge.

This study was designed carefully to evaluate five different LLMs performance in construction hazard recognition from construction case images. The findings of this study reveal that LLMs perform rather poorly in identifying construction hazards with zero-shot learning as standalone tool. However, the hazard recognition trend demonstrates that the LLMs' performances are not too different from the human subjects from construction industry. LLMs show a similar trend in identifying common hazards and missing out uncommon hazards, similar to previous other studies [12, 39, 41].

The study highlights that although LLMs show a promise in identifying construction hazards, they may require some input, training, additional knowledge base, interventions, and human assistance to perform better as previous studies have demonstrated an improved hazard recognition performance by human subjects with assistance from different LLMs [17, 18].

In future work, we plan to explore the impact of fine-tuning LLMs using larger construction-specific datasets to assess their potential for improved hazard recognition. While our current study leveraged zero-shot learning to establish baseline performance, real-world applications would likely require models that have been trained on domain-specific data. We plan to expand our analysis to include a detailed examination of false positives and false negatives for each model, providing deeper insight into the specific types of hazards that LLMs fail to recognize or misclassify. This will help identify recurring patterns in model errors and guide improvements in hazard recognition accuracy through targeted fine-tuning and dataset enhancements.

Additionally, building on our analysis of hazard identification trends in Section 4.6, future work will incorporate qualitative insights from construction safety experts to better understand the nature of LLM errors. While our study quantitatively identified the types of hazards frequently missed by LLMs, expert consultation will provide deeper context on why certain hazards are overlooked and how models can be improved. Specifically, we will work with industry professionals, site safety managers, and researchers to identify common misclassifications, potential biases in AI hazard detection, and effective strategies for model training.

References

[1] Labor, U.S.D.o., *Civil Engineers : Occupational Outlook Handbook: : U.S. Bureau of Labor Statistics*, in *2021*. 2021.

[2] Statistics, U.S.B.o.L., *Construction and Extraction Occupations : Occupational Outlook Handbook: : U.S. Bureau of Labor Statistics*. 2021.

[3] Bls, *Injuries, Illnesses, and Fatalities*, in *U.S. Bureau of labor Statistics*. 2021.

[4] Waehrer, G.M., et al., *Costs of occupational injuries in construction in the United States.* Accident Analysis and Prevention, 2007. **39**(6): p. 1258-1266.

[5] Perlman, A., R. Sacks, and R. Barak, *Hazard recognition and risk perception in construction.* Safety Science, 2014. **64**: p. 22-31.

[6] Carter, G. and S.D. Smith, *Safety Hazard Identification on Construction Projects.* JOURNAL OF CONSTRUCTION ENGINEERING AND MANAGEMENT, 2006. **132**(2): p. 197-205.

[7] Haslam, R.A., et al., *Contributing factors in construction accidents.* Applied Ergonomics, 2005. **36**: p. 401-415.

[8] Jeelani, I., A. Albert, and J.A. Gambatese, *Why Do Construction Hazards Remain Unrecognized at the Work Interface?* Journal of Construction Engineering and Management, 2017. **143**(5): p. 1-10.

[9] Rozenfeld, O., et al., *Construction Job Safety Analysis.* Safety Science, 2010. **48**(4): p. 491-498.
[10] Guo, B.H., T.W. Yiu, and V.A. González, *Predicting safety behavior in the construction industry: Development and test of an integrative model.* Safety science, 2016. **84**: p. 1-11.
[11] Jeelani, I., et al., *Development and Testing of a Personalized Hazard-Recognition Training Intervention.* Journal of Construction Engineering and Management, 2017. **143**(5): p. 04016120-04016120.
[12] Uddin, S.M.J., et al., *Hazard Recognition Patterns Demonstrated by Construction Workers.* International Journal of Environmental Research and Public Health, 2020. **17**(21): p. 7788-7788.
[13] Kim, I., Y. Lee, and J. Choi, *BIM-based hazard recognition and evaluation methodology for automating construction site risk assessment.* Applied Sciences (Switzerland), 2020. **10**(7).
[14] Jeelani, I., et al., *Are Visual Search Patterns Predictive of Hazard Recognition Performance? Empirical Investigation Using Eye-Tracking Technology.* Journal of Construction Engineering and Management, 2019. **145**(1): p. 04018115-04018115.
[15] Li, X., et al., *A critical review of virtual and augmented reality (VR/AR) applications in construction safety.* Automation in Construction, 2018. **86**: p. 150-162.
[16] Uddin, S.M.J., et al., *ChatGPT as an Educational Resource for Civil Engineering Students.* Computer Applications in Engineering Education, 2024.
[17] Uddin, S.M.J., A. Albert, and M. Tamanna, *Harnessing the power of ChatGPT to promote Construction Hazard Prevention through Design (CHPtD).* Engineering, Construction and Architectural Management, 2024.
[18] Uddin, S.M.J., et al., *Leveraging ChatGPT to Aid Construction Hazard Recognition and Support Safety Education and Training.* Sustainability, 2023. **15**(9): p. 7121-7121.
[19] Zheng, J. and M. Fischer, *BIM-GPT: a Prompt-Based Virtual Assistant Framework for BIM Information Retrieval.* 2023/04/18.
[20] Rane, N., S. Choudhary, and J. Rane, *Integrating Building Information Modelling (BIM) with ChatGPT, Bard, and similar generative artificial intelligence in the architecture, engineering, and construction industry: applications, a novel framework, challenges, and future scope.* Bard, and similar generative artificial intelligence in the architecture, engineering, and construction industry: applications, a novel framework, challenges, and future scope (November 22, 2023), 2023.
[21] Zheng, J. and M. Fischer, *Dynamic prompt-based virtual assistant framework for BIM information search.* Automation in Construction, 2023. **155**: p. 105067.
[22] You, H., et al., *Robot-enabled construction assembly with automated sequence planning based on ChatGPT: RoboGPT.* Buildings, 2023. **13**(7): p. 1772.
[23] Prieto, S.A., E.T. Mengiste, and B. García de Soto, *Investigating the Use of ChatGPT for the Scheduling of Construction Projects.* Buildings, 2023. **13**(4): p. 857-857.
[24] Moon, S., S. Chi, and S.-B. Im, *Automated detection of contractual risk clauses from construction specifications using bidirectional encoder representations from transformers (BERT).* Automation in Construction, 2022. **142**: p. 104465.
[25] Amer, F., Y. Jung, and M. Golparvar-Fard, *Transformer machine learning language model for auto-alignment of long-term and short-term plans in construction.* Automation in Construction, 2021. **132**: p. 103929.
[26] Abril, D.E., M.A. Guerra, and S.D. Ballen. *ChatGPT to Support Critical Thinking in Construction-Management Students*. in *2024 ASEE Annual Conference & Exposition*. 2024.
[27] Zhao, T., et al., *Impact of ChatGPT on Student Writing in Construction Management: Analyzing Literature and Countermeasures for Writing Intensive Courses.* Proceedings of 60th Annual Associated Schools, 2024. **5**: p. 339-348.
[28] Wang, Y., et al. *Investigating the Potential of ChatGPT in Construction Management: A Study of Interpreting Construction Crane-Related Accident Reports*. in *International Symposium on Advancement of Construction Management and Real Estate*. 2023. Springer.
[29] Smetana, M., et al., *Highway Construction Safety Analysis Using Large Language Models.* Applied Sciences, 2024. **14**(4): p. 1352.
[30] Hussain, R., et al., *Conversational AI-based VR system to improve construction safety training of migrant workers.* Automation in Construction, 2024. **160**: p. 105315.
[31] Albert, A., M.R. Hallowell, and B.M. Kleiner, *Enhancing Construction Hazard Recognition and Communication with Energy-Based Cognitive Mnemonics and Safety Meeting Maturity Model: Multiple Baseline Study.* Journal of Construction Engineering and Management, 2014. **140**(2): p. 04013042-04013042.
[32] Genc, O. and O. Genc, *Assessing the role of AI in advancing construction sector industrial symbiosis research: a comparative study of leading digital*

assistants. Environment, Development and Sustainability 2024, 2024-12-04.
[33] Ahmadi, E., S. Muley, and C. Wang, *Automatic construction accident report analysis using large language models (LLMs).* Journal of Intelligent Construction, 2025/1/1. **3**(1).
[34] OpenAI. *ChatGPT can now see, hear, and speak*. 2023 [cited 2025 3/12/2025]; Available from: https://openai.com/index/chatgpt-can-now-see-hear-and-speak/.
[35] Claude. *Introducing the next generation of Claude*. 2024 [cited 2025 3/12/2025]; Available from: https://www.anthropic.com/news/claude-3-family.
[36] Naveed, H., et al., *A comprehensive overview of large language models.* arXiv preprint arXiv:2307.06435, 2023.
[37] Yacouby, R. and D. Axman, *Probabilistic Extension of Precision, Recall, and F1 Score for More Thorough Evaluation of Classification Models.* Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems, 2020/11.
[38] Orozco-Arias, S., et al., *Measuring Performance Metrics of Machine Learning Algorithms for Detecting and Classifying Transposable Elements.* Processes 2020, Vol. 8, Page 638, 2020-05-27. **8**(6).
[39] Albert, A., B. Pandit, and Y. Patil, *Focus on the fatal-four: Implications for construction hazard recognition.* Safety Science, 2020. **128**(August 2019): p. 104774-104774.
[40] Albert, A., M.R. Hallowell, and B.M. Kleiner, *Enhancing construction hazard recognition and communication with energy-based cognitive mnemonics and safety meeting maturity model: Multiple baseline study.* Journal of Construction Engineering and Management, 2014. **140**(2).
[41] Albert, A., et al., *Empirical measurement and improvement of hazard recognition skill.* Safety Science, 2017. **93**: p. 1-8.

Implementation of the Cybersecurity Incident Severity Scale (CISS) to Assess Cyber Incidents in the Construction Sector

Dongchi Yao[1], Bharadwaj R. K. Mantha[2], Borja García de Soto[1]

[1]S.M.A.R.T. Construction Research Group, Division of Engineering, New York University Abu Dhabi (NYUAD), Experimental Research Building, Saadiyat Island, P.O. Box 129188, Abu Dhabi, United Arab Emirates
[2]Department of Civil and Environmental Engineering, University of Sharjah, Sharjah, UAE
dongchi.yao@nyu.edu, rmantha@sharjah.ac.ae, garcia.de.soto@nyu.edu

Abstract –

As the construction industry adopts digital technologies, cybersecurity risks are rising. However, the absence of a standardized incident reporting framework has resulted in limited disclosure of cybersecurity incidents and a lack of a centralized database. This prevents construction companies from learning from past events and developing effective cyber risk management strategies. To address this issue, this study applies the Cybersecurity Incident Severity Scale (CISS) model to assess the severity of cyber incidents within the construction sector. The CISS model uses a structured, semi-quantifiable approach to generate an integrated score, evaluating dimensions such as safety and financial impacts to provide a comprehensive understanding of an incident's consequences. A real-world construction cyber incident serves as a case study to demonstrate the model's applicability. By promoting standardized reporting, the CISS model can improve data collection, ensure consistent reporting across organizations, and enable meaningful comparisons over time and across regions.

Keywords –

Cyber Incidents; Severity Score; Construction 4.0; Construction Projects; Cyber Attacks

1 Introduction

The Construction 4.0 integrates the Internet of Things (IoT), Building Information Modeling (BIM), artificial intelligence (AI), and other digital technologies. These have transformed the industry by improving efficiency and collaboration. Although phishing, ransomware, and other cyber risks are prevalent in every sector involving digitalization, the construction industry faces unique cybersecurity challenges. Construction projects often involve many stakeholders, such as owners, contractors, and suppliers, who rely on BIM and IoT to exchange data in real time. This high level of coordination enlarges the cyberattack surface because each partner or device might be an attack entry point. The supply chain's fragmentation, the tight project schedule, and the limited budget make it hard to implement unified cybersecurity standards and regulations for different parties. Cyberattacks not only lead to data loss but interrupt physical site operations, delay the project, and may cause physical and financial harm. Through IoT equipment and smart machines, increased connectivity exacerbates these vulnerabilities, impacting time, cost, and safety. As for the cloud-based BIM platform and collaborative tools based on sensitive data analysis, strong security is necessary. AI-driven analytics must be protected to avoid tampering. Furthermore, ongoing digitalization needs updates to protect emerging techniques and processes. Multiple stakeholders with different cybersecurity awareness, expertise, and standards make it complicated to maintain a unified defense [1].

To effectively manage these cyber incident risks, the construction sector must implement comprehensive, effective, and robust cybersecurity strategies informed by lessons from past incidents. However, such progress is limited, and one primary reason is the absence of a standardized incident reporting framework. Without such a framework, the disclosure and documentation of cyber incidents remain fragmented and inconsistent across various projects and organizations, leading to a lack of a centralized database. This also hinders the analysis of trends, the identification of recurring vulnerabilities, and the design of effective preventive measures. Consequently, the industry misses out on critical insights essential for developing cybersecurity practices to address existing and emerging threats [1].

Global cybersecurity incident reporting frameworks vary significantly, offering lessons for the construction sector. The EU's NIS2 Directive [2] mandates rapid notifications and uniform reporting across member states, while Singapore's Cybersecurity Act [3] imposes a two-hour window for incident reporting. China's draft measures emphasize quick initial and follow-up reports [4] and Australia and India also enforce stringent

timelines to safeguard critical infrastructure [5]. In the United States, the Cyber Incident Reporting for Critical Infrastructure Act [6] aims to unify reporting across sectors. Despite the push for standardization, current frameworks rely on qualitative evaluations that overlook the complex threat landscape in construction. Without standardized quantitative metrics, comparing data across projects remains difficult. This limits thorough data collection and obscures key vulnerabilities. A comprehensive, quantitative approach could enhance security status and strengthen cross-sector collaboration.

To address this challenge, Conard (2024) proposed the Cybersecurity Incident Severity Scale (CISS) [7] to provide a semi-quantifiable and standardized metric for evaluating cybersecurity incident severity. We apply this approach to construction projects by utilizing it to assess and score the severity of cybersecurity incidents that uniquely impact this sector. By incorporating data related to the cybersecurity incident of the project as the inputs, this model derives a single severity score, making the assessment more evidence-based and objective. With a standardized severity score derived from uniform data inputs, stakeholders can quickly assess and learn from cybersecurity incidents, facilitating the sharing of materials and lessons across groups and organizations. This, in turn, enhances the overall efficiency of cybersecurity incident data preservation and collection.

2 Related Works

2.1 Industrial Cybersecurity Solutions

Cybersecurity is a rapidly evolving field, where practical knowledge and insights often originate from industry solutions. For instance, Willis Towers Watson's (WTW) "Cyber Quantified" platform [8] helps organizations model and quantify cyber risk exposures using insurance analytics and real-world benchmarks, while Marsh & McLennan's "Cyber Risk Analytics Center" [9] utilizes proprietary datasets to evaluate potential losses and guide mitigation strategies. Widely recognized annual reports such as Verizon's Data Breach Investigations Report (DBIR) [10] highlight patterns in attack vectors, adversary tactics, and breach trends. CrowdStrike's Global Threat Report [11] similarly maps the evolving threat landscape across diverse sectors. The Threat Landscape publications [12] of an EU commercial agency called ENISA offer industry-driven insights that complement academic literature. From a broader industrial perspective, frameworks such as the IAEA's guidance on "Computer Security for Nuclear Security" [13] demonstrate how other critical infrastructure sectors assess and manage risk severity. Collectively, these sources demonstrate the prevalence of real-world cyber threats and the necessity for practical assessment mechanisms. However, when it comes to the construction industry, research on cybersecurity, particularly the systematic scoring and comparison of incident severity, remains sparse. This indicates a clear need for deeper investigation within this sector.

2.2 Lack of Incident Severity Scoring Studies Within the Construction Industry

Awareness of cybersecurity within the construction industry, though increasing, remains relatively underdeveloped compared to other sectors. This is evidenced by the limited body of research dedicated to the topic. A 2023 scoping review by Pargoo and Ilbeigi [14] identified only 45 studies focused on cybersecurity within the construction industry, which can be broadly categorized into: discussions, reviews and solutions.

Among specific solutions, several studies focus on risk assessment of the entities (software, hardware, stakeholders, etc.) in construction projects. Mantha and García de Soto [15], for example, utilized agent-based models to examine how vulnerabilities spread among stakeholders in a construction network of projects. However, this remains at a very preliminary stage, lacking real-world case study validation, and the score assignment process is relatively subjective. In another study [16], they applied the CVSS to score stakeholders' vulnerabilities in construction networks based on the Confidentiality, Integrity, and Availability (CIA) Triad. Nevertheless, its manual and subjective score assignment process limits swift adaptability across different project contexts. Subsequently, Mantha et al. [17] enhanced their agent-based model from [15] by introducing seven more detailed steps to assess vulnerability spread among entities in construction projects, but the quantitative aspect has not significantly improved, and a large portion of the vulnerability score assignments remains non-data-driven. More recently, Yao and García de Soto [18] proposed a machine-learning-centric approach to assess the probability of cyber risks across project phases but lacked extensive case studies for validation.

Besides each study's limitations, they all focus on the vulnerability of entities and the probability of cybersecurity incidents to these entities within projects, failing to address the potential impact of incidents on the projects themselves. This gap highlights the need for comprehensive research on assessing the impact of cybersecurity incidents within construction projects.

2.3 Scoring Methods in Other Industries

Scoring methods have been widely adopted across diverse fields to quantify and communicate the severity of incidents, offering insights for robust cybersecurity scoring in construction. In the financial sector, the Data Envelopment Analysis (DEA)-based system [19]

synthesizes financial ratios into a "credibility score," helping institutions gauge credit risk and streamline decisions. Similarly, the IHS Markit Banking Risk Rating System (RRS) [20] employs a 0-100 scale to evaluate structural risk factors, detecting critical vulnerabilities early on. While quantitative, these approaches may overlook nuanced qualitative elements crucial for a holistic risk perspective. In public health, the NIH Stroke Scale [21] exemplifies a more granular clinical tool by assessing neurological function in stroke patients, but its specialized application underscores the challenges of designing universally applicable scoring systems. Transportation also illustrates how scoring or categorization frameworks must remain flexible. For example, the SAE Levels of Driving Automation [22] provide standardized automation categories but require ongoing adjustments as technology rapidly evolves.

In cybersecurity, the Common Vulnerability Scoring System (CVSS) [23] ranks software vulnerabilities on a 0–10 scale, guiding remediation priorities but potentially missing broader organizational effects. Insurance approaches like Willis Towers Watson's "Cyber Quantified" [8] translate threats into financial impacts using proprietary analytics, focusing on underwriting yet illustrating how data-driven methods can be adapted elsewhere, including construction. Across critical infrastructure, the U.S. National Cyber Incident Scoring System (NCISS) [24] developed by CISA is recognized as a best-practice for evaluating and prioritizing cyber incidents. It assigns severity ratings based on threat actor capabilities, potential critical infrastructure impacts, and disruption scale, offering a structured approach to triage. These scoring methods each contribute valuable methodologies for assessing various types of risks and incidents. However, their applicability to the construction sector, particularly for capturing the multi-stakeholder, project-based nature of construction projects, remains an open question, emphasizing the need for tailored or adapted scoring systems.

2.4 The CISS Model by Conard

In 2024, Conard [7] proposed the Cybersecurity Incident Severity Scale (CISS) model, drawing from established frameworks with potential applications in the construction sector for severity assessment. It aims to foster a uniform incident reporting scheme for cybersecurity incidents across various industries. The model integrates elements from scoring systems in other industries, focusing on standardized, detailed, and practical assessments. It evaluates factors such as data type, system criticality, individual harms, and organizational impacts, providing a comprehensive understanding of cybersecurity incidents. Additionally, the model must be adaptable to evolving risks and offer granular classifications for precision. International standardization, like those in transportation and public health, is crucial to support global collaboration and consistent cybersecurity threat mitigation. CISS is especially relevant for construction, where cybersecurity breaches can cause financial losses, operational disruptions, and stakeholder repercussions. Its holistic, semi-quantitative approach captures the sector's multi-dimensional nature, ensuring precise severity evaluations. Applying the CISS model would help address gaps in consistent incident reporting and severity assessment.

3 The CISS Model

For a more standardized cybersecurity incident severity assessment model in the construction industry, we considered Conard [7]'s CISS framework as a starting point because it provides a structured calculation methodology and includes critical factors such as financial repercussions, operational disruptions, and harm to individuals. Regarding the calculation method, the CISS model quantifies an incident's impact using a structured assessment pathway incorporating severity and criticality assessments. This is also useful for construction, where cybersecurity incidents may not only disrupt IT systems but also affect physical safety, project schedules, and contractual obligations. Regarding the factors, these align well with the key cybersecurity risks in construction, where incidents often impact supply chains, project timelines, collaborative digital platforms (e.g., BIM), and IoT-enabled smart construction sites. As a preliminary study, we apply the proven CISS framework to assess cyber incidents in construction. This helps ensure that our model is compatible with other sectors while demonstrating its practicality for the construction industry. However, we also recognize the need for future adaptations, such as integrating more granular construction-specific considerations due to the complexity of multi-stakeholder networks, supply chain vulnerabilities, and regulatory requirements.

3.1 Overview of the CISS Score

The overview of the scoring mechanism is shown in Figure 1. The CISS score (Y) is calculated using a weighted approach based on severity score(S), criticality score (C), and their weights β_1, β_2 using Equation 1.

$$Y = \beta_1 S + \beta_2 C \quad (1)$$

S is computed based on the financial score, individual score, and operational score, whereas C is measured based on the National Critical Function (NCF) score and National Security Impact (NSI) score. More information about each score is provided below. However, as the purpose of this paper is to implement and apply this scoring system within the context of construction, only a brief overview is included. The reader is encouraged to

refer to [7] for a more detailed description.

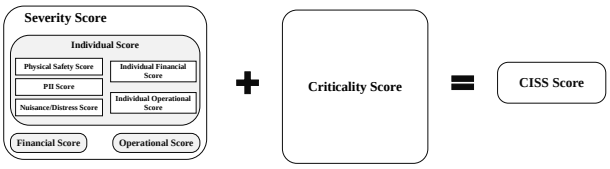


Figure 1. Overview of the CISS Scoring methodology adopted from [7]

Using the current estimation process, which incorporates a weighted approach and proposed scales, the overall CISS scores range from 0 to 10, where 0 indicates no impact, and 10 represents a very high impact. Table 1 outlines the various impact levels, corresponding score ranges, and their descriptions.

Table 1. Overview of different impact levels, corresponding score ranges, and description for CISS

Impact Level	Score Range	Description
No Impact	0	No observable impact or disruption to information and/or operations
Minor Impact	1-2	Minimal impact with no significant consequences; minor disruptions or inconveniences that do not affect core information and/or operations
Low Impact	3-4	Noticeable disruptions with limited consequences to information and/or operations
Medium Impact	5-6	Significant disruptions affecting critical information and/or operations
High Impact	7-8	Severe disruptions with major impacts on critical information and/or operations, potentially hindering recovery efforts
Very High Impact	9-10	Catastrophic disruptions causing long downtime, substantial losses, or permanent damage to information and/or operations

3.2 CISS Score Calculations

3.2.1 Severity Score (*S*)

According to [7], the Severity score evaluates the impact of an incident by combining three components, namely individual score, financial score, and operational score. This comprehensive metric provides insight into the scale and intensity of the incident. These are aggregated by considering an average, ensuring a balanced representation of the harm caused to individuals, financial systems, and organizational operations.

Individual Score: It evaluates the harm to affected individuals across five subcategories: physical safety, personally identifiable information (PII) breaches, nuisance/distress, financial harm, and operational harm. Each subcategory is scored on a scale from 0 (no impact) to 10 (extreme impact) and combined with population scores that account for the number of people affected. For example, physical safety measures range from minor injuries requiring first aid to fatalities, while PII breaches evaluate the sensitivity of compromised data, from basic contact information (low impact) to highly sensitive records like health data or legal information (high impact). Nuisance/distress measures emotional or psychological harm, while financial harm reflects direct monetary losses to individuals. Operational harm assesses disruptions to daily tasks, such as accessing essential services or managing personal responsibilities. These are also aggregated using the average of all the subcategories, reflecting both the severity of harm and the scale of those affected. This ensures a nuanced understanding of the incident's impact on individuals.

Financial Score: It quantifies monetary repercussions stemming from a security incident, using a scale from 0 (losses under $100K) to 10 (losses over $1B). These levels are informed by empirical breach reports, insurance data, and industry benchmarks, which ensures incremental levels of financial severity. Assessors factor in direct expenses such as investigation, legal actions, and response measures, as well as indirect costs tied to reputational damage or diminished customer trust. Because incidents can grow in scope as hidden liabilities appear, the Financial Score is recalculated whenever new data emerges. This dynamic nature allows stakeholders to grasp evolving fiscal burdens, guiding decisions on resource allocation and risk mitigation strategies. Ultimately, this score provides a consistent metric for comparing economic impacts.

Operational Score: It measures the extent to which an incident disrupts an organization's core functions on a scale of 0 to 10. It accounts for both the breadth (how many processes or systems are affected) and the duration (how long these disruptions persist). Minor issues such as short outages in non-critical applications receive lower scores, while incidents that halt production lines, obstruct customer services, or compromise safety protocols could reach the highest. Recovery progress influences the score, which may shift as incident response teams implement fixes or uncover new complications. By examining factors like system interdependencies, infrastructure resilience, and contingency procedures, the score provides a structured view of how deeply an incident impacts essential operations.

3.2.2 Criticality Score (*C*)

The criticality score, according to [7], evaluates the systemic importance of the affected systems by combining two dimensions, National Critical Function (NCF) and National Security Impact (NSI), into a single metric. Each dimension is scored on a 0–10 scale, with 0 indicating no impact and 10 representing a severe, prolonged effect on core national assets. The final

criticality score is an average of these two components.

The NCF evaluates the disruption to essential services, such as power grids, telecommunications, water supply, or financial transactions. The scoring scale ranges from short, localized interruptions, such as a brief internet outage scoring near 0–2, to multi-day failures across multiple interdependent services, such as energy, transportation, and emergency response, scoring 8–10. The framework often references established guidelines, such as the U.S. Cybersecurity and Infrastructure Security Agency (CISA)'s NCF pillars (Connect, Distribute, Manage, Supply) [25], but it is adaptable to country-specific classifications. The cascading effects of a single disruption can create ripple effects in other sectors, particularly if critical infrastructure is shared, thereby amplifying the overall severity.

The NSI assesses the sensitivity of any exposed or compromised data from unclassified but sensitive information to highly classified documents that, if leaked, could threaten national security. The scoring scale assigns higher values as data classification levels escalate. For example, a score of 0 signifies that no classified data was involved, a score of 4 might reflect unauthorized access to need-to-know information, and a score of 8–10 indicates a breach of Secret or Top Secret materials. The global adaptability of this scale allows organizations to align scoring with their own national standards, as classification systems vary by country. For instance, organizations may categorize data as "Confidential," "Secret," "Top Secret," or use equivalent local labels.

4 Implementation

To illustrate the applicability of the CISS model, we analyzed a real incident from a leading engineering and contracting firm in the UAE. During the 24-month construction phase of a $5 million commercial building project, phishing attempts occurred approximately every two months. One incident highlighted the effectiveness of the company's proactive cybersecurity measures. The attack was detected by the company's spam filtering system and promptly reported to the IT team, allowing for swift containment with only minor impacts. Thanks to robust defenses, including employee awareness training and immediate response protocols, the company ensured no data was compromised, and there were no significant operational, financial, or reputational damages. This incident underscores the critical role of structured cybersecurity practices in managing threats, maintaining operational continuity, and safeguarding stakeholder trust with minimal disruption.

4.1 Severity Score (S)

Individual Score. The phishing incident's individual impact was assessed across multiple dimensions to determine its severity. For physical safety, no harm occurred and thus no one was affected. In the Personally Identifiable Information (PII) category, PII with moderate sensitivity (e.g., contact details, employment information) was breached, affecting 10 employees. For nuisance/distress, about 100 employees (mostly IT staff, managers, and those who reported phishing attempts) were affected. Personal communication data (e.g., personal emails and messages) was stolen, causing inconveniences and disrupting daily routines. No financial losses were reported, so no one was financially impacted. For operational harm, about 150 employees (mainly IT and operational staff) experienced disruptions while managing phishing emails, delaying tasks like accessing online communication services. Table 2 shows a detailed breakdown of these dimensions. The integrated individual score, the average of all dimensions, is 1.5.

Table 2. Breakdown of Individual Score

Dimension	Impact Score	# of People Impacted	People Score	Ave. Score	Indv. Score
Physical Safety	0	0	0	0	1.5
PII Impact	2	10	2	2	
Nuisance/ Distress	3	100	2	2.5	
Financial Harm	0	0	0	0	
Operational Harm	4	150	2	3	

Financial Score. Although the phishing attacks led to some minor incidental costs (e.g., investigation and staff time), the overall financial impact remained below the $100k threshold set by [7]. As such, the Financial Score for this incident was 0, largely due to the company's proactive defenses (e.g., spam filtering) that effectively mitigated the threat.

Operational Score. In contrast, the Operational Score is 5, indicating significant disruptions to the firm's operations. The phishing incident required operational adjustments, such as workarounds and managing delays, underscoring its major impact on day-to-day processes even if it did not result in substantial financial losses.

The severity score can be computed by taking the average (as suggested in [7]) of the individually estimated scores as follows. However, it is worth noting that several other methods can be used, more of which are detailed in the discussion section.

$$Severity\ Score = \left(\frac{Individual\ Score + Financial\ Score + Operational\ Score}{3}\right) = \left(\frac{1.5 + 0 + 5}{3}\right) = 2.17$$

4.2 Criticality Score (C)

The incident occurred in a normal commercial

building project, which had no direct impact on critical infrastructure or cascading effects on national operations. Therefore, the NCF score is 0. Similarly, no sensitive or classified information was exposed, so the NSI score is 0, so the incident posed no risk to national security.

$$Criticality\ Score = \frac{NCF\ Score + NSI\ Score}{2} = \frac{0+0}{2} = 0$$

4.3 Overall CISS Score

The overall CISS score is derived from the Severity Score and Criticality Score with a weighted combination of 0.9 and 0.1, respectively, as suggested by [7]. Although several other weights can be considered, extensive set of correlative experiments conducted showed the significance of these weights and hence the usage for this implementation.

$$CISS\ Score = 0.9 * Severity\ Score + 0.1 * Criticality\ Score$$
$$= (0.9 * 2.17) + (0.1 * 0.0) = 1.95$$

The final CISS score is 1.95, indicating the incident had a small impact without significant consequences. It only caused minor disruptions or inconveniences without affecting core information or operations.

5 Discussion

5.1 Model Efficacy

The CISS score of 1.95 indicates that the phishing attack on the company only caused a minor impact. This score closely aligns with the reality of the case study, reflecting the effectiveness of proactive mitigation measures such as spam filtering systems and employee awareness. These measures successfully prevented any significant financial, operational, or security consequences. The score's consistency with the facts of the case demonstrates the efficacy of the proposed model in accurately assessing the impact of the phishing incident. By quantifying the relatively low severity and negligible criticality of the incident, the model effectively underscores its ability to provide a realistic and data-driven evaluation of cybersecurity risks. This correlation between the calculated score and the actual case outcomes further validates the model's robustness and utility in real-world scenarios.

5.2 Suggestions for Model Adaptation

Scoring Granularity. The severity score should be derived through systematic analysis of evidence across multiple hierarchical layers, each corresponding to specific causative factors or events, inspired by frameworks like the IHS Markit RRS [20] and the NIH Stroke Scale [21]. For example, the scoring can incorporate metrics such as individual repercussions and assessing the impact on personnel. It can be quantified by evaluating both the number of affected individuals and the severity of harm incurred. This multi-dimensional approach ensures a highly granular assessment, enabling the identification and consideration of nuanced factors that might otherwise be overlooked.

Adaptive Evolution. The model must prioritize adaptability to address the evolving cyber risks in construction. Inspired by frameworks like the SAE Levels of Driving Automation [22], which are updated to reflect technological advancements, the scoring model should also evolve in terms of its structure, dimensions, and inputs. This ensures responsiveness to emerging technologies, such as AI-powered design tools, and to evolving threats targeting interconnected systems. By maintaining this flexibility, the model can integrate new threat vectors and technologies, providing an up-to-date framework for assessing cybersecurity incidents in construction projects. This adaptability enhances resilience and promotes industry-wide collaboration.

Threshold Customization. Our study used predefined thresholds as illustrative baselines to reflect typical financial, operational, and stakeholder impacts. While these cutoffs effectively demonstrate how the model can be applied in practice, they should not be viewed as universally applicable. Different regions, project scales, and regulatory environments may require notably different threshold values to capture the true impact of cybersecurity incidents. We therefore recommend that future applications of this model begin with a calibration phase, during which these thresholds are adjusted based on local conditions, expert judgments, and available incident data. This can ensure that the model retains its relevance and accuracy as it is extended to broader contexts or more diverse construction project scenarios.

5.3 Insights and Implications

The CISS provides a standardized framework for reporting and assessing cybersecurity incidents. Using structured inputs, it ensures consistent data collection, produces a unified severity score, and facilitates reliable documentation across different organizations. The resulting quantitative metric helps prioritize and triage incidents, enabling meaningful comparisons of severity over time and across regions. By analyzing these scores, stakeholders can learn from the past, promote best practices, and respond to evolving cyber threats. CISS can also align cybersecurity efforts across public and private sectors, helping policymakers develop targeted regulations and allocate resources effectively. Through its common framework for assessment and reporting, CISS fosters transparency and accountability among clients, partners, and regulators. This standardized approach enhances industry resilience, promotes global

coordination, and protects critical infrastructure, making it valuable for the construction industry and beyond.

Many modern construction contracts, such as those using standard forms like FIDIC or NEC, now include clauses requiring specific cybersecurity standards and rigorous protocols for data protection. They may mandate compliance with recognized frameworks or set requirements for incident reporting. By integrating the CISS model into these contractual clauses, contractors can standardize severity assessments, ensuring compliance and fostering transparency. This alignment clarifies responsibilities and promotes uniformity across projects, facilitating more robust monitoring, response, and documentation of cyber incidents. Furthermore, embedding the CISS score into contractual deliverables encourages proactive adoption of best practices, reinforcing project security and stakeholder trust. Future work can investigate additional ways to link CISS metrics with emerging contractual obligations.

5.4 Limitations and Future Work

While the CISS model's applicability to construction has been demonstrated, several limitations exist, and further improvements are needed.

(1) The model's broad applicability can sometimes overlook industry-specific nuances due to the lack of context-specific factors in the model. In the construction sector, critical factors such as project delays, supply chain disruptions, and stakeholder confidence play a significant role but are not directly accounted for within the model's current dimensions. Addressing these industry-specific aspects would make the model more effective in reflecting the unique risks and impacts of construction-related cybersecurity incidents, enhancing its accuracy and relevance in this context.

(2) Another limitation of the CISS implementation is its reliance on weighted averaging, which can oversimplify the cybersecurity incidents. By converting impact dimensions into a single score through weighted averaging, granular interactions and cascading effects may be overlooked, hiding critical insights. However, this approach serves as a starting point to demonstrate how the CISS model can be applied in construction. In the future, more sophisticated weighting schemes or multi-criteria decision-making approaches could be integrated into the model to capture the complexity of cyber incidents more accurately in construction.

(3) Another technical limitation is the model's inability for real-time updates, a key aspect given fast-paced IT developments and evolving cyber threats. Although the current CISS model relies on static or post-incident data, construction environments shift quickly. By enabling real-time data inputs like IoT sensor feeds and continuous project updates, the model could provide near-instantaneous threat detection and adaptive scoring, aligning more closely with proactive cybersecurity strategies. As IT technology advances and threats grow more sophisticated, real-time adaptability will be crucial to safeguarding construction projects by enhancing situational awareness and rapid response capabilities.

(4) Using a single case study may limit the model's generalization, which may overlook the diversity of potential incident types, project scales, and organizational contexts. Consequently, certain scoring factors and contextual details within the model may not fully capture all conditions, which necessitates more extensive data and multiple examples to validate and refine the scoring model. Further research with a range of real-world incidents would help ensure that the scoring methodology is robust and reflective of the complex realities faced by different construction projects.

6 Conclusion

This study addresses the research gap regarding the lack of cybersecurity incident severity assessments and the absence of standardized measurement and reporting within the construction industry. By applying the CISS model, this study provides an integrated score for evaluating the severity of cyber incidents in construction projects. The model assesses severity based on financial, operational, and individual impacts, while criticality considers NCFs and the potential exposure of classified information. These dimensions collectively provide a comprehensive evaluation of a cybersecurity incident's impact and significance. Through a case study involving a real-world phishing incident at a construction company, the model demonstrates its applicability in assessing cybersecurity incidents within the construction context. The implications of this research include promoting the adoption of an existing structured framework for consistent and standardized incident reporting, enabling uniform data collection, supporting meaningful incident comparisons over time and across diverse contexts, fostering collaboration across organizations and regions, and ultimately contributing to industry-wide efforts to enhance cybersecurity resilience and coordination.

Acknowledgments

This research was partially supported by the Center for Sand Hazards and Opportunities for Resilience, Energy, and Sustainability (SHORES), funded by Tamkeen under the NYUAD Research Institute Award CG013; the Center for Cyber Security at New York University Abu Dhabi (CCS-AD), funded by Tamkeen under the NYUAD Research Institute Award G1104; and the Center for Interacting Urban Networks (CITIES), funded by Tamkeen under the NYUAD Research Institute Award CG001. We also thank ALEC Engineering & Contracting LLC, particularly Mr. Sabyasachi Jana, for providing project-related information.

References

[1] Deloitte, "Building Cybersecurity in the Construction Industry." Accessed: Sep. 30, 2023. [Online]. Available: https://www2.deloitte.com/ce/en/pages/real-estate/articles/ce-building-cybersecurity-in-the-construction-industry.html

[2] "DHS-CISA & DG CONNECT Cyber Incident Reporting Comparison Report," DHS, DG CONNECT, Mar. 2024. Accessed: Nov. 20, 2024. [Online]. Available: https://www.dhs.gov/publication/comparative-assessment-dhs-harmonization-cyber-incident-reporting-federal-government

[3] Baker McKenzie, "Global Data Privacy and Cybersecurity Handbook," Baker McKenzie. Accessed: Nov. 22, 2024. [Online]. Available: https://resourcehub.bakermckenzie.com/en/resources/global-data-privacy-and-cybersecurity-handbook/asia-pacific/singapore/topics/breach-notification-requirements

[4] China Briefing, "China's Cybersecurity Regulator Issues Draft Measures on Incident Reporting," China Briefing. Accessed: Nov. 15, 2024. [Online]. Available: https://www.china-briefing.com/news/chinas-cybersecurity-regulator-issues-draft-measures-on-incident-reporting/

[5] Australian Cyber Security Centre (ACSC), "Report a Cyber Security Incident," Australian Cyber Security Centre (ACSC). Accessed: Nov. 29, 2024. [Online]. Available: https://www.cyber.gov.au/report-and-recover/report/report-a-cyber-security-incident

[6] "Cyber Incident Reporting for Critical Infrastructure Act (CIRCIA) Reporting Requirements; Extension of Comment Period," Federal Register. Accessed: Oct. 23, 2024. [Online]. Available: https://www.federalregister.gov/documents/2024/05/06/2024-09505/cyber-incident-reporting-for-critical-infrastructure-act-circia-reporting-requirements-extension-of

[7] C. F. Conard, "Quantifying the Severity of a Cybersecurity Incident for Incident Reporting," Graduate Theses, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States, 2024. Accessed: Dec. 17, 2024. [Online]. Available: https://hdl.handle.net/1721.1/157124

[8] Willis Towers Watson, "Cyber Quantified," Willis Towers Watson. Accessed: Mar. 03, 2025. [Online]. Available: https://www.wtwco.com/en-us/solutions/products/cyber-quantified

[9] Marsh McLennan, "Marsh McLennan Cyber Risk Intelligence Center," Marsh McLennan. Accessed: Mar. 03, 2025. [Online]. Available: https://www.marshmclennan.com/solutions/cyber-resilience/cyber-risk-intelligence-center.html

[10] Verizon, "2024 Data Breach Investigations Report," Verizon, 2024. Accessed: Feb. 27, 2025. [Online]. Available: https://www.verizon.com/business/resources/reports/dbir

[11] CrowdStrike, "CrowdStrike 2024 Global Threat Report," CrowdStrike, 2024. Accessed: Mar. 01, 2025. [Online]. Available: https://www.crowdstrike.com/global-threat-report

[12] European Union Agency for Cybersecurity (ENISA), "ENISA Threat Landscape 2024," ENISA, Sep. 2024. Accessed: Mar. 01, 2025. [Online]. Available: https://www.enisa.europa.eu/publications/enisa-threat-landscape-2024

[13] International Atomic Energy Agency, "Computer Security for Nuclear Security," IAEA, 2021. [Online]. Available: https://www.iaea.org/publications/13629/computer-security-for-nuclear-security

[14] N. Salami Pargoo and M. Ilbeigi, "A Scoping Review for Cybersecurity in the Construction Industry," *Journal of Management in Engineering*, vol. 39, no. 2, p. 03122003, Mar. 2023, doi: 10.1061/JMENEA.MEENG-5034.

[15] B. R. K. Mantha and B. García de Soto, "Cyber Security Challenges and Vulnerability Assessment in the Construction Industry," in *Proceedings of the Creative Construction Conference 2019*, Budapest University of Technology and Economics, 2019, pp. 29–37. doi: 10.3311/CCC2019-005.

[16] B. R. K. Mantha and B. García de Soto, "Assessment of The Cybersecurity Vulnerability of Construction Networks," *Engineering, Construction and Architectural Management*, vol. 28, no. 10, pp. 3078–3105, Nov. 2021, doi: 10.1108/ECAM-06-2020-0400.

[17] B. R. K. Mantha, M. S. Sonkor, and B. García de Soto, "Investigation of the cyber vulnerabilities of construction networks using an agent-based model," *Developments in the Built Environment*, vol. 18, p. 100452, Apr. 2024, doi: 10.1016/j.dibe.2024.100452.

[18] D. Yao and B. García De Soto, "Cyber Risk Assessment Framework for the Construction Industry Using Machine Learning Techniques," *Buildings*, vol. 14, no. 6, p. 1561, May 2024, doi: 10.3390/buildings14061561.

[19] A. B. Emel, M. Oral, A. Reisman, and R. Yolalan, "A credit scoring approach for the commercial banking sector," *Socio-Economic Planning Sciences*, vol. 37, no. 2, pp. 103–123, Jun. 2003, doi: 10.1016/S0038-0121(02)00044-7.

[20] IHS Markit, "Banking Risk Ratings Methodology," S&P Global. Accessed: Dec. 08, 2024. [Online]. Available: https://www.spglobal.com/_assets/documents/marketplace/ihs-markit-banking-risk-ratings-methodology.pdf

[21] Naureen, "NIH Stroke Scale - NurseBrain®," NurseBrain®. Accessed: Dec. 12, 2024. [Online]. Available: https://nursebrain.com/2021/12/nih-stroke-scale/

[22] R. Harrington, C. Senatore, J. Scanlon, and R. Yee, "The Role of Infrastructure in an Automated Vehicle Future," *Bridge*, vol. 40, pp. 48–55, Jun. 2018.

[23] S. Radack, "The Common Vulnerability Scoring System (CVSS)," National Institute of Standards and Technology.

[24] Cybersecurity and Infrastructure Security Agency (CISA), "CISA National Cyber Incident Scoring System (NCISS)," CISA, 2023. Accessed: Mar. 02, 2025. [Online]. Available: https://www.cisa.gov/news-events/news/cisa-national-cyber-incident-scoring-system-nciss

[25] Cybersecurity and Infrastructure Security Agency (CISA), "National Critical Functions (NCF) Framework," CISA, 2024. Accessed: Feb. 25, 2025. [Online]. Available: https://www.cisa.gov/topics/risk-management/national-critical-functions

Development of a Construction-Information System by Measuring, Processing, and Utilizing 3D Data Toward Creating a Digital Twin

Satoshi Kubota[1] and Aika Yamaguchi[1]

[1]Faculty of Environmental and Urban Engineering, Kansai University Graduate School, Japan
skubota@kansai-u.ac.jp

Abstract

Because the scale and conditions of construction sites vary widely and it can be difficult to use a one-size-fits-all method to acquire data, a method for transferring and utilizing three-dimensional (3D) data from one stage to another in construction projects is urgently needed. In this study, we developed a construction-information system based on 3D point cloud data, taking into consideration the characteristics and conditions of construction sites and the definition of information systems in order to efficiently distribute and utilize 3D data. The use cases of the system were applied to the digital twin of a construction site. 3D point cloud data constructed using UAVs and camera-based photogrammetry were used to visualize the progress at the construction site and to understand the transition of earthworks. The results suggested the necessity of changing the measurement and processing methods according to the conditions of the construction site and the intended use of the data, as well as devising ways to reduce the impact on the current workflow.

Keywords –

digital twin, three-dimensional point cloud data, information system, construction practice

1 Introduction

The Ministry of Land, Infrastructure, Transport and Tourism (MLIT) of Japan is promoting "i-Construction," in which information and communication technology (ICT) is fully introduced in construction projects. However, construction sites are diverse and it can be difficult to use a one-size-fits-all method for acquiring three-dimensional (3D) data and 3D point cloud data to understand the site conditions. In urban spaces, digital twins are being promoted to collect data related to civil infrastructure and urban activities and visualize them in a three-dimensional space. An evolution has been transition from using each digital tools to the more integrated framework of digital twins [1][2]. However, in the construction field, there have been few initiatives to collect and manage information on construction sites with the aim of creating digital twins that reproduce information from construction sites in a virtual space, and it remains unclear which types of data can be acquired on construction sites. It is expected that the collection and visualization of data will advance i-Construction; eliminate interference with electric power, gas, and sewage systems in underground buried structures; and assist the engineer's decision-making process in inspection and diagnosis. However, no methods for transferring data between the planning, surveying, design, construction, inspection, maintenance, and renewal stages of civil infrastructure have been established.

In this study, a construction-information system is proposed in which 3D data and 3D point cloud data are distributed and utilized at small and medium-sized construction sites. The construction information that can be obtained at construction sites is considered and defined, a construction-site digital twin based on 3D point cloud data is considered, and use cases for utilizing construction information are discussed. Furthermore, the issues related to the acquisition, processing, management, and utilization of construction information at each stage of the use case are clarified.

2 Issues of Construction Information

2.1 Current Status and Issues Related to Construction Information

MLIT has published a public survey manual and management guidelines, which include surveying standards for 3D data at construction sites, but detailed information on shooting conditions, survey methods, and survey equipment for different site characteristics are not provided, making it difficult to measure and maintain data in a uniform manner at all sites.

Small and medium-sized construction companies

face the problem that the use of 3D data and i-Construction equipment has not progressed. Suzuki et al. proposed a measurement method with the aim of addressing coverage and cost issues [3]. However, a simple 3D data–surveying method and a 3D data–construction method involving inexpensive unmanned aerial vehicles (UAVs) and cameras are necessary to enable small and medium-sized construction companies to introduce ICT and carry out construction projects efficiently. Tanaka et al. [4] and Takahashi et al. [5] proposed and validated a measurement method involving a UAV-mounted camera or laser scanner. However, when determining the target structure and performing measurements, it is necessary to follow a survey method that is appropriate to the scale of the site and the surrounding conditions.

Kanzaki demonstrated how to utilize 3D models in the planning, construction, and maintenance stages and discussed the effects of their introduction but did not describe the data transfer from planning to construction and from construction to maintenance, nor the effects of this [6]. Fujita et al. proposed a method for assigning attributes to point cloud data but did not consider the data volume and format of point cloud data in consideration of operation [7]. Therefore, it is necessary to demonstrate how data can be transferred from the construction stage to the maintenance management stage as well as how data can be utilized between stages.

Building information modeling was used from the design process to the construction process as research for displaying various information in 3D space and using it for decision-making. Methods for utilizing 4D models by linking 3D models with scheduling information and for developing a 4D integrated construction-site planning system have been proposed [8][9]. A digital twin used in a bridge inspection has also been discussed [10].

2.2 Problem solving in this research

To address the issue of surveying according to the scale and conditions of the site, this study considers methods that are appropriate to the characteristics of the site, where construction equipment and people are present and data-feature points are few, when measuring the construction site with a terrestrial laser scanner (TLS) or camera. To construct 3D data using inexpensive equipment, we propose the acquisition of 3D point cloud data using cameras shooting from high locations in addition to UAVs, so that the data can be used even at construction sites where there are limitations on the equipment used. A construction-information system is conceived to solve the problem of passing data from the construction stage to maintenance management stage. The points to be considered in the measurement, processing, management, and utilization of data at each stage of pre-construction, construction, post-construction and maintenance are clarified.

3 Concept of a Construction-Information System Using 3D Data

3.1 Construction-Site Digital Twin

In this study, a digital twin is the visualization of a construction site in a digital space that is created by processing data collected at the construction site. The digital twin collects and accumulates daily progress and construction records as data from the start to the completion of construction. It is also assumed that a 3D construction site based on 3D point cloud data will be constructed, and the daily accumulated data will be displayed and used for construction planning and other purposes. Here, the digital twin has three roles: to provide “past records” as a data storage base that can be referred to when data are needed in later processes, by accumulating work records and changes in site data up to the present; “understanding the present” by reproducing the three-dimensional construction space digitally and reflecting the situation in the real world; and “predicting construction progress from the data accumulated thus far.”

In order to use ICT technology to advance construction management and realize a digital twin, it is important to save the data generated at the construction site as a log from the day-to-day construction. However, there is no research on the construction of a digital twin by accumulating data acquired at the construction site from the planning stage to the completion of construction. The 3D data related to the social infrastructure of the digital twin has issues such as surveying methods and 3D data construction according to the site conditions, and verification of accuracy due to differences in objects and surveying methods. Images related to urban activities, the action history of workers and logs of construction equipment are important information that should be acquired and retained at construction sites for use in safety management and health management. However, it is difficult to introduce these systems to all construction sites due to the need to store large amounts of data, the lack of data standardization to accumulate and utilize the data generated on a daily basis, and the high costs involved.

There are no current government initiatives to realize a digital twin of construction sites, where the data acquired from the planning stage to the completion of construction are accumulated. The reasons for this are as follows: the type of data to be acquired and stored at construction sites has not be specified; the work of artificially acquiring data without automatically obtaining logs from i-Construction equipment and wearable devices places a burden on workers; and the use cases for storing data are not clear. The use case for

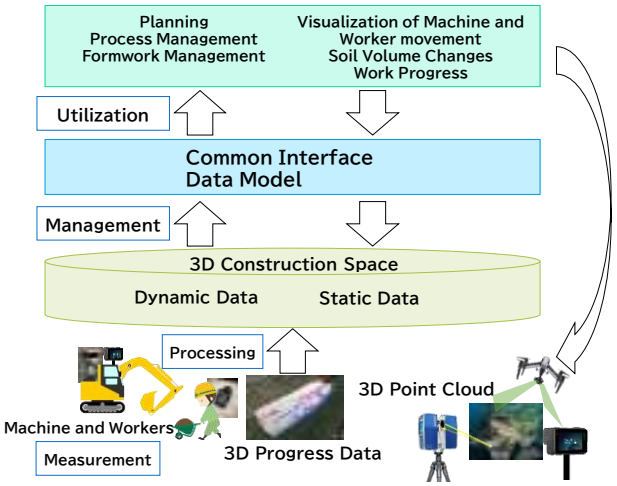


Figure 1. Concept of construction information system

storing data is not clear.

The construction site digital twin collects and accumulates data in accordance with data definitions based on the idea of digital threads that accumulate data transitions. The digital thread is a concept that accumulates data attributes and their transitions, analyses useful data from the accumulated data, and uses the results to realize construction sophistication through data utilization. An image of the construction site digital twin is shown in Figure 1. In the construction site digital twin, daily progress and construction records are collected and stored as data during the entire construction period of a civil engineering structure, from the start of construction to construction and completion. The accumulated data is used to construct a three-dimensional construction site based on 3D point cloud data, and as feedback to the site, the daily accumulated camera images and action logs are displayed and used for construction planning.

3.2 Use Cases

The use cases of the construction-site digital twin are process management, quality management, and safety management. In process management, 3D point cloud data of the site topography are acquired on a daily basis by measuring equipment such as TLS and UAVs and then superimposed on past data, enabling each day's progress to be monitored. Because daily records are accumulated, they can be checked regularly and corrected at the construction stage before construction defects are discovered after completion. In quality control, when a construction defect or deficiency occurs, the data collected on the digital twin can be used to reconstruct the situation at the time and identify the cause of the construction error. Furthermore, information from

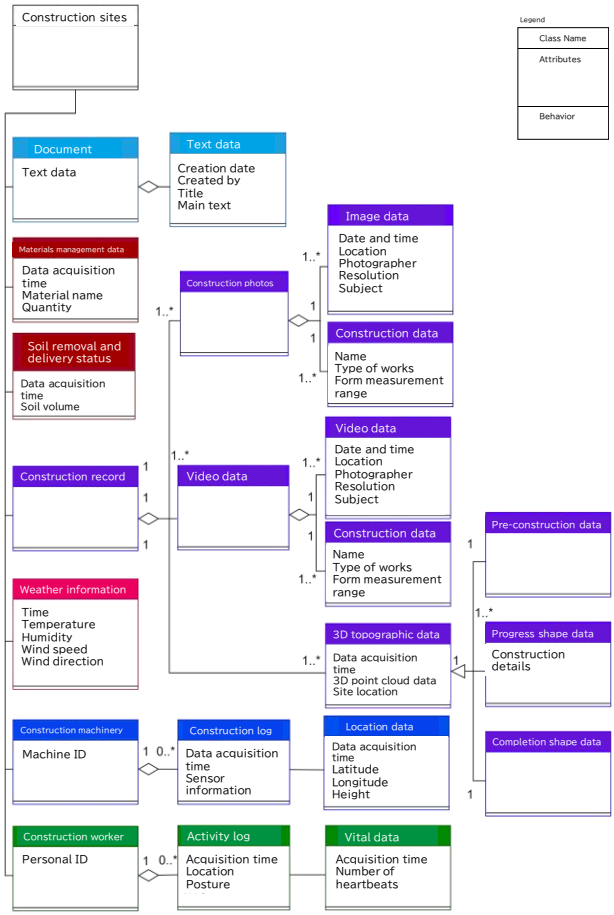


Figure 2. Class diagram of data acquiring at the construction site

the construction site is aggregated in the digital twin, enabling a remote presence in the digital space. In safety management, the physical condition of workers is managed by quantitatively evaluating changes in their physical condition based on vital data acquired using smart devices. For this purpose, movement history and changes in heart rate are recorded using Global Navigation Satellite System (GNSS) information. The location information of construction machinery and workers is used to analyze situations where contact accidents are likely to occur.

3.3 Proposal for a Construction-Information System

In this section, a digital twin-oriented construction-information system (Figure 1) is proposed to manage data generated at construction sites. It was utilized in the use cases described above. At the construction site, the shape and changes of the structure are acquired by TLS and UAVs, and by cameras and wearable terminals to reduce the workload of artificially acquiring the behavior of

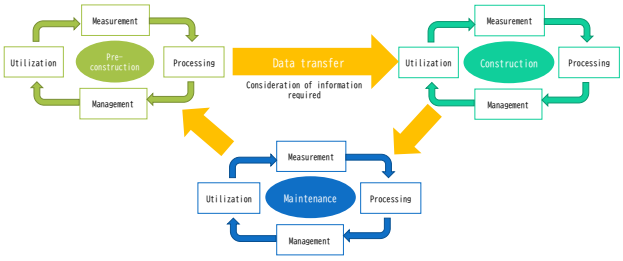


Figure 3. The information-systems cycle in construction projects

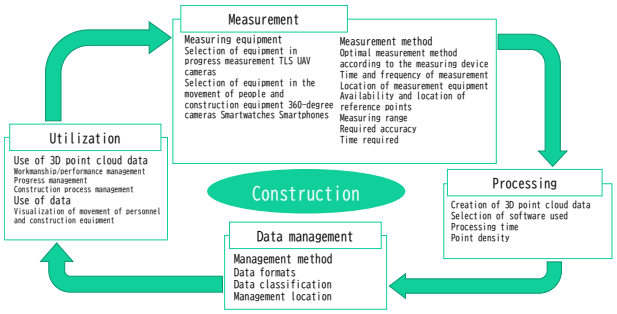


Figure 4. Collection, processing, communication, and use of information during and after construction

construction equipment and workers. The data to be acquired and accumulated at the construction site are classified with reference to the data hierarchy in the local dynamic map [9] in automatic operation, based on the characteristics of the construction site where the situation is constantly changing. In other words, the data are classified as “dynamic information” or “static information” according to the speed of change in the acquired data, while the acquisition frequency is defined so that the data acquired at the construction site are always the latest. The data to be managed were examined based on the MLIT guidelines, as shown in the class diagram in Figure 2. Figure 2 was described using UML (Unified Modeling Language). Construction photographs and 3D point cloud data are classified as “construction records.” The 3D point cloud data for the ground survey, which are acquired before construction, and the 3D point cloud data used for the formwork inspection at the time of completion are each acquired once. In contrast, the topographic data acquired to record the daily progress during construction are managed on a daily basis together with the details of the construction work, thereby serving as a construction record. By classifying these data, it becomes easier for the user to distinguish the data by purpose. In addition, the metadata that summarize the measurement and processing methods of the data are also managed together, providing a reference when proceeding to the next step in the process.

As shown in Figure 3, the construction-information system implements the process of collecting information by measuring the construction site or structure, processing the collected data, communicating by managing the data and utilizing the data at each stage of pre-construction, construction, and post-construction as well as maintenance management. This cycle is repeated for each stage of construction and maintenance management, and information is accumulated accordingly. When the pre-construction, construction, post-construction, and maintenance management processes are completed and the next stage is reached, the necessary information is reviewed and the data are passed on. The data acquired in the maintenance management stage are used as part of the completion drawings and construction plans before repair work is carried out.

Figure 4 shows the items to be considered at each stage of measurement, processing, management, and utilization of information in the construction-information system during and after construction. In the “collection of information” stage, the measurement range and equipment locations are set according to the measurement target, the existence and location of reference points, and the timing and number of measurements in order to meet the required accuracy, specified in the workmanship management guidelines, workmanship management supervision and inspection guidelines, and the work piece calculation guidelines. In the case of 3D point cloud data, the TLS point cloud data measured at multiple locations are combined and 3D point cloud data are constructed based on videos and photographs. Software is selected that is appropriate for the equipment used, and the density of the 3D point cloud data is taken into account. In the “information transfer” stage, data are passed from the construction stage to maintenance management stage in accordance with the data specification (Figure 2). In the “use of information” section, the constructed 3D point cloud data and visualized soil volume data are used for daily progress management, construction-process management, and formwork management. It is also possible to visualize changes in soil volume from two 3D point cloud data measured at different times and to visualize the movements of people and construction equipment from 3D point cloud data and GNSS information. In the maintenance management stage, the 3D point cloud data of structures captured when construction was completed can be used to visualize damage as well as for periodic inspections.

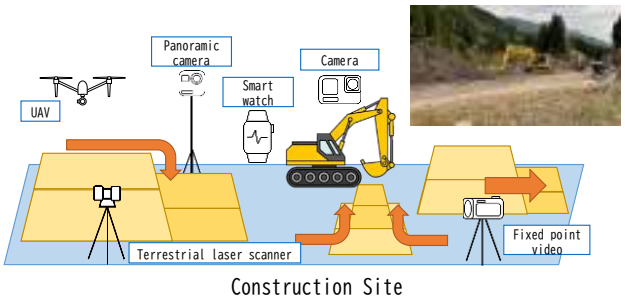


Figure 5. Experiment site



(a) Panoramic view of heavy equipment with camera attached



(b)An image acquired from the heavy equipment camera

Figure 6. Experimental

4 Implementation of Construction Information Systems

4.1 Experimental Overview

In this section, the cycle of measurement, processing, management, and utilization of the construction-information system is verified at the construction site, and the points to be considered in each process toward the construction of a digital twin are clarified. Figure 5 and 6 show the experiment. Assuming a small to medium-sized construction site, measurements were taken before and after construction with different embankment shapes. An Inspire2 UAV (DJI) was used when human access was restricted, and an action camera mounted on a monopod was used for photogrammetry from a height in areas where UAV flight was restricted or where the area was too narrow. From these, 3D point cloud data were generated using structure from motion/multi-view stereo processing. In aerial photogrammetry, the measurement time, 3D point cloud data construction time, and TLS 3D point cloud data are compared with the accuracy of the data as the most accurate values by changing the measurement equipment and altitude.

4.2 Experiment

4.2.1 Measurement of Construction Sites

To verify the accuracy of photogrammetry and visualize progress using the UAV-mounted camera and a camera mounted on a monopod, four black-and-white markers were used as feature points on the photographs: two wooden stakes were used as marking points and two wooden stakes as verification points. The positions of the fixed points and verification points were not changed before and after the construction, and the same values were used: the UAV flew one round-trip at 30 m and two round-trips at 20 m, depending on the angle of view of the camera. Photogrammetry by UAV and camera was carried out twice, once before and once after construction, in order to visualize the construction progress, using 3D point cloud data.

4.2.2 Data Processing

The data measured by the camera were given the absolute coordinate values of the six fixed points, so that the actual dimensions of the site could be measured. The accuracy and quality of the data during processing were considered to be the highest. In the public survey manual, the standard is 80% overlap and 60% sidelap, but in this study, the number of photographs was reduced by changing the shooting altitude, considering that the measurement and processing time for constructing the 3D point cloud data from camera images is proportional to the number of photographs used, given that the construction progress is the subject of this study. In the constructed 3D point cloud data, the UAV photogrammetry was carried out in the vertical direction, so depending on the flight route and altitude, the vegetation became a barrier and data on the ground surface were lost. When constructing the 3D point cloud

Table 1 Accuracy assessment by comparison of photogrammetric results with TLS point cloud data.

	Percentage of points within the required accuracy of ±200 mm (%)			Measurement Time	Processing Time	Data (GB)
	Land 1	Land 2	Land 3			
UAV (Height 20 m)	92.9	95.9	99.9	3 m19 s	24 m	2.45
UAV (Height 30 m)	93.9	96.2	99.9	4 m58 s	8 m	0.89
Camera (Height 3 m)	99.8	96.8	99.9	5 m58 s	8 h30 m	10.81
Camera (Height 8 m)	98.5	95.7	99.8	3 m43 s	1 h50 m	2.86

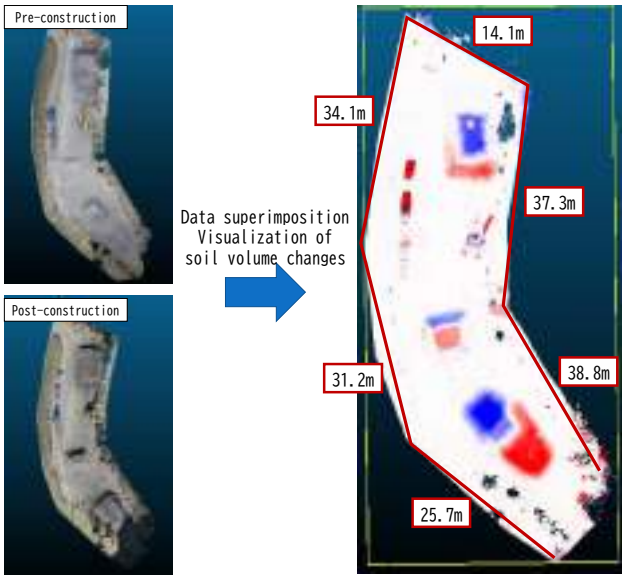


Figure 7. Visualization of construction progress with color display

data, any people, vehicles, and building materials reflected in the data were manually clipped out, thereby reducing noise.

4.2.3 Data Management

The processed data were classified according to the class diagram: the 3D point cloud data were classified as 3D topographic data in the construction record. The 3D point cloud data before construction were classified as pre-construction topographic data and the post-construction data as progress topographic data, and the acquisition time, measurement location, and construction implementation details of each data point were recorded.

4.2.4 Utilization of Data

To quantitatively monitor the daily progress at the construction site, the accuracy for this study was set according to the MLIT's Draft Guideline for the Calculation of Earthworks Completion Using Stereo Photogrammetry (Ground Moving Object) and Draft Guideline for the Calculation of Earthworks Completion Using Ground Photogrammetry (Moving Image Type), using 3D point cloud data captured by the TLS as the most accurate value. The horizontal and elevation range were within ±200 mm.

As a verification method, the difference between the 3D point cloud data captured by the TLS and the 3D point cloud data captured by aerial photogrammetry was calculated using the 3D point cloud data-processing software CloudCompare. The accuracy evaluation results are shown in Table 1. UAV aerial photogrammetry and high-altitude camera photogrammetry surveys conducted in this study, more than 92% of the 3D point cloud data points satisfied the required accuracy of ±200 mm. This indicates that the overlap and sidelap in the photogrammetric surveys were adequate. Noise and deficiencies due to vegetation and photogrammetric errors resulted in some points failing to meet the required accuracy, so it is necessary to cut out unnecessary data and change the number and location of the fixed points.

The distance between the points of the 3D point cloud data before and after construction is calculated as the construction progress at the site, and the change in soil volume is visualized by a color display. As shown in Figure 7, red indicates a positive change in height after construction compared with before construction, blue indicates a negative change, and white indicates no change. The color becomes darker if the change in soil volume is large, and lighter if it is small. If 3D measurements are taken at each stage of the construction progress, they can also be used for construction-progress management.

The third use case is to understand the construction status by tracing back the time period from the data recorded throughout the construction. The camera images were positioned and oriented on top of the 3D point cloud data, which are topographic data. In addition to the data acquired before and after the construction, the topographic data visualizing the progress of the slide shown earlier are also displayed, and the acquisition time (shooting time) for each data point was set. As shown in Figure 8, by moving the time axis, the site conditions and the data acquired at that time can be displayed.



Figure 8. Construction-information system

5 Considerations

5.1 Considerations on the Construction Information System

For photogrammetry using the UAV-mounted camera, the flight altitudes were 20 m and 30 m. In this study, to compare 3D point cloud data at different flight altitudes, two round-trips were made over the site at an altitude of 20 m, where the angle of view is narrower, and one-round trip was made at an altitude of 30 m, with the same lap rate. At an altitude of 30 m, the flight time and the construction time of 3D point cloud data were shorter than at an altitude of 20 m. Increasing the altitude increases the area that can be photographed, reduces the number of images, and shortens the time for photographing and constructing the 3D point cloud data. Although the accuracy of the 3D point cloud data decreases as the altitude increases, the topographic data measured on a daily or weekly basis is not required to have high accuracy for the purpose of understanding the daily progress. In the photogrammetric measurement at an altitude of 3 m using a camera mounted on a monopod, the construction site was photographed in two rounds, with the camera facing forward in the direction of travel in the first round and facing sideways in the second round, in order to reduce the loss of 3D point cloud data. Depending on the site conditions, it is necessary to devise ways to extract feature points, such as setting the walking route and image cropping intervals when taking photographs. By mounting the camera on a monopod, it is possible to take photographs from high altitudes, even in areas where UAVs have flight restrictions or weather conditions are not met.

5.2 Considerations on the Feasibility of Use Cases

In this chapter, use cases of process control, quality control, and safety control are examined to clarify the feasibility and issues. In the process control, 3D point cloud data and workers' vital data were used to determine the quantity of work, work progress, and workload of embankment construction. The results suggest that the UAV can be utilized for workmanship control. Therefore, the front wrap rate of 60% and sidelap rate of 40% set for the measurement using the UAV-mounted camera and the walking route for the measurement using the elevated camera were appropriate. On the other hand, 7.1% of the points did not meet the required accuracy at an altitude of 20 m for the UAV and 4.3% at an altitude of 8 m for the camera. It is necessary to walk around the measurement target multiple times and change the camera angle to avoid blind spots caused by structures and construction equipment. For the soil volume, the coordinates of the cross section were evaluated using the coordinates of the cross section of the 3D point cloud data from the UAV flight altitude of 30 m in the embankment area and the TS coordinates.) In this study, errors were also found at the top and bottom corners of the embankment by TS surveying. This difference may have affected the construction progress, volume and volume.

The volume of soil was calculated because the volume of soil is used to calculate the progress of construction in earthwork. Since each embankment was deformed, the volume of soil did not change significantly, but the volume of soil before and after deformation differed due to the use of heavy machinery to shape and compact the soil. This difference in soil volume can be used to calculate the compaction ratio, which is one of the criteria for earthwork.

In this study, the accuracy is low because 3D point cloud data by SfM/MVS processing was used, but if the measurement accuracy is high, a more accurate evaluation is possible. To visualize the progress of construction, two sets of data, one before and the other after construction, were superimposed and displayed in color to grasp the transition of earthwork. The 3D point cloud data before and after construction were given absolute coordinates, which eliminated the need to superimpose the data. In this study, two sets of data were used for construction progress, but daily progress can be visualized with three or more sets of data. In the experiment, the distance between two selected data points was calculated, so that the worker can select two data points at worker discretion and grasp the daily progress of construction and the progress since the start of construction. In this study, the 3D point cloud data by TLS was used as the most accurate value, but it can be used for visualization of construction progress in the same way as photogrammetric data by camera.

In quality control, time synchronization of the acquired images was performed to reproduce the on-site conditions. This allows us to confirm the site conditions at that time from multiple viewpoints. Video data was replayed on 3D point cloud data using TerraExplorer. We set the shooting position, orientation, and height, and we believe that we were able to reproduce the construction site digitally. However, the data acquired at the construction site must be brought back to the office for processing, and it takes time for the data to be reflected. A system is needed to upload the data acquired at the construction site to the cloud and display detailed information on the site status in real time.

For safety management, 3D point cloud data created based on worker positions and heart rate data were superimposed on the underlying 3D point cloud data. The color of the points changes according to the heart rate value, enabling quantitative judgment of changes in the worker's physical condition. The superimposition of the position of the construction machinery on these data can be used to prevent accidents involving contact between the construction machinery and the worker. Vital data can be visualized as 3D point cloud data, and it is effective to change the display method depending on the application. However, there are issues such as setting threshold values to capture changes in workload and physical condition, and managing the time axis to display the worker's position in real time.

In Japan, local governments have released 3D point cloud data on the Web. However, these data are available only at the operational stage or after construction, and there are no examples of releasing 3D point cloud data during construction. If the data measured in this study during construction and 3D point cloud data, along with metadata, are made publicly available on the Web, it is expected to promote the use of 3D data and new applications at many construction sites. Therefore, it is necessary to consider how to manage and store the enormous amount of data acquired at construction sites.

6 Conclusion

In this study, a construction-information system was proposed with the aim of distributing and utilizing 3D data and 3D point cloud data at small and medium-sized construction sites where ICT-assisted construction may be challenging. The system enables the construction of a digital twin and specifies the data to be acquired. In verifying the use case, the measurement method and required accuracy of the UAV-mounted camera and an elevation camera to monitor the construction progress were clarified. In addition, soil volume changes were visualized by data superimposition and measuring differences in distance, using 3D point cloud data.

In the future, the usefulness of the construction-information system will be evaluated through practical application at several sites. Furthermore, to manage the enormous amount of data acquired daily at construction sites, a system for sharing weekly 3D point cloud data and images extracted from videos among construction workers as well as a method for visualizing the site situation by using mesh data generated from 3D point cloud data will be studied.

References

[1] Albalkhy, W., Kamaoui, D., Ducoulombier, L., Lafhaj, Z. and Linner, T. Digital twins in the built environment definition, applications, and challenges. *Automation in Construction*, 162, 2024.

[2] Soman, R. K., Farghaly, K., Mills, G. and Whyte J. Digital twin construction with a focus on human twin interfaces. *Automation in Construction*, 170, 2025.

[3] Suzuki, R. Maehara, H., and Taira, K. Mobile Camera Localization Utilizing Automatic Tracking Total Station. *Journal of the Japan Society of Photogrammetry, and remote sensing*, 58(6):322-327, 2019.

[4] Tanaka, S., Kubota, S., Imai, R., Nakamura, K., Yamamoto, Y., Tsukada, Y., Taniguchi, H., and Nakahara, M. Accuracy evaluation of measurement with laser scanner equipped UAV for earthworks construction management. *Journal of Japan Society of Civil Engineers, Ser. F3 (Civil Engineering informatics)*, 74(2):II_125-II_135, 2018.

[5] Takahashi, G., Tsunogai, R., Miyamoto, S., Matsuura, Y., Takiguchi, Y., Ogawa, M., Miyahira, E., and Hayashi, T. Consideration to topography survey of river bed using UAV, *Journal of Japan Society of Civil Engineers, Ser. G (Environmental Research)*, 75(5):I_313-I_322, 2019.

[6] Kanzaki, K. Utilization of the three-dimensional model to a construction production system, *Journal of Japan Society of Civil Engineers, Ser.F4 (Construction management)*, 73(4):I_55-I_63, 2017.

[7] Fujita, Y. Ogata, S. Wo, C. and Kobayashi, I.: Attribute assigned road point cloud for using in construction life cycle, *Journal of Japan Society of Civil Engineers, Ser.F3 (Civil Engineering informatics)*, 70(2):I_144-I_151, 2014.

[8] Wang, H. J., Zhang, J. P., Chau, K. W. and Anson, M.: 4D dynamic management for construction planning and resource utilization, *Automation in Construction*, 13(5):575-589, 2004.

[9] Ma, Z., Shen, Q. and Zhang, J. : Application of 4D for dynamic site layout and management of construction project, *Automation in Construction*, 14(3):369-381, 2005.

[10] Ye, C., Butler, L., Calka, B., Iangurazov, M., Lu, Q.,

Gregory, A., Girolami, M., Middleton, C.: A Digital Twin of Bridges for Structural Health Monitoring, *Proceedings of the 12th International Workshop on Structural Health Monitoring*, 2019.

[11] ETSI TR 102 863 V1.1.1：*Intelligent Transport Systems(ITS)；Vehicular Communications；Basic Set of Applications；Local Dynamic Map(LDM)；Rationale for and guidance on standardization*, 2011.

RAG-Enhanced Safety Information Retrieval for Construction: Integration of Large Language Models with Domain-Specific Information

Xianxiang Zhao[1], Advik Mehta[2], Falak Sethi[2], Brian Gue[3], Qipei Mei[4], Lingzi Wu[1]

[1]Department of Construction Management, University of Washington, United States
[2]Department of Computer Science, University of Alberta, Canada
[3]PCL Industrial Management Inc, Canada
[4]Department of Civil and Environmental Engineering, University of Alberta, Canada

seanzhao@uw.edu, advik@ualberta.ca, fsethi@ualberta.ca, bmgue@pcl.com, qipei.mei@ualberta.ca, lingwu@uw.edu

Abstract –
In the construction industry, critical safety information is often scattered across numerous documents, standards, and regulations, making it challenging for practitioners to access and comprehend safety knowledge in their daily operations efficiently. To address this challenge, we propose an intelligent and reliable question-answering system for information retrieval and response generation on the construction health, safety, and environment documents via retrieval-augmented generation. Specifically, our system combines a fine-tuned LLaMA-3-8B base model with a vector database constructed using embedding models, enabling accurate information retrieval and enhancing the generated responses' reliability. Initial validation using cosine similarity analysis demonstrates promising results, with our system achieving a cosine similarity score of 0.936, outperforming the LLAMA3-8B base model's score of 0.884 in processing construction safety documentation. The preliminary findings show that: 1) our RAG-enhanced system provides safety information access, and 2) our specialized preprocessing techniques effectively synthesize and retrieve safety information, reducing fragmentation and access time.

Keywords –
Construction Safety; Information Retrieval; Large Language Model (LLM); Retrieval-Augmented Generation (RAG); Question Answering System; Vector Database

1 Introduction

Construction remains one of the most hazardous industries globally, with persistent workplace accidents and injuries despite technological advancements [1]. A significant challenge in ensuring construction safety lies in the scattered nature of safety information across numerous documents, standards, and regulations, making it difficult for practitioners to efficiently access and utilize comprehensive safety information in their daily operations.

Current approaches to construction safety documentation typically rely on traditional documentation systems and manual searches, which have proven inefficient and time-consuming for construction practitioners [2]. These conventional methods often require workers to navigate through extensive documentation to find specific safety information, a process that can be particularly challenging given the dynamic and transient nature of construction operations [3]. The construction industry would benefit from an intelligent and reliable system that can quickly retrieve relevant safety information from various sources and present it in a user-friendly format through natural language interaction.

Traditional construction safety information management relies heavily on manual document searching and interpretation, which is time-consuming and prone to human error. Current approaches lack the capability to efficiently integrate multiple safety information sources while providing natural language interaction. To bridge this gap, this study leveraged large language model (LLM), a type of artificial intelligence system trained on vast amounts of text data, and retrieval-augmented generation (RAG) technology to enable precise information retrieval from various safety document sources. By integrating LLM with RAG, our system provides automated, accurate responses to safety-related queries using construction safety documentation as its knowledge base.

The key contributions of this research included: (1)

proposal and development of a novel RAG-based system that enhanced construction safety information accessibility through automated retrieval and natural language generation, and (2) empirical validation of the system's accuracy and effectiveness in a controlled construction safety management environment.

The remainder of this paper is organized as follows: Section 2 reviews relevant literature, Section 3 presents the prototype development, Section 4 details system validation and evaluation, with conclusions and future directions discussed in Sections 5 and 6.

2 Literature Review

This section reviews the development and limitations of safety information retrieval systems in construction, focusing on the progression from traditional keyword-based approaches to advanced language models. We examine how different technological solutions have attempted to address the challenges of efficient information retrieval, particularly highlighting the persistent limitations in query processing and response generation that impact the practical utility of these systems in construction safety applications.

Many natural language processing (NLP) applications have been developed to retrieve construction safety information, each with limitations. Early research focused primarily on discrete tasks such as incident report classification and safety information extraction from unstructured text [4,5]. These traditional NLP approaches have typically treated natural language understanding (NLU) and natural language generation (NLG) as separate components, resulting in fragmented solutions. In 2023, Ricketts et al. highlighted that fully automating safety processes through NLP could be problematic without a strong underlying system [6], particularly when handling complex safety queries that require both accurate understanding and contextual response generation.

The emergence of LLMs has introduced new possibilities for addressing these limitations. Unlike conventional NLP methods, LLMs leverage pre-training on vast, diverse corpora, enabling generalization across various tasks without task-specific data preparation [7,8]. However, while LLMs have demonstrated superior capabilities in analyzing safety incidents [9], commercially available solutions remain too general-purpose for specialized construction safety applications and raise significant data security concerns.

Current construction safety information retrieval primarily relies on manual searches and keyword-based retrieval, which have shown significant limitations in practice. Studies have demonstrated that these document-centric systems require users to navigate through multiple sources and extensive text documents to find specific safety information, thus failing to meet construction workers' diverse needs [10,11]. For example, when workers need to access relevant safety protocols during on-site operations, they often need to search through various safety manuals, regulatory documents, and incident reports separately, making the process time-consuming and risking incomplete information retrieval.

Recent research has demonstrated how RAG technology can enhance model reliability by incorporating accurate, context-specific information during the generation process in healthcare, education, and materials science domains [12–14]. In healthcare, RAG has improved clinical decision-making, pharmacogenomics interpretation, and disease-specific information delivery through specialized applications like GastroBot [12]. In education, RAG has enhanced intelligent tutoring systems and domain-specific knowledge delivery, particularly in technical fields like geotechnical engineering [13]. In materials science, RAG has supported engineering analysis and design processes by enabling access to comprehensive materials databases and enhancing knowledge graph integration [14]. Despite these promising applications across various fields, no study has implemented RAG to enhance construction safety information retrieval, according to our knowledge.

To summarize, research gaps exist in the following two critical aspects: 1) current approaches lack the ability to integrate accurate information retrieval with natural language interaction seamlessly, and 2) existing systems are unable to effectively synthesize and retrieve safety information from multiple sources, leading to fragmented and time-consuming information access processes.

3 System Development

Our proposed system was built upon the LLaMA-3-8B foundation model and integrated the BGE-large embedding model for semantic search capabilities, as shown in Figure 1.

The system architecture comprises three main components: (1) a vector database powered by the BGE-Large embedding model for semantic search functionality, (2) a fine-tuned LLaMA3-8B language model for natural language understanding and response generation, and (3) a context integration module that combines retrieved information with query processing.

The system processes user queries through these components in a streamlined workflow. When a user submits a query, it simultaneously triggers a semantic search in the vector database to retrieve relevant safety documentation while being processed by the language model. The context integration module then combines these elements to generate accurate, contextually grounded responses based on the safety documentation. This architecture ensures reliable safety information

delivery while maintaining natural language interaction capabilities.

The development of our system followed two essential steps to ensure its effectiveness. First, we established a comprehensive data processing pipeline to prepare construction safety documentation for the system, as detailed in subsection 3.1. Second, we conducted careful model selection and fine-tuning to optimize the system's performance in the construction safety domain, as detailed in subsection 3.2.

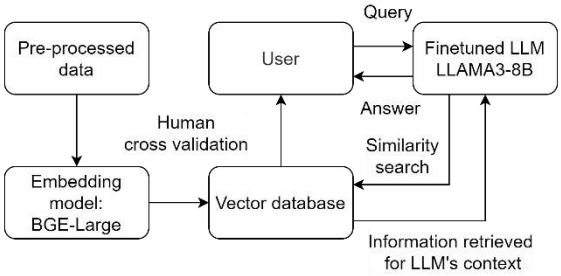


Figure 1 System Design

3.1 Data Processing Pipeline

The data processing pipeline has three key components: data collection, document pre-processing, and information organization and retrieval. First, we established a comprehensive health safety and environment (HSE) information data source as our foundation, as detailed in subsection 3.1.1. Next, a document pre-processing layer converted these diverse document formats into standardized JSON representations while preserving critical safety information, as detailed in subsection 3.1.2. Finally, we implemented an information organization and retrieval system that enabled efficient access to safety information through vector-based search and maintained traceability to original documents, as detailed in subsection 3.1.3.

3.1.1 Data Collection

We obtained a comprehensive collection of health safety and environmental (HSE) documentation from a leading construction company for our information retrieval system. The dataset comprised 1,299 pages and 327,336 words of safety protocols, guidelines, and procedures, as shown in Table 1. This extensive documentation set served as the source material for our information retrieval system. According to industry experts, our documentation set represents standard industry HSE protocols and procedures.

Table 1 HSE Data Source

Document Type	Count
Core HSE Manuals	3
Operational Forms	23
PPE Specifications	11

3.1.2 Document Pre-processing

Our document pre-processing transformed raw safety documents into structured data, as shown in Figure 2. The process consisted of two main steps: content extraction and visual processing.

In the content extraction step, we leveraged the Unstructured library to systematically parse PDF documents [15]. This library enabled precise extraction of three types of content while maintaining the structural relationships between different document elements: tables, images, including diagrams and illustrations, and text.

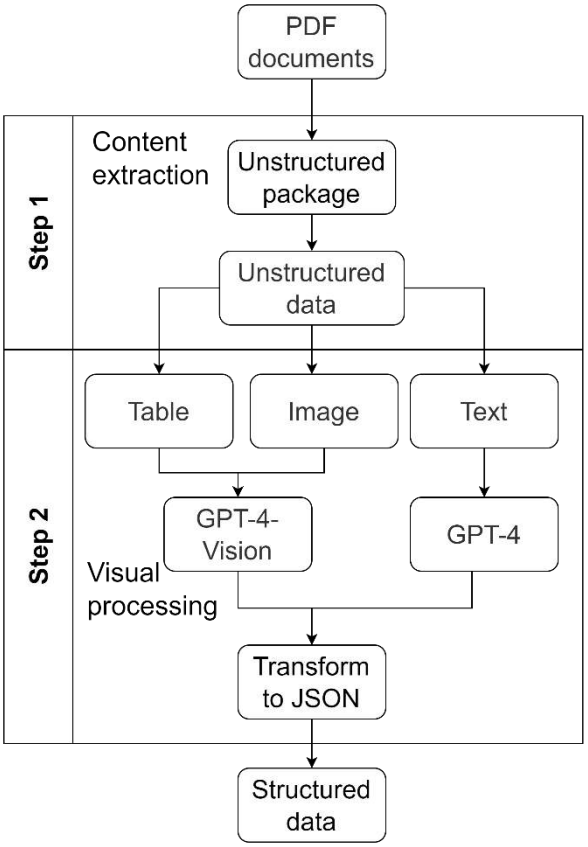


Figure 2 Document Pre-processing Layer

The extracted tables and images presented unique challenges. We addressed these through the implementation of GPT-4-Vision [16], which generated detailed textual representations of the visual content. This approach ensured that critical safety information contained in visual formats was preserved and made accessible for subsequent processing.

All processed content was then consolidated into a

standardized JavaScript Object Notation (JSON) structure. This transformation was crucial for construction safety documentation, where the relationships between different content types (e.g., textual instructions, diagrams, and tables) often had significant safety implications. The resulting structured data served as the foundation for generating the fine-tuning dataset and constructing the vector database.

3.1.3 Information Retrieval

The information retrieval process is the key component of our system, implementing similarity search and traceability mechanisms to effectively process and access construction safety documentation.

Our information retrieval system utilizes the BGE-large-en-V1.5 embedding model to convert HSE document into high-dimensional vectors [17,18], enabling similarity search capabilities. We selected this model for its superior performance in capturing technical and domain-specific relationships in construction safety content.

To effectively manage and trace information and their reference source, we developed a traceability framework with multi-vector retrieval capabilities. We designed and implemented a custom JSON mapping system that preserves structural relationships while facilitating efficient retrieval. Our system pre-processes documents, extracts nodes, and preserves comprehensive metadata to maintain bidirectional traceability between processed content and source documents.

3.2 Model selection and Fine-tuning

Model selection and fine-tuning enhances our system's capability from general knowledge to domain-specific expertise in construction safety information retrieval. Fine-tuning complements our RAG framework by enabling the model to better interpret construction safety terminology, recognize domain-specific relationships, and effectively integrate retrieved content into coherent responses. This synergy between fine-tuning and RAG is essential, as neither component alone would achieve the accuracy and relevance needed for construction safety applications. This process consisted of two main stages: base model selection and model fine-tuning, as detailed in subsections 3.2.1 and 3.2.2.

3.2.1 Base Model Selection

Our base models were selected considering the following reasons: 1) less than 8 billion parameters due to GPU memory limitation of our training hardware (RTX A6000), and 2) open source for data privacy protection and security when processing construction safety documents. Based on the context window and massive multitask language understanding (MMLU) score as listed in Table 2, we selected Llama-3-8B. MMLU score is a comprehensive benchmark that evaluates models across 57 subjects, including science, engineering, and professional knowledge.

Table 2 Model Selection Criteria and Performance

Model	Context Window	MMLU
LLaMA-3-8B	**8K**	**66.6**
Gemma-7B	8K	64.3
GPT-Neo-2.7B	2K	25.0

3.2.2 Model Fine-tuning

Our model's fine-tuning consists of two main steps: (1) Q-A pairs generation from structured dataset using Langchain Tuna, and (2) model fine-tuning through LLAMA Factory. Figure 3 illustrates this fine-tuning pipeline.

First, we utilized Langchain Tuna to generate high-quality training data from our structured construction safety dataset. This tool created 11,837 structured question-answer (Q-A) pairs that comprehensively cover various aspects of construction safety protocols [19]. These Q-A pairs are designed to maintain technical accuracy and contextual relevance for safety documentation. Figure 4 presents an example of these generated Q-A pairs.

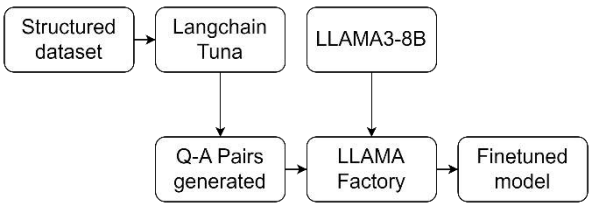


Figure 3 Fine-tuning steps

As our second step, our model's fine-tuning leverages the LLAMA Factory framework [20] to enhance the LLaMA-3-8B model's understanding of construction safety protocols. We chose Low-Rank Adaptation (LoRA) technology [21] for the fine-tuning, as it enables efficient domain adaptation while preserving the model's fundamental capabilities. We used LLAMA Factory using GPUs with CUDA v12 (RTX A6000), with carefully selected hyperparameters shown in Table 3 based on empirical experiments and model characteristics:

Table 3 Fine-tuning Parameters

Parameter	Value
Batch Size	2
Learning Rate	5.00E-05
Training Epochs	3
LoRA+ Ratio	16

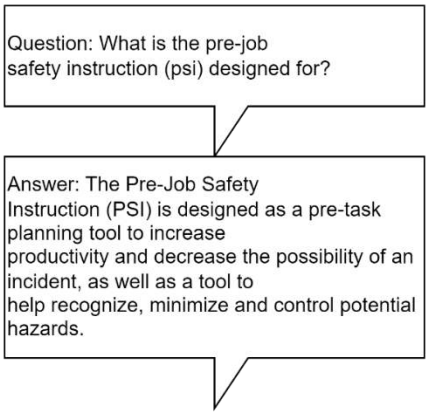


Figure 4 An example of Q-A Pairs generated by Langchain Tuna

While our system architecture incorporates both RAG and fine-tuning, it is important to clarify their distinct yet complementary roles in enhancing construction safety information retrieval. RAG primarily addresses the challenge of information access by retrieving relevant context from the vector database, while fine-tuning addresses the challenge of domain understanding by adapting the model to construction-specific terminology and concepts. Fine-tuning is essential in our approach because construction safety documentation contains specialized terminology, regulatory references, and technical concepts that general-purpose models may not adequately understand. Without domain adaptation through fine-tuning, even a model with perfect retrieval capabilities would struggle to correctly interpret construction-specific queries or synthesize retrieved information into coherent, domain-appropriate responses.

4 Validation and Evaluation

We compared our model with the base LLaMA-3-8B model to evaluate its performance. Our validation and evaluation followed three steps: test case selection, quantitative analysis, and qualitative analysis, as detailed in subsections 4.1, 4.2, and 4.3.

4.1 Test Case Selection

For test case selection in our validation and evaluation, we identified a representative scenario focusing on crisis response protocols from a safety operational manual. We used one specific test question ("What are the steps for Crisis Response?") as our validation case. This case was specifically selected to evaluate our system's ability to handle complex procedural information—a seven-step protocol with sequential dependencies and decision points. Such protocols represent a particularly challenging test case for safety information systems as they require both accurate retrieval of each procedural step and correct preservation of the sequential relationships between steps. This type of complex procedural information is abundant in construction safety documentation and essential for proper safety management. By successfully processing this challenging protocol, our system demonstrates its reliability for handling the diverse information needs in construction safety applications. This case was chosen to evaluate our model's comprehensive understanding capabilities in two key aspects: (1) the ability to process and interpret complex textual information in safety procedures, and (2) the capability to accurately comprehend and convert visual flowchart representations into structured data. The source information was originally presented as a visual flowchart (Figure 5), which was pre-processed using GPT-4V for text conversion and subsequently stored in our vector database.

We designed the test case query as "What are the steps for Crisis Response?" This query was selected to assess the model's capabilities in: comprehending and articulating complete protocol sequences, handling interconnected steps in safety procedures, and maintaining accuracy in complex information retrieval scenarios.

The selection of this test case is based on the real-world case, as protocol-based information represents a substantial portion of construction safety documentation. The ability to accurately retrieve complete step sequences is crucial for maintaining safety compliance in real-world applications.

We present the comparative responses from the LLaMA-3-8B model and our model in Figure 6 and Figure 7, respectively, demonstrating their performance differences in handling this complex query.

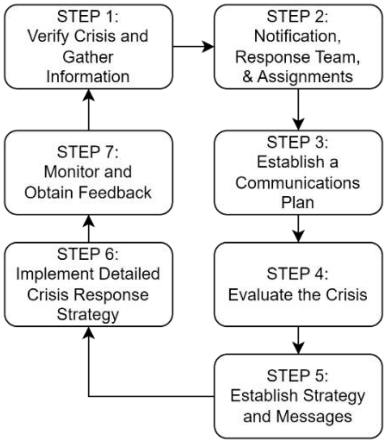


Figure 5 An example adapted from the graph information in the PDF document

4.2 Quantitative Analysis

To evaluate our model's performance quantitatively, we employed two complementary approaches: ROUGE (Recall-Oriented Understudy for Gisting Evaluation) for assessing text generation quality [22], and cosine similarity for measuring semantic relevance between responses and reference documents [23].

The ROUGE metrics we adopted provide a comprehensive evaluation framework that includes precision (accuracy of generated content), recall (coverage of reference content), and F1 score (harmonic mean of precision and recall). This multidimensional analysis through ROUGE metrics offers a complete view of the system's performance beyond simple text overlap. Our evaluation examined three ROUGE variants: ROUGE-1 calculates unigram overlap to assess content coverage, ROUGE-2 measures bigram overlap to evaluate fluency, and ROUGE-L employs the longest common subsequence algorithm to evaluate structural similarity. Additionally, cosine similarity provides a direct measure of semantic alignment between the model's outputs and the source safety documentation.

Following standard practice in NLP evaluation [24], we pre-processed all texts for consistent tokenization and normalization. As shown in Table 4, the experimental results demonstrate substantial improvements across all metrics, indicating that our model better captures both content accuracy and structural coherence in construction safety responses.

As shown in Table 4, our model demonstrates substantial improvements across all evaluation metrics. It is important to note that each ROUGE score reported represents the F1 measure, which balances precision (the percentage of generated content that is accurate) and recall (the percentage of reference content successfully captured). The significant improvements in ROUGE scores (44.9%-60.4%) indicate that our model not only produces more accurate content but also achieves superior coverage of the essential safety information. These comprehensive metrics collectively validate our system's effectiveness for construction safety information retrieval and response generation tasks.

The cosine similarity improvement from 0.884 to 0.936, while appearing modest in absolute terms, is particularly meaningful considering the high baseline. As visualized in Figure 6 and Figure 7, this enhancement manifests in the model's responses being more focused and domain aware. The base model's responses (Figure 6) tend to be more generic, while our model (Figure 7) provides more precise, construction safety-specific responses with clear source references.

Table 4 Model Selection Criteria and Performance

Metric	Our Model	Base Model	Improvement
ROUGE-1	70.8%	10.4%	60.4%
ROUGE-2	47.6%	2.7%	44.9%
ROUGE-L	67.7%	7.3%	60.4%
Cosine Similarity	93.6%	88.4%	5.2%

4.3 Qualitative Analysis

From a qualitative perspective, the comparison between model responses (Figure 6 and

Figure 7) clearly demonstrates the superiority of our model. When processing the crisis response flowchart, our model accurately captured and reproduced the exact procedural steps and decision points from the source documentation. As shown in Figure 7, our model provides direct references to the original safety documentation, with each reference corresponding to specific sections in the source HSE manuals. These references enable users to verify information and access additional context from the original documents, thereby enhancing the traceability and reliability of the provided information.

For instance, when asked about emergency response procedures, our model correctly identified the sequence "Initial Assessment → Hazard Evaluation → Response Selection" without introducing additional, unsourced steps or procedures. In contrast, the base model tended to supplement the flowchart information with general safety protocols do not present in the original documentation, demonstrating hallucination tendencies. The base model's responses often included generic safety knowledge and theoretical frameworks that, while potentially valid, were not part of the source material. Our model's ability to strictly adhere to the source content while maintaining the logical flow of the cyclic graph structure showcases its enhanced reliability for safety-

critical information retrieval tasks.

```
>>> What are the steps for Crisis Response
Crisis response is a structured approach to managing
 unexpected events or crises that threaten an organi
zation,
its reputation, or its stakeholders. The following s
teps outline a comprehensive crisis response plan:

1. **Detection and Initial Response**:
        * Identify the crisis situation quickly.
        * Activate the crisis management team (CMT)
or incident response team.
        * Establish communication channels with affe
cted parties (e.g., employees, customers, media).
        * Provide initial guidance and reassurance t
o stakeholders.
2. **Assessment and Situation Analysis**:
        * Gather facts about the crisis situation.
        * Assess the severity of the crisis and its
potential impact on the organization.
        * Identify key stakeholders affected by the
crisis (e.g., employees, customers, regulatory agenc
ies).
3. **Communication Strategy**:
        * Develop a clear and consistent communicati
on strategy to address stakeholder concerns.
        * Designate a single point of contact for me
dia inquiries and other external communications.
        * Establish protocols for internal communica
```

Figure 6 Answer from llama3-8B base model



Figure 7 Answer with source references

5 Discussion and Conclusion

This study addressed the critical challenge of efficiently retrieving and processing construction safety information, traditionally hindered by complex documentation structures and technical terminology barriers. We proposed an RAG architecture that combines fine-tuned LLaMA-3-8B with specialized document preprocessing capabilities. Our system demonstrated superior performance in handling construction safety queries through a natural language interface while maintaining high accuracy in information retrieval.

Unlike previous studies that rely on commercial LLMs with data privacy concerns, our approach leverages a locally deployable, smaller model (LLaMA-3-8B) enhanced through domain-specific fine-tuning and RAG architecture. Our approach provides enhanced security for sensitive safety information and greater customizability through construction-specific adaptation.

The reliability of our system was systematically validated through multiple complementary approaches. The quantitative evaluation showed substantial improvements across all metrics, with our model achieving significantly higher ROUGE scores than the base model and improved cosine similarity from 88.4% to 93.6%. These technical improvements manifested in more precise, construction safety-specific responses with better source reference integration, as evidenced in our system's outputs (Figure 7). Particularly noteworthy was our system's ability to correctly process and present the seven-step crisis response protocol, demonstrating its capability to handle complex procedural information with high fidelity. This performance validates both the retrieval capabilities of our RAG architecture and the domain understanding provided by fine-tuning.

This research addresses critical gaps in construction safety information management through two key contributions: (1) the development of a RAG architecture that seamlessly integrates accurate information retrieval with natural language interaction, enabling intuitive access to safety information, and (2) the implementation of specialized document preprocessing techniques that effectively synthesize and retrieve safety information from multiple sources, significantly reducing information fragmentation and access time.

This research has the following limitations. First, while the system excelled at information retrieval, it did not address broader information management challenges such as information creation, organization, and updating. Additionally, its performance varied across different types of construction safety documentation, and its effectiveness in actual construction site conditions needed further validation. Second, the system's capabilities were confined to English-language documentation, restricting its utility in international construction contexts. These limitations suggested opportunities for future research in developing more comprehensive and multilingual implementations.

6 Acknowledgements

This project was supported by the University of Alberta Research Exploration Fund. We would also like to express our gratitude to PCL Industrial Management Inc. for their continued support throughout the projects.

7 References

[1] G.M. Waehrer, X.S. Dong, T. Miller, E. Haile, Y. Men, Costs of occupational injuries in construction

in the United States, Accident Analysis & Prevention 39 (2007) 1258–1266. https://doi.org/10.1016/j.aap.2007.03.012.

[2] J. Teizer, Right-time vs real-time pro-active construction safety and health system architecture, CI 16 (2016) 253–280. https://doi.org/10.1108/CI-10-2015-0049.

[3] L. Wu, S. AbouRizk, Towards Construction's Digital Future: A Roadmap for Enhancing Data Value, in: S. Walbridge, M. Nik-Bakht, K.T.W. Ng, M. Shome, M.S. Alam, A. El Damatty, G. Lovegrove (Eds.), Proceedings of the Canadian Society of Civil Engineering Annual Conference 2021, Springer Nature Singapore, Singapore, 2023: pp. 225–238. https://doi.org/10.1007/978-981-19-1029-6_17.

[4] R. Ganguli, P. Miller, R. Pothina, Effectiveness of Natural Language Processing Based Machine Learning in Analyzing Incident Narratives at a Mine, Minerals 11 (2021) 776. https://doi.org/10.3390/min11070776.

[5] M. Locatelli, E. Seghezzi, L. Pellegrini, L.C. Tagliabue, G.M. Di Giuda, Exploring Natural Language Processing in Construction and Integration with Building Information Modeling: A Scientometric Analysis, Buildings 11 (2021) 583. https://doi.org/10.3390/buildings11120583.

[6] J. Ricketts, D. Barry, W. Guo, J. Pelham, A Scoping Literature Review of Natural Language Processing Application to Safety Occurrence Reports, Safety 9 (2023) 22. https://doi.org/10.3390/safety9020022.

[7] S.A. Prieto, E.T. Mengiste, B. García de Soto, Investigating the Use of ChatGPT for the Scheduling of Construction Projects, Buildings 13 (2023) 857. https://doi.org/10.3390/buildings13040857.

[8] H. Aladağ, Assessing the Accuracy of ChatGPT Use for Risk Management in Construction Projects, Sustainability 15 (2023) 16071. https://doi.org/10.3390/su152216071.

[9] M. Smetana, L. Salles De Salles, I. Sukharev, L. Khazanovich, Highway Construction Safety Analysis Using Large Language Models, Applied Sciences 14 (2024) 1352. https://doi.org/10.3390/app14041352.

[10] A.M. Basahel, Safety Leadership, Safety Attitudes, Safety Knowledge and Motivation toward Safety-Related Behaviors in Electrical Substation Construction Projects, IJERPH 18 (2021) 4196. https://doi.org/10.3390/ijerph18084196.

[11] C.I. Arum, T.O. Osunsanmi, C.O. Aigbavboa, Appraisal of the Challenges to Ensuring Occupational Health and Safety Compliance within the Nigerian Construction Industry, Mocs (2019) 488–495. https://doi.org/10.29173/mocs130.

[12] Q. Zhou, C. Liu, Y. Duan, K. Sun, Y. Li, H. Kan, Z. Gu, J. Shu, J. Hu, GastroBot: a Chinese gastrointestinal disease chatbot based on the retrieval-augmented generation, Front. Med. 11 (2024) 1392555. https://doi.org/10.3389/fmed.2024.1392555.

[13] A. Tophel, L. Chen, U. Hettiyadura, J. Kodikara, Towards an AI Tutor for Undergraduate Geotechnical Engineering: A Comparative Study of Evaluating the Efficiency of Large Language Model Application Programming Interfaces, (2024). https://doi.org/10.21203/rs.3.rs-4658661/v1.

[14] M.J. Buehler, Generative Retrieval-Augmented Ontologic Graph and Multiagent Strategies for Interpretive Large Language Model-Based Materials Design, ACS Eng. Au 4 (2024) 241–277. https://doi.org/10.1021/acsengineeringau.3c00058.

[15] Unstructured-IO/unstructured, (2024). https://github.com/Unstructured-IO/unstructured (accessed December 13, 2024).

[16] U. Lee, Y. Jeong, J. Koh, G. Byun, Y. Lee, H. Lee, S. Eun, J. Moon, C. Lim, H. Kim, I see you: teacher analytics with GPT-4 vision-powered observational assessment, Smart Learn. Environ. 11 (2024) 48. https://doi.org/10.1186/s40561-024-00335-4.

[17] P. Zhang, S. Xiao, Z. Liu, Z. Dou, J.-Y. Nie, Retrieve Anything To Augment Large Language Models, (2023). https://doi.org/10.48550/arXiv.2310.07554.

[18] S. Xiao, Z. Liu, Y. Shao, Z. Cao, RetroMAE: Pre-Training Retrieval-oriented Language Models Via Masked Auto-Encoder, (2022). https://doi.org/10.48550/arXiv.2205.12035.

[19] Introducing Tuna - A Tool for Rapidly Generating Synthetic Fine-Tuning Datasets, LangChain Blog (2023). https://blog.langchain.dev/introducing-tuna-a-tool-for-rapidly-generating-synthetic-fine-tuning-datasets/ (accessed December 21, 2024).

[20] Y. Zheng, R. Zhang, J. Zhang, Y. Ye, Z. Luo, Z. Feng, Y. Ma, LlamaFactory: Unified Efficient Fine-Tuning of 100+ Language Models, (2024). https://arxiv.org/abs/2403.13372 (accessed December 13, 2024).

[21] E.J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, W. Chen, LoRA: Low-Rank Adaptation of Large Language Models, (2021). https://doi.org/10.48550/arXiv.2106.09685.

[22] C.-Y. Lin, ROUGE: A Package for Automatic Evaluation of Summaries, in: Text Summarization Branches Out, Association for Computational Linguistics, Barcelona, Spain, 2004: pp. 74–81.

https://aclanthology.org/W04-1013 (accessed December 21, 2024).
[23] F. Rahutomo, T. Kitasuka, M. Aritsugi, Semantic Cosine Similarity, (n.d.).
[24] S. Bird, Natural Language Processing with Python, 2009.

Enhancing Data Management Efficiency in Precast Concrete Manufacturing Through a Custom Database Solution

Alaa Abu Nokta[1], Mohammad Darwish[1], and Mohamed Al-Hussein[1]

[1] Department of Civil and Environmental Engineering, University of Alberta, Edmonton, AB, Canada
abunokta@ualberta.ca, mdarwish@ualberta.ca, malhussein@ualberta.ca

Abstract

It is crucial in complex production environments, such as precast concrete manufacturing, to have efficient data management. Construction enterprises must navigate and organize vast amounts of data as part of their critical decision-making functions. The research described in this paper uses a case study to examine the implementation of a custom database solution for data management in a precast concrete manufacturing facility. The database is developed to address existing organizational issues and production inefficiencies stemming from the absence of a centralized data management system. By automating and organizing data retrieval, the database streamlines data flow and enhances the accessibility and integrity of critical information. Moreover, a dashboard is developed to display essential performance indicators, such as project tracking, estimates of future revenue, and production schedules. The system manages various types of data, from management to production information. The case study highlights the importance of integrated data management systems in improving productivity, quality control, and overall performance in the construction industry.

Keywords

Data Management, Precast Concrete, Modular Construction, Database

1 Introduction

The application of data-driven approaches in construction projects has increased significantly in recent years, reflecting a growing recognition of the crucial role these approaches can play in project lifecycle management [1]. This expansion of data-driven applications within the construction industry is primarily attributable to the emergence and maturation of knowledge-based management [2], which facilitates innovation, ensures effective quality control [3], and can be used to manage large, complex datasets in an efficient manner [4]. Construction projects typically operate in complex environments, necessitating precise coordination among stakeholders, materials, and production processes. Consequently, the use of integrated database systems has become increasingly critical as a means of handling the substantial volumes of data generated throughout a facility's lifecycle [5].

In today's construction industry, efficient management of data, which is extensive and sourced from multiple stages and stakeholders, remains a challenge [6]. Traditionally, data management and information exchange in construction have been performed manually [7], resulting in wasted time, increased costs, and frequent data loss and inconsistencies stemming from human error [8]. Therefore, there is a critical need to transition from manual methods to robust technologies that ensure consistency and quality, as well as automated database management solutions that minimize these risks and support precise, informed decision-making.

Precast concrete manufacturers face unique industry-specific challenges, such as the need for optimized resource allocation, streamlined production workflows, and enhanced project monitoring, and this further emphasizes the importance of database solutions explicitly tailored to the needs of this subsector [9]. Data-centric structures have been demonstrated to be particularly effective in safeguarding data quality, maintaining security, and ensuring that critical information is accurate and readily accessible [10]. Moreover, the diversity and complexity of the data sources associated with construction necessitate comprehensive, versatile databases populated with data from real-life projects [11].

The studies outlined above provide a foundation for understanding the benefits of database integration in construction project management. In this context, this paper presents a data management framework that provides a solution to the industry's need for historical data tracking and a centralized data system. The main objective and contribution of the paper is the development of a structured database tailored to enterprises producing ribbed-structure precast concrete panels. The developed database enhances data integrity,

ensures consistent quality control, and enables efficient data storage while supporting informed decision-making.

The database was developed using data collected through a case study conducted at a precast concrete manufacturing facility in Edmonton, Canada. The database encompasses records of various aspects of management, production, geometric information of the panels, installation, material tracking, project tracking, and quality control.

2 Background

2.1 Literature review

Knowledge management plays a crucial role in project-based industries such as construction. However, the distinctive characteristics of construction projects often pose challenges with respect to the effective capture and utilization of information [3]. Dawood and Marasini [9] emphasized the significance of tailored database solutions in addressing specific challenges faced by precast concrete manufacturers, highlighting advantages such as optimized resource allocation, streamlined production workflow, and improved project monitoring. Similarly, Mathew Zachariah et al. [10] underscored the capability of data-centric structures in preserving data quality and security, ensuring that critical information remains accurate and accessible.

Christensen et al. [11] noted that the diversity of data significantly influences a database's overall quality. Supporting this perspective, Batselier and Vanhoucke [12] emphasized that comprehensive and versatile databases incorporating data from real-life projects are essential for integrated project management. Such databases facilitate real-time project tracking and reporting, thus enabling informed decision-making and proactive problem-solving [13, 14].

Recent studies have highlighted the value of dashboards in visually representing real-time data. Nimbarte et al. [15] emphasized the critical role dashboards play in improving user engagement and clarity in decision-making. They noted that dashboards effectively enhance communication among team members and offer a centralized platform for monitoring project progress, financial metrics, and key performance indicators (KPIs).

The Entity Relationship (ER) diagram has also been widely recognized as an essential component of database design. ER diagrams graphically illustrate relationships among entities such as people, objects, locations, or events, providing the foundational structure for relational databases [16]. ER diagrams are especially useful for clearly depicting data structures and relationships [17].

In summary, the literature underscores the utility of customized database solutions in addressing the specific challenges associated with precast concrete manufacturing, emphasizes the role of knowledge management in enhancing data quality and security, and highlights the advantages of dashboards and ER diagrams in improving data visualization and system design. Collectively, these insights establish the foundation for the structured database solution presented herein.

2.2 Case company's concrete panel system

The case company produces a wide range of precast concrete products and custom-built solutions for the residential and commercial markets. Its unique reinforced ribbed concrete panel system features vertical and horizontal ribs. The panel's structure includes layers of plywood, rebar, Styrofoam, wire mesh, and concrete. The panels are designed to provide structural and thermal strength, enhance durability, and reduce concrete usage. Figure 1 shows the layers of the panel's ribbed structure.



Figure 1: Breakdown of the case company's concrete panel system

2.3 Case company's project lifecycle

The case company's projects typically follow the lifecycle shown in Figure 2. The cycle begins with consultative meetings between management and the client to clarify the project's scope of work and to specify design criteria. The project management team then issues a quote to the client. If the client accepts the quote, a contract is negotiated and executed, and the project proceeds to the engineering stage. At this point, the



Figure 2: Project lifecycle

permit and production drawings are issued for production. The project then advances to production, at which stage precast panels are formed and poured. The project then proceeds to the final stage, installation.

2.4 Problem statement

Variations in product specifications, as well as the different attributes linked with each project, underscore the need for a system that combines work orders and organizes data flow within the enterprise while also serving as a resource for data reference. Such a system must support the following functions:

- Store client information (e.g., contact information).
- Input new project information into the system as projects are received from clients.
- Update the status of projects during each phase of their lifecycle.
- Generate production orders once a project is approved.
- Store panel information and bills of materials.
- Track concrete orders and the compressive strength testing results for the concrete panels manufactured.
- Store material and supplier information to support procurement and issue purchase orders.
- Provide reports to track sales and provide overall performance insights.

3 Methodology

The primary objective of the research presented in this paper was to create a comprehensive database solution that addresses the need to have a system that protects data integrity, tracks performance, forecasts production schedules, and generates reports as needed. MS Access was used to structure the database due to its user-friendly interface, cost-effectiveness, and ability to create tables, queries, forms, and reports. Another reason for choosing MS Access is that the case company is relatively small, such that having a single employee manage the database is sufficient. Its cost-effectiveness makes it an ideal fit for the case company's needs, offering robust functionality without the need for extensive IT resources. Moreover, given that there are plans to link the database to a simulation model that will be created using the Simphony.NET in the future, the interoperability between MS Access and the Simphony.NET simulation environment is another notable advantage.

Tables were built to contain critical information related to clients and projects, and they were connected using the appropriate relationship types. Additionally, the database was designed to facilitate tracking of project status and to generate reports of relevant project information. Each table is crucial in organizing and managing specific aspects of projects, clients, and resources. A structured database design ensures data integrity, efficient data retrieval, and comprehensive reporting for various stakeholders within the organization. In order to appropriately structure the database, the case company's workflow was studied. The flowchart shown in Figure 3 was built to represent the company's production process.

The process begins with gathering information about the project as the basis for establishing a scope of work. The project management team prepares a quote and issues it to the client. Upon the client's approval, the project will proceed to the engineering phase, which involves preparation of production drawings of the panels and a bill of materials. Purchase orders are issued based on materials' suppliers and prices, the bill of materials, and the project information. Following this phase, concrete is ordered. Following forming and pouring in the off-site production facility, quality control testing is conducted, marking the completion of the production phase. The final stage is the installation of precast components at the site.

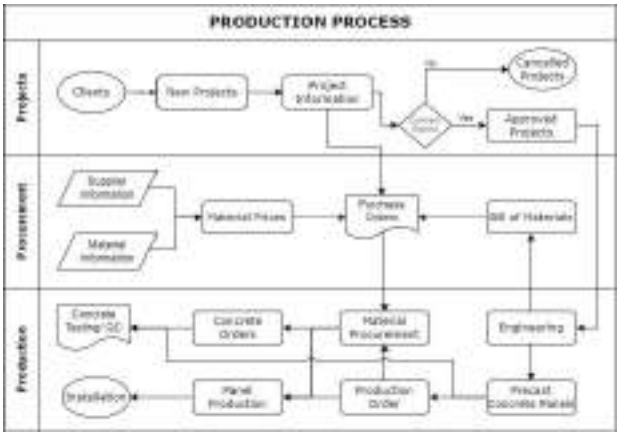


Figure 3: Production flowchart

4 Database development

4.1 Entity-Relationship diagram

An ER diagram was created to support the development of a database skeleton, complete with relationships between the various entities across different tables. The ER diagram served as a crucial blueprint for creating the database, visually representing the database's structure and the relationships among various entities. This diagram not only aided in understanding the intricacies of the data model, but also formed the foundation for effective database design. The ER diagram guided the decision-making process by outlining entities, relationships, and attributes, ensuring a coherent and organized database. Figures 4 and 5 illustrate the different entities, their attributes, and the relationships among them.

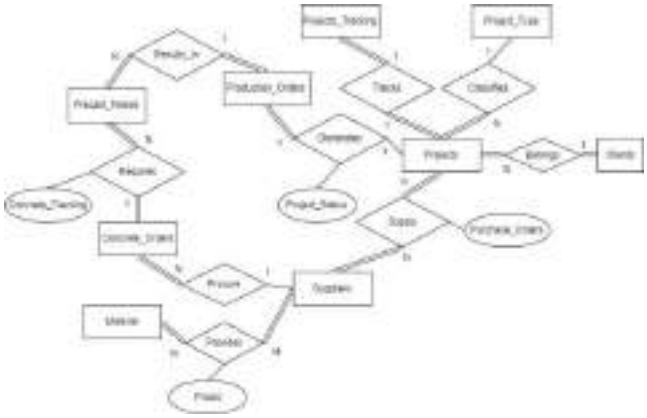


Figure 4: ER Diagram

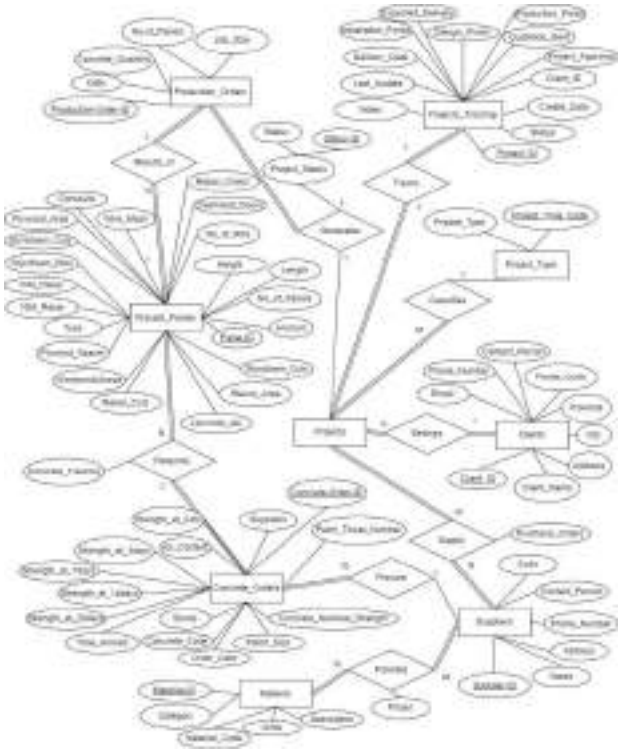


Figure 5: High-level ER Diagram

4.2 Database tables

The database tables (designed in MS Access) and their assigned attributes closely follow the proposed ER diagram. What follows is a summary of the tables and their use:

- *Clients*: This table is crucial for maintaining client information, including contact details. It provides centralized client data storage, facilitating access for various departments.
- *Project_type*: This table defines various project types, serving as a point of reference for categorizing and searching projects. It also ensures consistency in project-type definitions across the database.
- *Projects_info*: This table stores detailed project information, including cost estimates and scopes of work. It centralizes project-specific details for efficient project management and facilitates easy tracking and reporting on project-related metrics.
- *Project_tracking*: This table provides a comprehensive overview of project progress by tracking project status and key dates. It enables real-time monitoring of project timelines and milestones and facilitates communication among team members by documenting important project events.
- *Production_orders*: This table manages production orders and tracks concrete quantities. It organizes orders and links them to projects, streamlining the production process. Additionally, it provides insights into resource allocation and project-specific production requirements.
- *Precast_panels*: This table details the design specifications of precast panels. It offers a comprehensive view of panel characteristics, aids in quality control, and supports the accurate estimation of material and resource requirements.
- *Concrete_orders*: This table manages concrete orders, recording details such as supplier information and concrete specifications (e.g., strength). It tracks concrete usage across projects for resource planning purposes and facilitates quality control by recording concrete strength at various stages.
- *Panels_concrete_tracking*: This table tracks concrete usage for precast panels. It provides a detailed account of concrete application on panels and aids in troubleshooting and quality assurance for panel-related concrete work.
- *Project_status*: This table defines various project statuses, standardizes project status definitions for consistent reporting, and enhances communication by providing a common language for tracking project progress.
- *Materials*: This table centralizes information about materials used in projects. It tracks material details to support efficient inventory management and facilitate project cost analysis and budgeting.
- *Suppliers*: This table stores details about suppliers, including contact information and rates. It enables quick access to supplier information for procurement activities and supports the establishment of relationships via purchase orders and prices.
- *Prices*: This table records prices of materials from

different suppliers. It facilitates cost estimation and budgeting for projects and streamlines the process of selecting suppliers based on pricing information.

- *Purchase_orders*: This table manages purchase orders, including details such as quantity, price, and delivery information. It tracks procurement activities, ensuring timely and accurate order fulfillment. Additionally, the table provides transparency in financial transactions related to project expenses.

4.3 Entity relationships

The relationships between tables in the database help to connect the data and facilitate a cohesive project and client management system. The Clients table is a foundational element, establishing connections with multiple other tables. The Projects_info table is linked to the Clients table through the "Client_Code," connecting each project to a specific client. The Project_type table provides a reference point for categorizing projects, which in turn determines the "project_type_code" in the Projects_info table.

The Project_tracking table is connected to the Projects_info table through the "Project-ID," enabling seamless tracking of project statuses and key dates. The Production_orders table is linked to both the Projects_info and Clients tables, providing a bridge between production activities and the specific projects and clients they pertain to. The Precast_panels table also establishes connections with the Production_orders and Projects_info tables, ensuring that details about precast panels are associated with the corresponding projects and production orders.

The Concrete_orders and Panels_concrete_tracking tables connect to the Projects_info and Production_orders tables, recording information about concrete supply and usage in relation to specific projects and production orders. Additionally, the Prices and Purchase_orders tables are linked to the Suppliers table, forming relationships that aid in tracking material costs, procurement details, and supplier information. These interconnected relationships create a comprehensive network, allowing for efficient information retrieval, accurate project progress tracking, and seamless coordination among the various elements of the company's operations.

4.4 Database dashboard

A dashboard for this database was developed to showcase the results by enhancing the user experience and providing valuable insights into the company's projects and financial status. The dashboard consists of twelve forms for entering information into the database and twelve reports for information retrieval.

Figure 6 provides an overview of the developed dashboard.

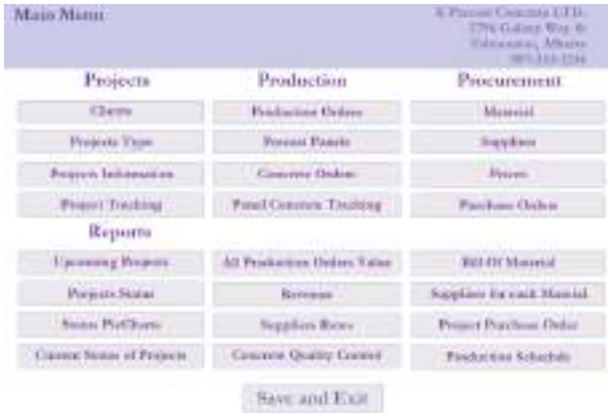


Figure 6: Developed dashboard

Below is a brief description of the reports generated by the developed dashboard:

- *Upcoming Projects*: Displays a list of approved future projects, providing details such as the project's name, client, and expected completion date.
- *Project Status*: Utilizes a bar chart and table to categorize projects based on their status (e.g., in progress, completed, or pending approval).
- *Status Pie Chart*: Incorporates a pie chart reflecting the distribution of projects across different statuses, providing both the project count and their cumulative value.
- *Current Status of Projects*: Allows users to input a specific project status, generating a report that details the projects in that particular status, including all relevant information.
- *Production Value*: Presents a comprehensive report listing the details of projects progressing to production orders, along with their associated values.
- *Revenue Report*: Generates a report showcasing completed projects and their values, enabling quick assessment of the revenue generated.
- *Suppliers Rates*: Includes a table summarizing the prices of materials from different suppliers, facilitating comparisons and procurement decisions.
- *Concrete Quality Control*: A detailed report linking concrete orders to the panels they were used for, including testing results for quality control purposes.
- *Bill of Materials*: Offers a breakdown of materials required for each project by panel, with a summary of total materials for efficient project planning and resource allocation.
- *Suppliers for Each Material*: Enables users to input a material type, generating a list of suppliers to aid in procurement decisions.

- *Project Purchase Order*: Lists all purchase orders related to the queried project, aiding in tracking of procurement activities.
- *Production Schedule*: Displays projects scheduled for production within a given date range, aiding in resource planning and scheduling.

The developed reports and dashboard provide a user-friendly interface for efficient project and client management, informed decision-making, and comprehensive data analysis.

5 Results

The developed database was implemented at the case company and was found to greatly improve project performance by streamlining processes such as purchase order management, reducing material waste, and providing more accurate production forecasts. Tailored reports, such as the "Upcoming Projects" and "Project Status" reports, were found to offer valuable insights into project progress and financial implications. The database, integrated with a dynamic dashboard, enabled real-time monitoring of KPIs and visual representations (e.g., "Status PieChart") that aided in identifying trends and bottlenecks in the project lifecycle. Figures 7, 8, 9, and 10 display some of the key outputs from the generated reports, including the exported concrete quantity report, project status percentage and value charts, and summary of upcoming projects.

It should be noted that the database is specifically designed to handle the unique requirements of precast concrete panel production. The database was found to improve communication and is now in use by the case company as an essential decision support tool. The developed database with user-friendly dashboard offers several benefits:

- Efficient monitoring: Tracks project status and performance metrics to support timely interventions.
- Enhanced productivity: Provides quick access to information for streamlined decision-making.
- Strategic planning: Offers a clear view of KPIs, helping management to identify areas for improvement.

While the use of a database for information storage and retrieval in construction management is not novel, the database developed in this study is structured to accommodate the unique design specifications and geometric information associated with ribbed precast concrete components. This approach ensures optimal performance and addresses the distinct needs of this construction subsector. Future work will involve creating a simulation model representing the case company's production process and integrating it with the developed database for estimation purposes.



Figure 7: Concrete quality control report

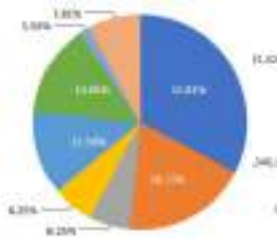


Figure 8: Percentage-based project status chart

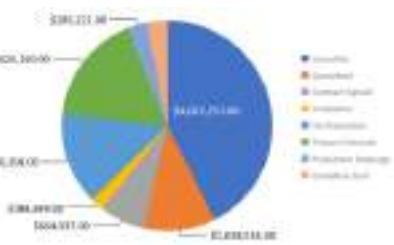


Figure 9: Financial value-based project status chart

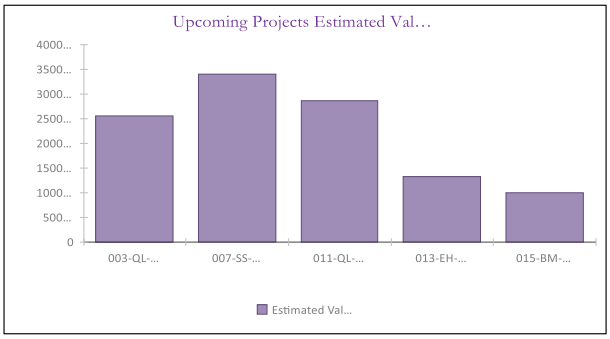


Figure 10: Chart representing the value of upcoming projects alongside their expected revenues

References

[1] Y. Huang, Q. Shi, J. Zuo, F. Pena-Mora, and J. Chen. Research status and challenges of data-driven construction project management in the big data context. *Advances in Civil Engineering*, 2021:674980, 2021.

[2] B. Ozorhon, C. G. Karatas, and S. A Demirkesen. web-based database system for managing construction project knowledge. *Procedia - Social and Behavioral Sciences*, 119:377–386, 2014.

[3] G. Eken, G. Bilgin, I. Dikmen, and M. T. Birgonul. A lessons learned database structure for construction companies. *Procedia Engineering*, 123:135–144, 2015.

[4] B. Thalheim. Component construction of database schemes. In *Lecture Notes in Computer Science: Proceedings of the International Conference on Conceptual Modeling*, 2503:20–34, 2002.

[5] L. Li, S. Msaadi, H. Manier, and M.-A. Manier. The creation of the hydrogen supply chain decision database. *Energies*, 16(24):8081, 2023.

[6] M. Bilal, O. L. Oyedele, J. Qadir, K. Munir, O. S. Ajayi, O. O. Akinade, and M. Pasha. Big data in the construction industry: A review of present status, opportunities, and future trends. *Advanced Engineering Informatics*, 30(3):500–521, 2016.

[7] M. Kantardzic. *Data Mining: Concepts, Models, Methods, and Algorithms*. Canada: John Wiley and Sons, Inc., 2011.

[8] H. T. T. Le, V. Likhitruangsilp, and N. Yabuki. A BIM-integrated relational database management system for evaluating building life-cycle costs. *Engineering Journal*, 24(2):75–86, 2020.

[9] N. Dawood and R. Marasini. An integrated database for real-time management of stockyard "stockman": A case study in precast concrete products industry. In *Proceedings of the CIB-W78 International Conference*, 2001.

[10] G. C. Mathew Zachariah, S. Santhosh, S. Anandhrosh, S. Thomas, and C. Mathew. Database and modern database technology. *International Research Journal on Advanced Engineering and Management*, 2(12):3680–3686, 2024.

[11] D. Christensen, R. C. Antolini, and J. W. McKinney. A review of estimate at completion (EAC) research. *Cost Analysis and Management*, 12(1):41–62, 1995.

[12] J. Batselier and M. Vanhoucke. Construction and evaluation framework for a real-life project database. *International Journal of Project Management*, 33(3):697–710, 2015.

[13] M. Martínez-Rojas, N. Marín, and M. A. Vila. The role of information technologies to address data handling in construction project management. Journal of Computing in Civil Engineering, 30(4): 04015064, 2015.

[14] W. Yandong and G. Jianya. A construction schema for provincial spatial database of China. Geo-spatial Information Science, 12(1):25–32, 2009.

[15] A. D. Nimbarte, N. Smith, and B. Gopalakrishnan. Human factors evaluation of energy visualization dashboards. Ergonomics in Design, 32(1):39–47, 2021.

[16] S. Hettiarachchi, C. Sugandhika, A. Kathriarachchi, S. Ahangama, and G. T. Weerasuriya. A scenario-based ER diagram and query generation engine. In *Proceedings of the 4th International Conference on Information Technology Research*, Moratuwa, Sri Lanka, 2019.

[17] C. Ordonez, S. Tahsin Al-Amin, and L. Bellatreche. An ER-flow diagram for big data. In *Proceedings of the IEEE International Conference on Big Data*, pp. 5795–5797, Atlanta, GA, USA, 2020.

Enhancing Circular Construction Through Structural Steel Reuse: Case Study of Champlain Bridge

Dina Abouhelal[1] and Amin Hammad[1]

[1]Gina Cody School of Engineering and Computer Science, Concordia University, Canada

dina.abouhelal@mail.concordia.ca , amin.hammad@concordia.ca

Abstract – The construction industry plays a critical role in tackling the challenges of climate change, carbon emissions, and resource consumption. Current practices predominantly adhere to a linear supply chain model that gives insufficient attention to the End-of-Life (EOL) phase of structures. Transitioning to Circular Construction (CC) is essential for minimizing waste and mitigating negative environmental impacts. While recycling structural steel is widely utilized at the EOL, it is energy-intensive and may degrade material quality. In contrast, steel component reuse presents a more sustainable option, requiring minimal alterations and significantly reducing embodied energy and Greenhouse Gas emissions. The main aim of this study is to advance CC by developing a framework for the catalyzing factors that can enhance the efficacy of the deconstruction process and structural steel component reuse. These factors were gathered through a comprehensive literature review. The identified factors were then mapped to a high-profile deconstruction project, the original Champlain Bridge deconstruction, to assess the challenges and initiatives involved in improving deconstruction efficiency and optimizing the reuse of structural steel components. The data of the deconstruction process and structural steel reuse were obtained through series of meetings with the Jacques Cartier and Champlain Bridges Incorporated (JCCBI) team. The numerical data regarding quantities and weights of structural steel was analyzed to assess the types of reusable steel components. Finally, the paper highlights future research areas for improving the efficiency of the value chain processes, maximizing the steel component reuse, improving the decision-making at the EOL.

Keywords –

Deconstruction; Circular Construction; Infrastructure Sustainability; Structural Steel Reuse; Adaptive Reuse; End-of-Life Management; Champlain Bridge.

1 Introduction

Effective construction waste management has gained paramount importance in recent years due to increasingly stringent regulations focusing on material recycling and reuse [1]. The construction industry is responsible for about 40% of the global natural resource exploitation [2, 3]. The current practices in the construction industry follow a linear supply chain model (take-make-dispose) with minimum consideration of the End-of-Life (EOL) phase of the structures [4]. There is a need to change this model to a circular model that aims to keep resources in use for as long as possible, which will minimize waste and reduce the negative environmental impact [5, 6]. Adopting Circular Construction (CC) is highly dependent on the way of handling the structure's components at the EOL phase, where two main options are available: demolition or deconstruction [7].

Demolition involves dismantling the structure, with most steel being recycled, although some steel end up in landfills due to contamination or recycling limitations. Deconstruction is a more methodical approach that focuses on carefully disassembling the structure to maximize the recovery of valuable components [8]. Following deconstruction, recovered materials can be managed through two main pathways: reuse or recycling. In practice, a combination of reuse and recycling often occurs. Recycling is a common approach for handling materials at the EOL, but it entails energy-intensive processes, such as steel melting and fabrication, which can elevate energy consumption. Additionally, these processes may compromise material quality, potentially leading to increased demand for new materials in future applications [9]. Alternatively, reuse often presents a more sustainable approach, as it involves minimal physical alterations [10]. Furthermore, the embodied energy and Greenhouse Gas emissions of the structures made of reused components are 63% and 56% lower, respectively, than those of the new components [11]. However, implementing deconstruction and reuse practices necessitates radical changes in the construction industry, moving away from the traditional approach of fixed-lifespan structures to a more flexible approach that

facilitates deconstruction and component reuse [12]. This will require the adoption of new business models and technologies that prioritize deconstruction and support CC [13].

While several studies have recognized the importance of the EOL phase of existing steel structures, and the opportunity of their component reuse, there is a significant gap in the development of a comprehensive methodology that integrates the catalyzing factors necessary to facilitate the reuse of structural steel after deconstruction [14 – 16]. Accordingly, this paper aims to promote CC by developing a framework that enhances deconstruction efficiency and facilitates the reuse of structural components, incorporating key factors and actionable insights to support sustainable deconstruction strategies. The study is structured into two main stages. The first stage involves identifying the catalyzing factors that promote the efficient reuse of structural components, encompassing technological, regulatory, and economic influences, which were sourced from various scholarly references. The second stage involves mapping these factors within a high-profile deconstruction project, the original Champlain Bridge deconstruction project in Montreal, Canada, to identify successes and challenges associated with the deconstruction process and the reuse of structural steel components. Finally, the research highlights future directions for enhancing value chain processes, promoting component reuse, and improving decision-making through the integration of advanced technologies.

2 Catalyzing Factors for Enhancing Structural Component Reuse

This section focuses on catalyzing factors aimed at enhancing the efficiency of the deconstruction process and maximizing the reusability of structural components. These factors were determined through a comprehensive review of recent research papers, conference proceedings, books, and review articles.

2.1 Structure Attributes

(1) Structural Design: The design of structures plays a crucial role in determining both the efficiency of the deconstruction process and the potential reusability of their components. Steel structures are not only highly recyclable, but also easy to dismantle and store [17]. The modular design of structural components is a potential approach for designing structures that can easily adapt, be moved, dismantled, and reused [18]. Mechanical connections such as bolts and fasteners can facilitate easier disassembly during the deconstruction process. Properly designed connections that allow for straightforward removal without damage to the components or surrounding materials simplify the extraction of reusable parts [19].

(2) Component Condition: The condition of the components and desired reuse application influence the level of refurbishment required. The examination of several key subfactors related to a structure's condition, including structural integrity, historical value, and cost-effectiveness informs decisions and strategies in managing deconstruction, emphasizing practicality and sustainable outcomes [20].

(3) Location: The strategic location provided easy access for heavy machinery and transportation, enabling efficient removal and transport of salvaged materials. Efficiently matching of supply and demand for reused components reduces logistics costs in deconstruction projects [21].

2.2 Business Attributes

(1) Willingness to Reuse: Finding a client willing to take a risk and pay a premium for CC or reuse solutions decreases the cost risks for value chain actors [22].

(2) Organization Competences and Circular Readiness: The project team’s compatible skills, and innovative approach facilitate creative solutions for implementing structural component reuse, resulting in high reuse rates and reduced transition costs and time for efficient circular processes [23].

2.3 Value Chain Activities

(1) Deconstruction Planning: Effective deconstruction planning is crucial to minimize the risks of deconstruction processes and ensure the maximization of reusable structural components [24]. Digital technologies enhance the efficiency in inventorying and tracking components. Also, visualizing the deconstruction process over time using 4D simulation aids in efficient resource allocation, minimizes disruptions, and improves cost efficiency through identifying and eliminating non-value-added activities [25]. Quality management ensures that salvaged components meet standards for reuse. The use of heavy machinery enables the safe dismantling of structures. Implementing a reusing and recycling program promotes sustainability by diverting materials from landfills and integrating them into new construction projects [19]. For certifying reused steel or quality control procedures in the steel industry, both ISO [26] and ASTM [27] standards provide essential frameworks. These standards involve evaluations such as material non-destructive testing and inspections of structural integrity to ensure the steel's reliability and safety in new applications. Environmental monitoring ensures compliance with regulations and minimizes environmental impacts [21].

(2) Safety Planning: Safety planning in the

deconstruction process is critical for mitigating risks and protecting both workers and the environment. This involves conducting risk assessments, training programs, regular site safety audits, and an emergency response plan [20, 21].

(3) Logistics Optimization: Careful logistics planning minimizes environmental and economic costs associated with component transportation, preserving the value potential of reusable components and reducing transportation expenses [28]. Value chain actors can efficiently access information on available reusable components and their locations. Moreover, verifying the environmental benefits of refurbished components in Environmental Product Declarations enhances demand for building component reuse. Furthermore, establishing metrics for tracking progress, component recovery rates, and types of components salvaged, enables real-time adjustments and reporting on reuse success for future projects [19].

(4) Storage Conditions: Good-quality storage enables the value of salvaged components to be preserved and minimizes risk of additional costs [21].

(5) Lean Processes: Lean processes could meticulously identify each step involved in deconstruction, allowing for the elimination of unnecessary tasks, optimization of workflows, and minimization of wastes [19]. Quality control measures, such as regular inspections and evaluations, enforce rigorous standards for reusable materials and minimize defects. Moreover, training programs provide workers with the necessary skills and knowledge in best refurbishment practices, reducing errors and enhancing safety [29].

2.4 Project Management

(1) Public Consultation and Stakeholders' Involvement: An effective advertising campaign for sustainable deconstruction educates the public and promotes awareness and support. Collaboration among stakeholders streamlines processes and maximizes resource reuse. Working with nearby city administrations aligns regulations and enhances regional planning. Pilot projects provide experiential learning, helping refine best practices. Influencing industry attitudes encourages broader adoption of sustainable practices, driving long-term resource conservation in the built environment [30].

(2) Performing Feasibility Study: The feasibility study can improve deconstruction efficiency by evaluating methods that align with sustainability goals, focusing on minimizing environmental impact and maximizing component recovery [3].

2.5 Legislations

Regulatory support is crucial with policies that streamline permitting processes for projects incorporating salvaged components, thereby facilitating the efficient reuse and promoting environmental sustainability in the construction sector. Sustainability programs initiated by governments prioritize sustainable construction practices, incentivizing the reuse of components from deconstruction projects. Legislative financial incentives encourage the usage of reused components in construction, increasing profitability and accelerating the construction sector's transition to circular practices [31].

Several initiatives are being actively implemented to enhance the reuse of steel components, focusing on reducing waste and advancing sustainability within the construction sector. For example, the Toronto Green Standard (TGS) sets environmental performance standards to accelerate GHG reduction, aiming for near-zero emissions by 2030. Higher tiers mandate embodied carbon assessments (e.g., LCA) to measure material and construction emissions. Compliance is enforced through the Site Planning Approval process [32].

3 Champlain Bridge Deconstruction and Steel Component Reuse

The original Champlain Bridge is one of the largest infrastructure projects in North America. It was built in 1962 by Jacques Cartier and Champlain Bridges Incorporated (JCCBI). It has a total length of 3.4 km, with six lanes for vehicular traffic, three in each direction. The original Champlain Bridge faced several significant structural problems that contributed to its eventual deconstruction. The primary issue was severe corrosion, exacerbated by de-icing salts and an inadequate drainage system in its original design. The absence of a proper drainage system in the bridge's design led to the accumulation of water and salt on critical structural elements, worsening corrosion and compromising the integrity of the bridge. Throughout its service life, the bridge required frequent and significant maintenance activities, including repainting the steel structure, repairing piers, replacing the concrete deck with a steel deck, and reinforcing girders. Despite these efforts, the bridge's deterioration continued to worsen, compromising its performance and reliability [33]. In 2013, the need to replace the original Champlain Bridge was further confirmed when sensors detected unusual deformations in the edge girders. In response, the bridge was closed, and a 75 tonne, 56-meter steel "super beam" was installed to reinforce a cracked girder, reducing traffic capacity by one lane. Five months later, the super beam was replaced with 50 tonnes of steel modular trusses. Despite these measures, the aging bridge's condition continued to deteriorate, highlighting the need for a full replacement [34].

The planning of the original Champlain Bridge's

deconstruction began in 2017. The planned deconstruction cost was around Can$225 million, while the overall investment cost was Can$400 million. Design-build of the deconstruction was chosen as the project delivery method, and work on-site began in July 2020. The deconstruction process finished in November 2023.

3.1 Mapping the Catalyzing Factors to the Case of Champlain Bridge Project

The data were derived from the JCCBI website [34], the steel components catalogue [35], consultation reports [36], the competition report [37], alongside meetings with the JCCBI team.

(1) Structure Attributes: The main part of the original Champlain Bridge was steel truss cantilever with approach viaducts constructed of prestressed concrete beams supporting a prestressed concrete deck paved with asphalt. A total of 413 standardized steel components, with a total weight of 401 tonnes, were identified for potential reuse in other steel construction projects. Mechanical connections enable easier component separation during deconstruction, using techniques like heat cutting for large elements, unbolting, and removing rivets.

For the component condition, the bridge's deconstruction produces approximately 25,000 tonnes of steel, with structurally viable components for salvaging and repurposing. Moreover, some components were used for non-residential structures (e.g., car canopy) which were not subject to high structural requirements, reducing bureaucracy and enabling the simplification or avoidance of cost-intensive steps related to element recovery, inspections, repairs, and certification.

Pertaining to the bridge location, the proximity of the bridge to the surrounding infrastructures facilitates the deconstruction process and the reuse of components by providing easy access for logistics, as well as enabling efficient removal and transport of salvaged materials. Also, the bridge's legacy and historical significance added intrinsic value to its components, making them appealing for preservation and reuse in heritage projects. The urban setting increased demand for reused components, promoting interest from local builders and developers in acquiring salvaged components. Moreover, due to the environmental considerations, the contractor has to propose a greenhouse gas compensation strategy for ensuring a carbon-neutral project.

(2) Business Attributes: JCCBI set sustainability targets for handling and transportation, emphasizing the innovative reuse of components rather than traditional recycling. They also implemented continuous performance monitoring to improve the project's profitability. Furthermore, a sustainability decision support tool was created, incorporating environmental, economic, social, and technical factors. The tool was utilized to evaluate different scenarios and compare all the available deconstruction methods, possible ways for the reuse/recycling of the components, and transportation methods.

(3) Value Chain Activities: JCCBI enhanced logistics by devising a phased deconstruction strategy that prioritizes the safe removal of components while minimizing disruption to surrounding areas. A comprehensive assessment was conducted to evaluate the bridge's structural integrity and determine which components were suitable for reuse. Also, they developed a 4D model for the bridge's deconstruction to facilitate comprehensive planning, scheduling, and coordination. Visualizing the deconstruction process over time aids in efficient resource allocation, minimizes disruptions, and improves cost efficiency through identifying and eliminating unnecessary activities. Moreover, non-destructive tests were performed to evaluate the quality of the bridge components. The structural steel components were dismantled using specialized equipment, including cranes, barges, and excavators. Additionally, JCCBI implemented a reusing and recycling program as they directed the contractor to meet specific reuse and recycling targets, including 100% steel reuse or recycling and a 90% overall component recovery from the bridge, with penalties for non-compliance. The contractor was bound by the design-build requirements to recycle the components within a 150 km radius from the project site.

Moving to safety planning, the deconstruction process involved 200 specialists and laborers with specialized training in safe deconstruction practices, machinery operation, and emergency response protocols. Regular safety audits and inspections were carried out to ensure compliance with safety regulations and to address any emerging risks. An emergency response plan was established, detailing procedures dealing with accidents, injuries, or other unforeseen incidents.

For logistics optimization, a phased deconstruction strategy was developed to prioritize the safe removal of components while minimizing disruption. To facilitate data access and flows, a data labeling system was implemented for each reusable component. Additionally, a traceability protocol using QR codes was established to monitor the materials, enabling precise georeferencing of each component from its origin to its destination. Reusable components were systematically stored on-site in designated areas specifically allocated for sorting and inventory management. These storage areas were equipped with protective measures to shield components from adverse weather conditions and potential damage.

To mitigate deconstruction waste, JCCBI integrated specific clauses into the design-build contract of the deconstruction to establish a comprehensive

sustainability program. This program prioritizes the reuse of components wherever feasible, thereby reducing waste and advancing sustainable practices throughout the project. For quality control, they conducted regular inspections and evaluations of reusable steel components. Also, they provided training for employees on best practices for refurbishment.

(4) Project Management: The community was engaged through an extensive multiplatform advertising campaign to gather innovative ideas regarding sustainable development, environmental protection, and component reuse. The consultation involved 4,186 participants via various channels. An advisory committee was established consisting of representatives from different environmental and cultural organizations to provide strategic guidance and expertise throughout the project. Moreover, JCCBI searched similar projects and identified the Bay Bridge deconstruction project in San Francisco as a relevant example [33]. The California Department of Transportation initiated a program to repurpose components of the bridge but reused less than 1% of the bridge's steel. In contrast, JCCBI developed its reuse program earlier in the project, aiming to achieve a significantly higher percentage of component reuse.

In 2017, JCCBI commissioned the PTA Consortium (Parsons/Tetra Tech/Amec Foster Wheeler) to conduct a feasibility study and targeted environmental analysis for the Champlain Bridge deconstruction, assessing different options for deconstruction methods, transportation of components, and component recovery. The study evaluated various scenarios based on sustainable development criteria, focusing on technical feasibility, economic viability, environmental impacts, and social acceptability.

(5) Legislations: The Government of Canada has shown a robust commitment to sustainability, requiring all ministries and agencies to contribute to the "United Nations Sustainable Development Goals" and align with Canada's national sustainability strategy [38]. Consequently, the deconstruction of the original Champlain Bridge is designed to promote environmental performance for future projects, reflecting both national and international sustainability objectives. The deconstruction of the original Champlain Bridge and the reuse of its components have been supported by various financial incentives, including public funding sources such as local tax revenues, federal funds, fares, and bonds.

3.2 Data Analysis

The authors conducted an analysis of the data provided in the catalogue to gain insights into the quantities and weights of different types of reusable steel components of the Champlain Bridge [35]. The methodology involved classifying the reusable steel components based on their location, followed by calculating the amount of reused steel components and the weight percentage for each category. Figure 1 shows the histogram of the reusable steel components in terms of their quantity. The horizontal axis categorizes the different types of structural components in the Champlain Bridge, including Diagonal, Bracing, Vertical, Top chord, Lower chord, and Deck beam. The vertical axis represents the number of reusable steel components, showing how many structural elements of each type were identified for reuse. This histogram visually compares the distribution of reusable components across these categories.

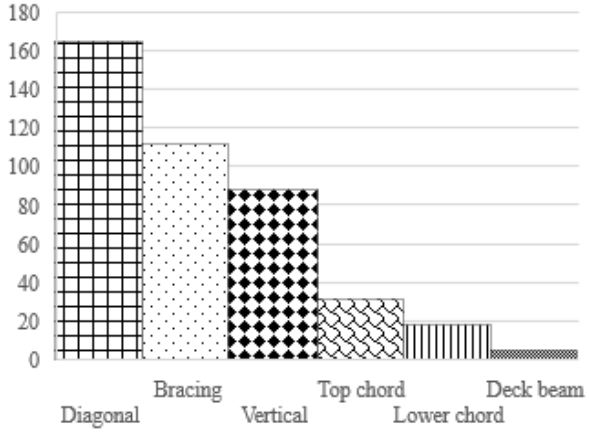


Figure 1. Quantity histogram of reusable steel components

The histogram shows that diagonal components are the most abundant, totaling 165 units. Bracing and vertical components follow with significant quantities of 112 and 88 units, respectively. Top and lower chord components have moderate quantities, at 31 and 18 units, while deck beams are the least numerous with only four units. Figure 2 shows the percentage of reusable steel components classified by their weight. The analysis of steel component weights reveals that the top chord components have the highest proportion by weight (23.7%). Lower chord components are significant in both quantity and weight, suggesting they are moderately heavy. Vertical components reflect a balance between their quantity and weight. While diagonal components are numerous, they are relatively lighter compared to components in other locations, indicating these components are heavier per unit. Deck beam components have low quantities but higher weight percentages (8.9%), suggesting they are dense despite their limited numbers.

The analysis indicates a varied distribution of steel in terms of quantity and weight, reflecting their different applications and structural roles. Understanding these distributions aids in planning the reuse of these components in future projects.

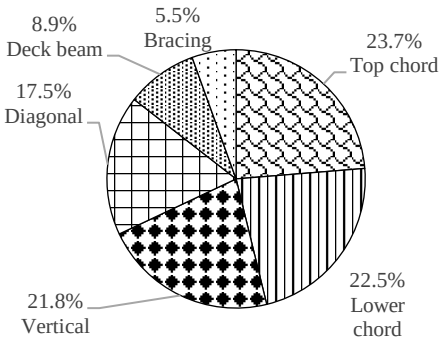


Figure 2. The weight percentage of reusable steel components

4 Future Research Areas

For improving the efficiency of the processes related to value chain activities, maximizing the reuse of components, and supporting the decision-making, the following future research areas can be considered:

(1) Automation in Deconstruction Planning and Optimization: AI algorithms can be applied for identifying the optimal sequence of steel components deconstruction based on their structural importance and condition. Also, it helps to simulate potential problems during the deconstruction process and automate the decision-making [39]. Furthermore, automated systems can classify components by category to facilitate repurposing [40].

(2) 3D Scanning and Digital Twin Technologies: Adopting 3D scanning and developing digital twin of the structure can assist in the efficient deconstruction processes and the precise planning for maximum reuse of steel components. It provides a highly accurate, real-time representation of the structure's condition, including the identification of critical components and potential structural weaknesses [41].

(3) Material Passport System: Material Passport is a digital record that details the composition, origin, lifespan, and condition of each component used in the structure. It has the potential to maximize the reuse of structure components, reducing costs, and environmental impact through matching deconstructed components to future construction projects. Radio Frequency Identification could be used to track the component throughout its lifecycle, facilitating the efficient identification and recovery of valuable components during deconstruction [42].

(4) Advanced Robotics: The use of autonomous robots or drones in the process of scanning, dismantling and sorting of bridge components can improve accuracy and reduce the associated risks. For instance, robots can be designed to safely remove steel components while conserving their integrity for reuse [43].

(5) Enhanced Workers Training with Virtual Reality (VR) and Augmented Reality (AR): VR and AR can ensure that workers are adequately trained for the deconstruction process. These technologies allow workers to simulate deconstruction scenarios and understand the best practices for handling components, improving safety, and maintaining the integrity of reusable components [44].

5 Conclusions

The transition toward CC is increasingly vital as the construction industry confronts pressing challenges related to climate change, carbon emissions, and resource depletion. Adopting a circular approach offers a sustainable solution by minimizing waste, conserving resources, and lowering embodied energy, making it a key strategy for addressing these critical issues. This study emphasizes the need for a shift towards CC, particularly through the reuse of structural steel components. It begins by identifying the key factors that could catalyze the deconstruction process and enhance the reuse of components. The factors include the structure attributes, business attributes, value chain activities, project management, and legislations. These factors were applied to the Champlain Bridge deconstruction project to evaluate process effectiveness, identify challenges, and define the potential for optimizing the reuse of structural steel components.

The analysis revealed several success criteria, including the development of a phased deconstruction strategy, comprehensive audits of structural integrity, and the integration of digital technologies for effective planning and coordination, thereby enhancing overall project efficiency. Additionally, incorporating specific sustainability clauses into the deconstruction contract significantly increases the potential for steel reuse. Furthermore, presenting a sustainability decision support tool to evaluate various deconstruction methods can facilitate informed decision-making. Engaging the community and collaborating with environmental organizations, along with organizing competitions for innovative reuse ideas, further fosters a culture of sustainability within the construction industry. Finally, the research provides future research areas for supporting the processes of the value chain activities, increasing the potential of reusing components, and enhancing the decision-making. The future research areas emphasize the integration of advanced technologies such as artificial intelligence for deconstruction planning, robotics for precision dismantling, 3D scanning for accurate modeling, and material passport systems for enhanced traceability. However, while this study provides valuable insights for enhancing deconstruction processes and the structural steel reuse, it lacks detailed quantitative

analysis for the environmental and economic benefits in comparison with the traditional steel recycling. Future research could focus on conducting scenario analysis for the EOL of structural steel utilizing sustainable life cycle approach. This would include a detailed comparison of the environmental and economic benefits of reuse relative to traditional recycling.

The research results could improve the decision-making process for EOL phase of structures by adding comprehensive qualitative analysis towards sustainable management and conservation of resources. These results align with Canada's international commitments to sustainable development, including the United Nations Sustainable Development Goals, particularly Goal 12 (Responsible Consumption and Production). Furthermore, this study provides numerical data and actionable insights that will support the Canadian government's objectives related to sustainability, economic efficiency, and environmental responsibility, contributing to a more sustainable construction industry.

Acknowledgement

The authors sincerely thank Mr. Moctar Sidibe, Senior Engineer at JCCBI, for his willingness to engage in meetings with our team, which significantly deepened our understanding of the deconstruction process and the reuse of structural steel components. We also extend our gratitude to other students in our research group (Hosna Ghorab and Golnaz Mirzaei) for their initiative in launching this collaborative effort.

References

[1] Geissdoerfer M., Pieroni M. P. P., Pigosso D. C. A., and Soufani K. Circular business models: A review. *Journal of Cleaner Production*, 277:123741, 2020.

[2] Sanchez B., Herthogs P., and Stouffs R. Identifying key parameters for BIM-based disassembly planning. In Proceedings of the International Symposium on Automation and Robotics in Construction, pages 32–39, 2023.

[3] Abouhelal D., Kamel W., and Bassioni H. Informatic Analysis and Review of Literature on the Optimum Selection of Sustainable Materials Used in Construction Projects: Critical review. *International Journal of Sustainable Construction Engineering and Technology*, 14(1):97–109, 2023.

[4] Guerra B. C. and Leite F. Circular economy in the construction industry: An overview of United States stakeholders' awareness, major challenges, and enablers. *Resources, Conservation and Recycling*, 170:105617, 2021.

[5] Pomponi F. and Moncaster A. Circular economy for the built environment: A research framework. *Journal of Cleaner Production*, 143:710–718, 2017.

[6] Nik-Bakht M., Pieroni M. P. P., Pigosso D. C. A., and Soufani K. Value Stream Mapping of Project Lifecycle Data for Circular Construction. In *Proceedings of the International Symposium on Automation and Robotics in Construction*, pages 1033–1042, 2021.

[7] Queheille E., Taillandier F., and Saiyouri N. Optimization of strategy planning for building deconstruction. *Automation in Construction*, 98:236–247, 2019.

[8] Akinade O. O., Kamardeen I., Akinade A. O., and Ajayi S. O. Waste minimisation through deconstruction: A BIM based Deconstructability Assessment Score (BIM-DAS). *Resources, Conservation and Recycling*, 105:167–176, 2015.

[9] Eberhardt L. C. M., Birkved M., and Birgisdottir H. Building design and construction strategies for a circular economy. *Architectural Engineering and Design Management*, 18(2):93–113, 2022.

[10] Zhuang G. L., Shih S. G., and Wagiri F. Circular economy and sustainable development goals: Exploring the potentials of reusable modular components in circular economy business model. *Journal of Cleaner Production*, 414:137503, 2023.

[11] J. Brütting, C. De Wolf, and C. Fivet, "The reuse of load-bearing components," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 225, no. 1, 2019.

[12] Cheng B., Huang J., Guo Z., Li J., and Chen H. Towards sustainable construction through better construction and demolition waste management practices: a SWOT analysis of Suzhou, China. *International Journal of Construction Management*, 23(15):2614–2624, 2023.

[13] Cristiano S., et al. Construction and demolition waste in the Metropolitan City of Naples, Italy: State of the art, circular design, and sustainable planning opportunities. *Journal of Cleaner Production*, 293:125856, 2021.

[14] Sanchez B. and Haas C. A novel selective disassembly sequence planning method for adaptive reuse of buildings. *Journal of Cleaner Production*, 183:998–1010, 2018.

[15] Mahpour A. Prioritizing barriers to adopt circular economy in construction and demolition waste management. *Resources, Conservation and Recycling*, 134:216–227, 2018.

[16] Etienne D., Lisa W., and Laetitia D. Evaluating 'reuse' in the current LCA framework - Impact of reuse and reusability in different life cycle stages. In *Proceedings of the IOP Conference Series: Earth and Environmental Science*, pages 1–12, 2022.

[17] Thai H. T., Ngo T., and Uy B. A review on modular construction for high-rise buildings. *Structures*, 28:1265–1290, 2020.

[18] Mao C., Shen Q., Shen L., and Tang L. Comparative study of greenhouse gas emissions between off-site prefabrication and conventional construction methods: Two case studies of residential projects. *Energy and Buildings*, 66:165–176, 2013.

[19] Minunno R., O'Grady T., Morrison G. M., Gruner R. L., and Colling M. Strategies for applying the circular economy to prefabricated buildings. *Buildings*, 8(9):125, 2018.

[20] Bertino G., Kisser J., Zeilinger J., Langergraber G., Fischer T., and Österreicher D. Fundamentals of building deconstruction as a circular economy strategy for the reuse of construction materials. *Applied Sciences*, 11(3):939, 2021.

[21] Riuttala M., Harala L., Aarikka-Stenroos L., and Huuhka S. How building component reuse creates economic value – Identifying value capture determinants from a case study. *Journal of Cleaner Production*, 443:141112, 2024.

[22] Dunant C. F., Drewniok M. P., Sansom M., Corbey S., Cullen J. M., and Allwood J. M. Options to make steel reuse profitable: An analysis of cost and risk distribution across the UK construction value chain. *Journal of Cleaner Production*, 183:102–111, 2018.

[23] Aarikka-Stenroos L., Chiaroni D., Kaipainen J., and Urbinati A. Companies' circular business models enabled by supply chain collaborations: An empirical-based framework, synthesis, and research agenda. *Industrial Marketing Management*, 105:322–339, 2022.

[24] Harala L., Alkki L., Aarikka-Stenroos L., Al-Najjar A., and Malmqvist T. Industrial ecosystem renewal towards circularity to achieve the benefits of reuse - Learning from circular construction. *Journal of Cleaner Production*, 389:135885, 2023.

[25] Sanchez B., Rausch C., Haas C., and Hartmann T. A framework for BIM-based disassembly models to support reuse of building components. Resources, *Conservation and Recycling*, 175:105825, 2021.

[26] ISO, "ISO 14040: Environmental Management - Life Cycle Assessment - Principles and Framework.," 2006.

[27] ASTM, "ASTM A370: Standard Test Methods and Definitions for Mechanical Testing of Steel Products," 2009.

[28] Whalen K. A., Milios L., and Nussholz J. Bridging the gap: Barriers and potential for scaling reuse practices in the Swedish ICT sector. *Resources, Conservation and Recycling*, 135:123–131, 2018.

[29] Adams K. T., Osmani M., Thorpe T., and Thornback J. Circular economy in construction: Current awareness, challenges and enablers. In *Proceedings of the Institution of Civil Engineers - Waste and Resource Management*, pages 15–24, 2017.

[30] Jayasinghe R. S., Rameezdeen R., and Chileshe N. Exploring sustainable post-end-of-life of building operations: A systematic literature review. *Engineering, Construction and Architectural Management*, 26(4):689–722, 2019.

[31] Devènes J., Bastien-Masse M., and Fivet C. Reusability assessment of reinforced concrete components prior to deconstruction from obsolete buildings. *Journal of Building Engineering*, 84:108584, 2024.

[32] Aileen Carroll, "Toronto Green Standard: Guide to TGS sustainable design requirements,"2025.

[33] Pollalis P. S. N. and Lappas D. *The Deconstruction of the Champlain Bridge Montreal, Canada: The Zofnass Program at Harvard*, volume 617. 2020.

[34] JCCBI. The Jacques Cartier and Champlain Bridges Incorporated | PJCCI. On-line: https://www.jccbi.ca, Accessed: 1/11/2024.

[35] JCCBI. "Catalog of reusable materials from the original Champlain Bridge," 2019. On-line: https://www.jccbi.ca, Accessed: 3/11/2024.

[36] JCCBI. "Champlain Bridge Deconstruction Public Consultation Report," 2019. On-line: https://www.jccbi.ca, Accessed: 3/11/2024.

[37] JCCBI. Material Reuse Competition | PJCCI. On-line: https://www.jccbi.ca, Accessed: 5/11/2024.

[38] ReNew Canada. "Building the structural steel reuse industry in Canada." On-line: https://www.renewcanada.net, 2024.

[39] Zhao J. and Liu Y. Application of artificial intelligence in construction and demolition waste management. *Journal of Construction Engineering and Management*, 146(10):04020119, 2020.

[40] Tiwari R. and Van der Meer T. Application of robotics and automation in the circular economy: A review of opportunities and challenges. *Waste Management*, 120:256–268, 2021.

[41] Bosch M. and Kallioras S. Digital twin technology and 3D scanning for construction and demolition projects: A review. *Automation in Construction*, 114:103173, 2020.

[42] Veleva V. and Ellenbecker M. Material passports: A new approach to building sustainability and resource efficiency. *International Journal of Life Cycle Assessment*, 19(5):976–987, 2014.

[43] Zhao X. and Hwang B. Applications of robotics in construction and demolition: A review of current trends and future directions. *Automation in Construction*, 106:102868, 2019.

[44] Bock T. and Linner T. The impact of virtual reality and augmented reality on the construction industry: A case study and review. *Journal of Construction Engineering and Management*, 141(7):04015010, 2015.

Reducing Carbon Emissions in 3D Printed RCC Slabs

Abhishek Patel1 and Benny Raphael1

1Indian Institute of Technology, Madras, India
abhi300894shek@gmail.com, benny@civil.iitm.ac.in

Abstract

Concrete 3D printing offers significant potential to revolutionize construction through improved efficiency and cost-effectiveness. However, high cement content in 3D printable mixes raises questions about the environmental sustainability of this technology. This study proposes an automated methodology for 3D printing Reinforced Cement Concrete (RCC) filler slabs using compressed polyethylene (PE) waste to reduce the carbon footprint. A gantry-style 3D printer with integrated pick-and-place functionality is designed to position recycled plastic fillers, reducing raw material consumption and emissions. A cradle-to-gate life cycle assessment is employed to compare three scenarios: conventional slabs, filler slabs with clay pots, and filler slabs with PE waste. The results show that the proposed method significantly reduce material use and carbon emissions, while mitigating the environmental impact of polyethylene waste mismanagement. This study demonstrates the potential of integrating 3D printing and automation to advance sustainable construction practices and lower the carbon footprint of RCC slabs.

Keywords – Concrete 3D printing, filler slabs, carbon footprint, carbon emission, automation, sustainable construction, recycling, polyethylene waste.

1 Introduction

Concrete 3D printing has potential to transform the construction industry in terms of productivity, cost effectiveness and environmental sustainability. 3D printing can help construction industry by improving productivity, overcoming skilled labor shortages [1–3], and increasing safety at off-site plants and on-site construction. 3D printing also eliminates the need for formwork, offering geometric flexibility and enabling significant material savings through structural optimization [4, 5]. Though Concrete 3D printing has various potential benefits, challenges like limited automation and dependency on manual labour, insufficient research on structural elements like slabs, and high cement content in 3D printable mixes persist.

The construction industry accounts for around 39% of global energy-related CO2 emissions, a figure expected to rise with urbanization [6]. Concrete, the second most used material worldwide due to its affordability and compressive strength [7], significantly contributes to this footprint, with cement production alone responsible for 5-8% of global CO2 emissions [8]. While concrete 3D printing reduces material waste and eliminates formwork, its high binder content elevates the carbon footprint, leaving its sustainability performance inconclusive. [9–11] have conducted studies to establish the sustainability performance of Concrete printing but they are limited to the non-structural components. The calculations for environmental impact in these studies lack consideration of electricity consumption and transportation of the materials. Inclusion of these aspects would give a clearer aspect on the sustainability performance of Concrete 3D printing process.

In flexural members like slabs and beams concrete tension zone has only one function to connect the concrete in the compression zone to the rebars in the tension zone, thereby facilitating shear transfer. This provides an opportunity to reduce the consumption of concrete in this region. Filler slabs reduce concrete usage in the tension zone by incorporating elements like clay pots and tiles between rebars, decreasing concrete consumption by 25-30%. While this lowers carbon emissions, the benefits are offset by the significant emissions from manufacturing these fillers [12]. Additionally, filler placement is time-intensive and requires meticulous supervision, inadequate oversight or poor workmanship can cause reinforcement bars to contact fillers, increasing the risk of corrosion. The world produces around 2.01 billion tons of solid waste, with 33% poorly managed and 11% incinerated, resulting in significant carbon emissions [13]. Waste management contributes 5% of global GHG emissions [14]. Using mechanically compressed plastic waste as fillers can mitigate emissions from clay pot and tile production while reducing those from incineration and mismanagement of polyethylene waste.

Most reported applications of concrete 3D printing

focus on architectural elements or compression members [15]. To maximize its benefits, efforts are now directed toward printing structural elements capable of resisting tension and flexure [16]. While concrete 3D printing can reduce labour and enhance productivity, reliance on manual tasks such as reinforcement placement and curing may undermine efficiency and compromise the benefits of automation. As structural slab casting is labour-intensive process, concrete 3D printing integrated with automation for ancillary works can help reduce the requirement for manual labour and improve efficiency of the process.

This study proposes a methodology to leverage the benefits of concrete 3D printing, automation and filler slabs to reduce the carbon footprint of the RCC slab. A gantry-type 3D printer integrated with a robotic gripper is proposed to automate the precise placement of filler elements and reinforcement cages and casting of filler slabs. The system's environmental performance is assessed through cradle-to-gate life cycle carbon emission assessment of a case study comparing embodied carbon emissions for different scenarios of a 6 m × 4 m RCC slab. Mechanically compressed polyethylene (PE) waste is proposed as a filler material to overcome the limitations of clay pots.

2 Materials and Methods

2.1 Slab Design & Mix Design

As an illustrative case study, an RCC slab of width 4 m and length 6 m is designed. The design follows the Indian Standard Code for Plain and Reinforced Concrete [14]. Assumptions made for the design are mentioned in Table 1.

Table 1 Assumptions for the design of RCC Slab

Design Parameters	Unit	Value
Grade of Concrete	N/mm^2	35
Reinforcement Grade	N/mm^2	415
Live Load	kN/m^2	2
Floor Finish	kN/m^2	1.5
Exposure Conditions	N/A	Mild
Support Condition	N/A	Simply Supported

The design summary is presented in Table 2.

Table 2 Design Summary

Design Parameters	Unit	Value
Thickness of the Slab	mm	250
Rebar Diameter	mm	10
Main Rebar Spacing	mm c/c	150
Distribution Rebar Spacing	mm c/c	180
Cover Provided at the Bottom of the Slab	mm	20
Neutral Axis Location considered from the bottom of the slab	mm	130

The mix design [17] and the material usage detail are presented in Table 3 and Table 4 respectively.

Table 3 Mix proportion for OPC + 30% LC2 concrete design.

Material	Quantity (kg/m^3)
Quartz Sand	1206
Cement	575
LC2 (Limestone Calcined Clay)	246
Water	288

Table 4 Detail of material usage in conventional and filler slab

Material	Unit	Filler	Conventional
Total Volume	m^3	4.37	6
Printable Concrete (tension zone and boundary of compression zone)	m^3	1.63	3.26
Conventional M35 Concrete (used in compression zone)	m^3	2.74	2.74
Volume of Voids	m^3	1.64	-

2.2 Proposed Concrete 3D Printing Setup

A gantry 3D printer with a robotic gripper is proposed for slab printing, comprising three main components: a

printing head with dual nozzles a pneumatic gripper for pick-and-place operations, and a motion control system. Concrete is pumped from the mixer to the printing head, where an auger mechanism ensures controlled extrusion. The dual nozzles handle 3D printable mix and regular concrete, respectively. The gantry operates in the x-y-z plane, facilitating rebar cage and filler element placement, with a workspace of 8 m × 8 m in the x-y plane and 1 m in the z direction. Workspace and the 3D
printer is shown in Figure 1 and Figure 2 respectively. An early prototype of the design integrating 3D printing with robotic pick and place features is shown in Figure 3.

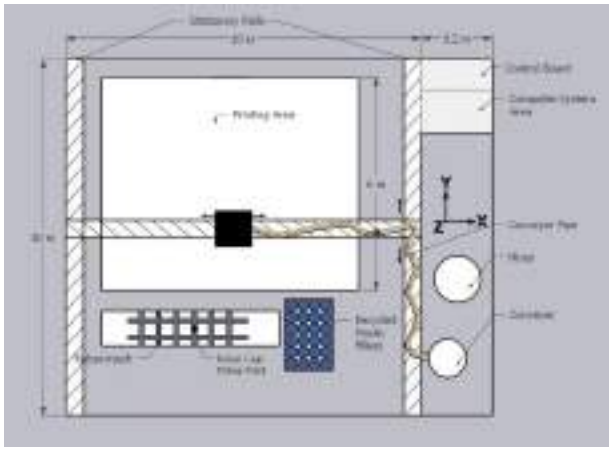


Figure 1 Plan view of the 3D printing system integrated with automated pick and place function.

The process flow is illustrated in Figure 4.

2.3 Filler Slab Printing Process

The slab is assumed to be printed as 6 panels, each 5m long and 1m wide. The printing has been carried out in three parts: cover layers, compression zone, and tension zone printing Figure 5.

The filler and rebar placement settings are shown in Figure 6. “2w” and “2b” are the main and distribution bar spacings, “G+a” indicates the distance between filler pickup point and the first void. The movement in the z-direction is assumed to be 100 mm per pick and place operation.

2.4 Life Cycle Assessment

The LCA used in the study is based on ISO 14040/44 [18, 19] guidelines. The consequential LCA (CLCA) approach is used to show the change in carbon emissions by adopting the proposed concept.

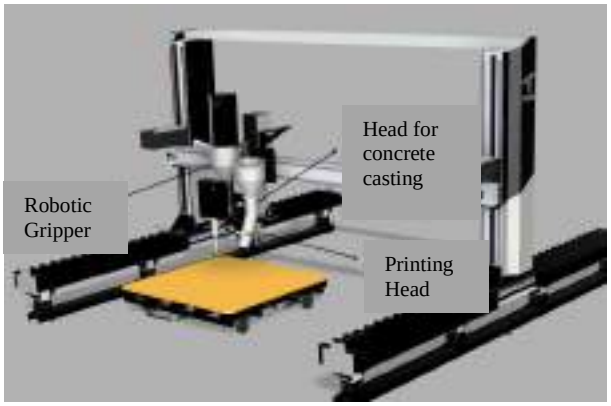


Figure 2 Plan view of the 3D printing system integrated with automated pick and place function.



Figure 3 An early prototype of the design integrating 3D printing with robotic pick and place features

2.4.1 Goal, Scope and System Boundary

This study aims to quantify the reduction in carbon emissions by adapting the proposed method of printing filler RCC slabs using recycled PE as fillers, compared to conventional slab printing. The results provide new insights for the industry and research community, promoting the integration of robotics, concrete 3D printing, and waste recycling to develop sustainable construction alternatives. The recyclable waste used is polyethylene (LDPE or HDPE), the most consumed plastic, with a gantry-type concrete 3D printer integrated with a gripper for pick-and-place operations employed in the process.

The functional unit in this study is the concrete 3D printing 1 m^3 volume of a 250 mm thick structural slab spanning 6 m X 4 m to allow comparisons between different printing scenarios.

Case-1: 3D printing of 6 panels of 4 m length and 1m width for a 6 m X 4 m RCC slab without any fillers.

Case-2: 3D printing of 6 panels of 4 m length and 1m width for a 6 m X 4 m RCC slab with clay pots as fillers.

Case-3: 3D printing of 6 panels of 4 m length and 1m width for a 6 m X 4 m RCC slab with Recycled PE waste as fillers.

Case-1 is considered the base case for the comparing the carbon emissions. The system boundary used for conducting LCA in this study is shown in Figure 7.
.

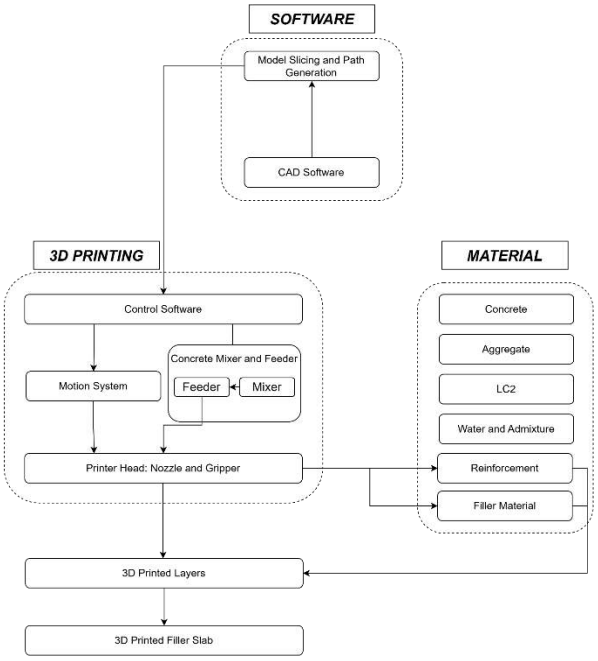


Figure 4 Process flow for 3D printing of filler slab

2.4.2 Life Cycle Inventory

Concrete production emissions are based on design mix [17] and carbon emission study of concrete printed mix [9]. Plastic waste recycling involves motors for washing, drying, and compressing PE waste, with emissions calculated from electricity use (Ecoinvent v3.9), the power calculation for washing and drying has been done based on study by [20], a 5 HP motor is assumed to compress 1 ton of waste in 10 seconds.

Clay pot production emissions is taken from [12] market data is used for volume to weight calculation, assuming 1 litre. pot weighs 1kg. Transportation emissions are calculated for a 10 km local distance using a 3-ton vehicle with 6 km/L mileage and Ecoinvent v3.9.

3D printing emissions account for electricity consumption, while avoided emissions are based on the incineration of PE waste which is calculated using data from Ecoinvent v3.9. Only 11% of the waste used as filler is assumed to be incinerated.

This study employs CML v4.8 2016 to assess life cycle carbon emissions, focusing on the reduction of cement and concrete content and related carbon emissions, as required by the baseline impact category.

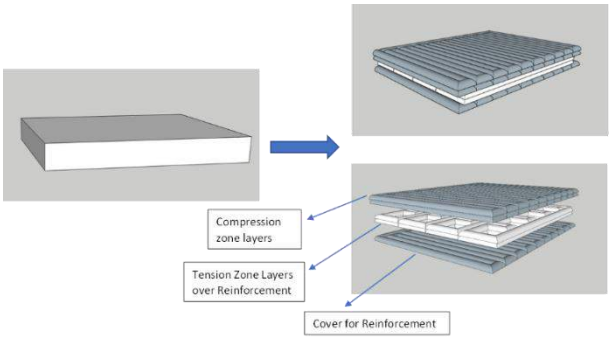


Figure 5 Representation of a Conventional Slab (Left). 3D printed slab (Right Top). Different layers of 3D printed slab (Right Bottom)

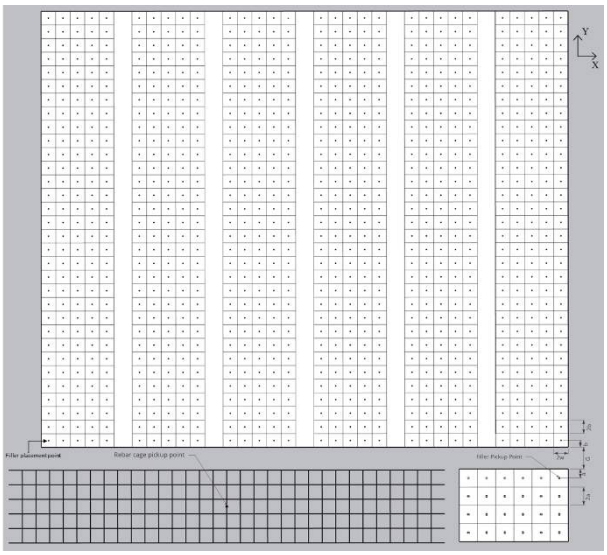


Figure 6 Plan for placement of fillers and rebar cage

3 Results and Discussions

3.1 Concrete 3D Printed Slab Output

The printing path and the placement of the filler are shown in Figure 8. The size of filler material for one void is 140 mm by 110 mm cross-section and 110 mm deep. Figure 9 shows a cross-section view of the 3D printed elements for the case of conventional printing and printing with fillers.

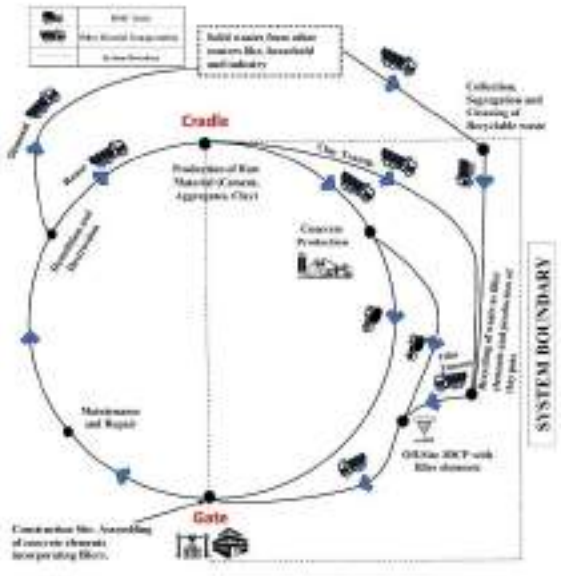


Figure 7 Schematic for life cycle assessment of concrete filler slab

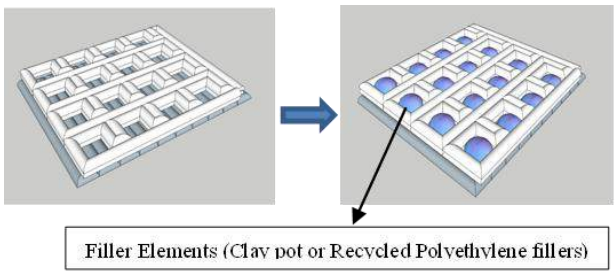


Figure 8 Printing path over rebars in tension zone (left), placement of fillers in voids (right)

3.2 Carbon Emission Assessment

3.2.1 Filler Production

As per the assumptions stated in section 2.4.2 the carbon emission assessment for clay and recycled PE filler production is mentioned in Table 5 and Table 6.

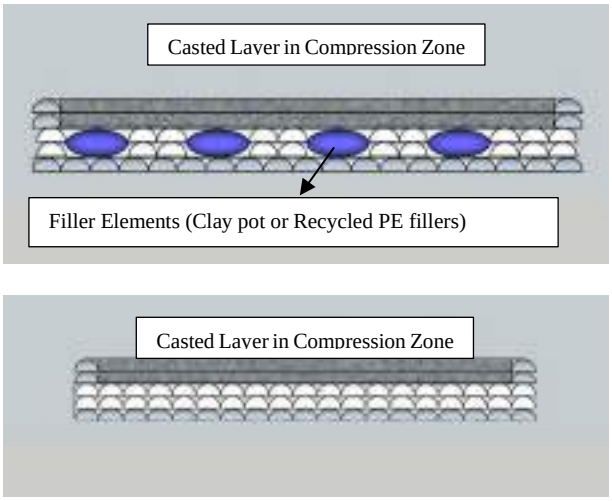


Figure 9 (a) Cross Sectional View of Slab in Scenario Case-1 and Case-3 (b) Cross Sectional View of Slab in Scenario Case-1

Table 5 Emissions generated due to the production of clay pot fillers.

Description	Unit	Value
Weight of the Clay Pots Required	kg	1636.4
Carbon emissions per kg of unit produced	$kgCO_2$ eq.	1.91
Total Carbon emissions for clay pots required	$kgCO_2$ eq.	3125.53

Table 6 Emissions generated due to the production of recycled plastic fillers.

Process	$kgCO_2$ eq. emissions
Transportation from scrap dealer to MRF	12.20
Washing of waste PE	4.07
Transportation to the 3D printing facility	12.20
Compaction using Automatic Baler	16.24
TOTAL	44.70

Using recycled polyethylene fillers reduces emissions by 98%. Additionally, the weight of the recycled polyethylene filler slab (total weight of PE fillers used is 2.08 tonnes) is 6% lower than the clay pot filler slab and 60% lower than the conventional slab, reducing the slab's dead load. This reduction helps minimize material consumption in the substructure.

3.2.2 Electricity Consumption in Printing the slab

The printer head operates at an assumed speed of 80 mm/s in both the x and y directions. For movement in the z-direction, rather than calculating the total distance, the number of layers is considered, with each layer transition taking approximately 1 second. The gantry system employs 1 kW, three-phase motors for the x, y, and z axes. The electricity consumption is determined by multiplying the power rating of the motors by the total operational time. The pump used for extruding concrete has a power rating of 5.5 kW.

The total electricity consumption for printing the filler slabs is presented in Table 7.

Table 7 Total Electricity Consumption for printing slab

Description	Filler Slab	Conventional Slab
Printing Process	36.62	57.48
Filler Placement	37.30	-
Rebar cage placement	0.28	0.28
Pumping Concrete	183.08	287.40
Total Electricity Consumption in kWh	257.28	345.16

3.2.3 Embodied Life Cycle Carbon Emission

Two concrete mixes were used to produce printed slabs. An M35 mix [21] with 30% Class F Fly Ash replacement, emitting 289 kg CO_2/m^3, was used for casting the compression zone, while an OPC-LC2 mix [17] was used for printing the cover, tension and boundary zones, with emissions data sourced from [9]. CO_2 emissions from plastic waste incineration and electricity consumption were derived from Ecoinvent v3.9 [22]. The emission assessment results are summarized in Table 8.

The table reveals that using recycled plastic waste as filler elements reduces carbon emissions by 37%. Conversely, clay pot filler slabs increase emissions by 55% due to the high carbon footprint of clay pot production (3125 kg CO_2 eq.), despite reduced material usage. Case 3 demonstrates the lowest emissions, attributed to avoided emissions from repurposing waste plastic. This highlights that filler material selection is crucial, as the production emissions can offset the benefits of reduced raw material consumption, as seen in Case 2.

Table 8 Life cycle carbon emission comparison for the three scenarios

Description	Case-1 (kgCO2 eq.)	Case-2 (kgCO2 eq.)	Case-3 (kgCO2 eq.)
Concrete Production for compression zone (M 35-35% fly ash replacement)	790.76	790.76	790.76
Concrete Production for tension zone (OPC+30% LC2)	2153.5	694.74	694.74
Production of Filler Element	-	4221.5	49.7
Electricity Consumption in 3D Printing	531.54	396.21	396.21
Avoided Emissions	-	-	-120.48
NET EMISSIONS	3475.78	5385.95	2189.64
NET EMISSIONS PER m^3	579.29	897.65	364.94

4 Conclusion

Concrete 3D printing has significant potential to transform the construction industry, but its full realization is hindered by challenges such as limited automation, minimal application in printing structural elements like slabs and beams, and the high cement content of 3D printable mixes. Hence, developing techniques to reduce raw material consumption is essential to minimizing the carbon emissions associated with production of concrete 3D printed elements. This study proposes an automated methodology for 3D printing RCC filler slabs using compressed polyethylene (PE) waste as an alternative to clay pots. A case study of

a 250 mm thick slab spanning 6 x 4 m is analyzed to evaluate the carbon footprint of 3D printing a RCC slab.

Following are the major findings of the case study:

- The findings suggest that the proposed 3D printing methodology can effectively reduce material usage and carbon emissions, offering a cleaner and more sustainable construction alternative. In the case study, utilizing polyethylene waste as a filler material decreased concrete consumption by 27% and carbon emissions by 37% compared to the baseline scenario of 3D printing slabs without filler elements.
- The lower carbon emissions in filler slabs with recycled PE waste, compared to those with clay fillers, are primarily due to the significantly reduced emissions from producing recycled PE fillers, which are 98% lower.
- Also, Recycling PE waste into slab fillers prevents the emissions that would have resulted from the incineration of plastic if the waste had not been recycled.
- Slabs with recycled PE waste as fillers have a lower dead load compared to those with clay pot fillers or no fillers, further contributing to material and carbon emission savings in substructure construction.

The study has assumed various transportation distances which may not be the real scenario. The equipment production related emissions were ignored in the calculation of carbon footprints because the equipment being used in all the three scenarios mentioned is same. Also, a cradle to gate approach for carbon emission assessment was used, because the main objective of the study was to compare the scenario in production phase. In future other phases like operational and demolition phases can also be studied. Also, the study demonstrates the potential using a single slab case, in future the results will be generalized using different cases and structures.

Nevertheless, these findings offer a path toward utilizing the advantages of concrete 3D printing and automation for sustainable construction. Automating tasks such as placing rebars and fillers reduces the risk of human error, addressing one of the key drawbacks of the conventional filler slab technique.

Acknowledgement

This project is funded by the Science and Engineering Research Board (SERB), India, through the Grant IMP/2018/000224. The Institute of Eminence Research Initiative Project on Technologies for Low-Carbon Lean-Construction (TLC2) and SBI Foundation also provided financial support for this research. Tvasta Manufacturing Solutions Pvt. Ltd., India is thanked by the authors for their support in providing 3D printers for the study.

References

[1] Bazli, M., Ashrafi, H., Rajabipour, A., Kutay, C.: 3D printing for remote housing: Benefits and challenges. Autom Constr. 148, 104772 (2023). https://doi.org/10.1016/j.autcon.2023.104772

[2] Kim, S., Chang, S., Castro-Lacouture, D.: Dynamic Modeling for Analyzing Impacts of Skilled Labor Shortage on Construction Project Management. Journal of Management in Engineering. 36, (2020). https://doi.org/10.1061/(ASCE)ME.1943-5479.0000720

[3] Weng, Y., Li, M., Ruan, S., Wong, T.N., Tan, M.J., Ow Yeong, K.L., Qian, S.: Comparative economic, environmental and productivity assessment of a concrete bathroom unit fabricated through 3D printing and a precast approach. J Clean Prod. 261, 121245 (2020). https://doi.org/10.1016/j.jclepro.2020.121245

[4] Asprone, D., Auricchio, F., Menna, C., Mercuri, V.: 3D printing of reinforced concrete elements: Technology and design approach. Constr Build Mater. 165, 218–231 (2018). https://doi.org/10.1016/j.conbuildmat.2018.01.018

[5] Martens, P., Mathot, M., Bos, F., Coenders, J.: Optimising 3D Printed Concrete Structures Using Topology Optimisation. In: High Tech Concrete: Where Technology and Engineering Meet. pp. 301–309. Springer International Publishing, Cham (2018)

[6] Miller, S.A., Horvath, A., Monteiro, P.J.M.: Readily implementable techniques can cut annual CO 2 emissions from the production of concrete by over 20%. Environmental Research Letters. 11, 074029 (2016). https://doi.org/10.1088/1748-9326/11/7/074029

[7] Gagg, C.R.: Cement and concrete as an engineering material: An historic appraisal

and case study analysis. Eng Fail Anal. 40, 114–140 (2014). https://doi.org/10.1016/j.engfailanal.2014.02.004

[8] Shanks, W., Dunant, C.F., Drewniok, M.P., Lupton, R.C., Serrenho, A., Allwood, J.M.: How much cement can we do without? Lessons from cement material flows in the UK. Resour Conserv Recycl. 141, 441–454 (2019). https://doi.org/10.1016/j.resconrec.2018.11.002

[9] Bhattacherjee, S., Basavaraj, A.S., Rahul, A.V., Santhanam, M., Gettu, R., Panda, B., Schlangen, E., Chen, Y., Copuroglu, O., Ma, G., Wang, L., Basit Beigh, M.A., Mechtcherine, V.: Sustainable materials for 3D concrete printing. Cem Concr Compos. 122, 104156 (2021). https://doi.org/10.1016/j.cemconcomp.2021.104156

[10] Mohammad, M., Masad, E., Al-Ghamdi, S.G.: 3D Concrete Printing Sustainability: A Comparative Life Cycle Assessment of Four Construction Method Scenarios. Buildings. 10, 245 (2020). https://doi.org/10.3390/buildings10120245

[11] Batikha, M., Jotangia, R., Baaj, M.Y., Mousleh, I.: 3D concrete printing for sustainable and economical construction: A comparative study. Autom Constr. 134, 104087 (2022). https://doi.org/10.1016/j.autcon.2021.104087

[12] Li, Y., Sharaai, A.H., Ma, S., Wafa, W., He, Z., Ghani, L.A.: Quantification of Carbon Emission and Solid Waste from Pottery Production by Using Life-Cycle Assessment (LCA) Method in Yunnan, China. Processes. 10, 926 (2022). https://doi.org/10.3390/pr10050926

[13] World Bank: WHAT A WASTE 2.0: A Global Snapshot of Solid Waste Management to 2050, https://datatopics.worldbank.org/what-a-waste/, (2018)

[14] Gautam, M., Agrawal, M.: Greenhouse Gas Emissions from Municipal Solid Waste Management: A Review of Global Scenario. Presented at the (2021)

[15] Raphael, B., Senthilnathan, S., Patel, A., Bhat, S.: A review of concrete 3D printed structural members. Front Built Environ. 8, (2023). https://doi.org/10.3389/fbuil.2022.1034020

[16] Cohen, Z., Carlson, N.: Piling and pressing: towards a method of 3D printing reinforced concrete columns. Construction Robotics. 4, 61–73 (2020). https://doi.org/10.1007/s41693-020-00029-6

[17] Rahul, A.V., Sharma, A., Santhanam, M.: A desorptivity-based approach for the assessment of phase separation during extrusion of cementitious materials. Cem Concr Compos. 108, 103546 (2020). https://doi.org/10.1016/j.cemconcomp.2020.103546

[18] ISO: ISO 14044 environmental management — life cycle assessment — requirements and guidelines. (2006)

[19] ISO: ISO 14040: Environmental Management-Life Cycle Assessment-Principles and Framework. (2006)

[20] Aryan, Y., Yadav, P., Samadder, S.R.: Life Cycle Assessment of the existing and proposed plastic waste management options in India: A case study. J Clean Prod. 211, 1268–1283 (2019). https://doi.org/10.1016/j.jclepro.2018.11.236

[21] Pillai, R.G., Gettu, R., Santhanam, M., Rengaraju, S., Dhandapani, Y., Rathnarajan, S., Basavaraj, A.S.: Service life and life cycle assessment of reinforced concrete systems with limestone calcined clay cement (LC3). Cem Concr Res. 118, 111–119 (2019). https://doi.org/10.1016/j.cemconres.2018.11.019

[22] Yadav, P., Bharati, D.K., Gautam, R.K.: Determination of water quality index and suitability of water bodies in Robertsganj, Sonbhadra district, Uttar Pradesh, India. Asian J Res Soc Sci Humanit. 7, 1 (2017). https://doi.org/10.5958/2249-7315.2017.00360.4

Design for Circular Disassembly of High-Performance Facade Systems

Naomi Grigoryan[1] and Alexandros Loutsioli Daskalakis[1]

[1] Schüco International, Virtual Construction Lab, United States
ngrigoryan@schuco-usa.com, aloutsioli@schuco-usa.com

Abstract –

The construction sector faces significant challenges in reducing emissions, with extending the lifespans of architectural products emerging as a key strategy. As stakeholders of the built environment focus on decarbonization, the concentration of efforts diverges towards either sustainable development of future products or sustainable preservation of the existing built environment. Critical intersections between development and preservation are the reuse, repurpose, repair, and refurbishment stages within a circular economy. This research investigates the integration of Design for Disassembly (DfD) principles into the lifecycle of high-performance facade systems, as complex architectural systems frequently utilize materials and methods that hinder circularity, such as fixed connections. Designing and advocating for future uses is challenging without explicit disassembly standards or supporting infrastructure. By analysing existing disassembly methodologies, a tailored approach was developed for high-performance façade design, leveraging Schüco KG's existing digital infrastructure to optimize disassembly workflows. This methodology focuses on the generation of practical disassembly outputs, for use cases across designing, manufacturing, and constructing façade systems. Coordinating with fabricators, a proof-of-concept for a Disassembly Application is developed to integrate with existing digital infrastructure. The case study on the AWS 75.SI+ window system highlights the feasibility and benefits of incorporating disassembly planning from the early design stage. This framework provides a scalable model for reducing construction waste, improving material reuse, and supporting circular economy goals for high-performance façade systems and across complex building systems.

Keywords –

Design for Disassembly; Life Cycle Analysis; Built Environment; Building Façade System; Case Study

1 Introduction

There is a substantial amount of embodied energy in existing buildings. Demolishing obsolete buildings in favor of higher-performing new construction is a popular path, but the energy cost of demolition and disposal must be addressed. Standard Architecture, Engineering, and Construction (AEC) education and practice center buildings with lifecycles of 30-50 years, programmed for demolition rather than renovation [1]. Frequent technological advancements increase the rate of technical obsolescence of façade systems. Instead of considering future technological upgrades as a design criterion, many façade systems are replaced in their entirety in favor of higher-performing systems. High-performance façade systems often contain aluminum and other critical materials in the Critical Raw Materials Act (CRMA) [2]. Smart façade systems with integrated technology contain critical raw materials such as cobalt, coltan, and lithium. The technology sector and transportation sector account for critical material recovery, but the building sector does not have a comprehensive recovery strategy. Demolition involves all materials going to landfill, with no planned recovery. Deconstruction, however, is the systematic recovery of materials with circular intention. Disassembly refers to the capacity for deconstruction in a system, meaning systems intended for deconstruction must be designed with easy part disassembly [3]. Design for Disassembly (DfD) is a research field and design practice developed within the manufacturing and construction industries, geared towards expanding material circularity. DfD considers the various parameters including material composition, component connection, data accessibility, material flow, stakeholder incentives, and local options. Developing infrastructure to facilitate DfD across construction would allow valuable materials to re-enter the Material Bank [4] and outline opportunities for future designs, reducing both resource extraction and construction waste.

Disassembly for architecture has decades of existing research and methods, with many digital solutions and case studies; however, there is a gap in disassembly

research pertaining to high-performance facade systems and associated components and processes. While façade systems often use critical and high-value materials such as aluminum and steel, high-performance systems rely on polymers and adhesives, that reach technical obsolescence sooner, to meet structural, thermal, and acoustic criteria. Although high thermal performance significantly lowers operational carbon, current practice uses fixed connections, permanently sealing non-circular materials to systems. Thus, although facade assemblies are historically among the most demountable, reconfigurable, and reusable building systems, the only End-of-Life options for these high-performance systems are recycling, incineration, and landfill.

This research begins with a brief review of published disassembly methodologies, identifying applicable modeling for high-performance facades. A methodology is developed to integrate into existing digital design and manufacturing infrastructure. The product catalog provides a wide range of façade systems with varying degrees of performance and disassembly. Various electronically operated systems assist with ventilation, solar shading, security, operability, and comfort. As with other technology devices, these smart components contain cobalt, lithium, and coltan. The documentation of disassembly data serves three main purposes: it improves the efficiency and effectiveness of deinstallation during fabrication, construction, maintenance, and at End-of-Life; it improves circularity and circularity documentation; and lastly, it assists with product design, improvement, and development related to product circularity. Next, the methodology and case study portions follow Schüco system AWS 75.SI+ to develop a disassembly application. The research aims to bridge existing construction disassembly theories and studies with AEC workflows.

2 Review of Published Methodologies

For this research, multiple published disassembly and work estimation methodologies were reviewed and assessed for applicability. The review included Durmisevic's Knowledge Model [5], BIM-DAS [6], Selective Disassembly Planning for Buildings (SDPB) [7], Disassembly Network Analysis (DNA) [8], Method Time Measurement (MTM) [9], Maynard Operation Sequence Technique (MOST®) [10], Adaptive Reuse Model (ARP) [11], AdaptSTAR [12], and the Parent Action Child (PAC) model [13]. With considerations of users, inputs, outputs, and level of automation, four main methodologies were selected. Durmisevic's Knowledge Model is one of the older and most referenced disassembly methodologies. Outputs include disassembly sequences, connection-type knowledge, and an independence and exchangeability aggregated score. While it considers small-scale connection information applicable for facade systems, the manual process cannot be easily integrated into existing workflows [5]. The BIM-DAS method is a Building Information Modeling (BIM) based Deconstructability Assessment Score, developed to assess and minimize construction waste from the design stage. Outputs include the Deconstruction Score, the Reuse Score, and finally, the Deconstructability Assessment Score (0% being the lowest, 100% being the highest). The scores are not appropriately weighted, and the focus is on larger scale building connections than facade systems [6]. Consideration of component end-of-life conditions is necessary for disassembly effort assessment. The PAC model divides products into assemblies and subassemblies, further categorized into parents, actions, and children. Considering disassembly failures (DFs), the PAC model outcomes include the Disassembly Effort Index (DEI) and the circularity index (CI). The DEI represents the time required to perform an action and disassemble a component, and the CI represents the circularity of the components after disassembly [13]. Disassembly failures are categorized into three types: Type One occurs during product use and are related to the children in the PAC model, Type Two is associated with actions in the PAC model, and Type Three affects the parents, meaning both the action and one or more children are impacted. MOST® is a predefined motion time system used in industrial settings to establish the standard time required for a worker to perform a task. Tasks are broken down into individual motion elements and assigned Time Measurement Units (TMUs), where 100,000 TMUs equal one hour [10]. This information can be integrated into the PAC model to assess the DEI, providing a comprehensive understanding of the disassembly process and identifying areas for efficiency improvements [13].

3 Methodology

The existing disassembly methodologies described in the methodology review portion were assessed for criteria related to facade system product lifecycles: intended user groups, required inputs, outputs, and transparency. Existing methodologies were also selected based on the availability of case studies. A methodology was defined by analysing the physical components and life cycles of facade products, as well as the existing product documentation and company-wide digital infrastructure. Existing information necessary to disassembly was mapped, and remaining information was identified.

Building Façade Systems (BFS) are composed of elements and components. Façade elements are the functional assemblies: external walls, doors and windows.

Façade systems are combinations of multiple elements. Components are the parts that make up the elements, such as the frame, handle, or reinforcement profile. The functions of components are structural support, thermal and acoustic control, aesthetic and resilient finish, and integration to other components through connection [3]. The disassembly of façade systems is dependent upon the separation and removal of components for replacement and reuse. Components can serve multiple functions, which constitutes a closed façade system. If the functions of components are separate, they are considered functionally independent. A component can be removed from the element if it is functionally independent, and a system comprised of functionally independent components is considered an open façade system [5].

As shown in Figure 1, there is currently no selective disassembly and recovery integrated into the standard facade system product life cycle. All façade building materials are either sent to recycling or landfill. Highlighted in green is the "closed loop" of material circularity, stages A3 - D. Life cycle stages aligned according to ISO 14040 [14].



Figure 1. The typical product life cycle of a facade system product.

Figure 2 shows how building materials gain the opportunity to be diverted from the two previous End-of-Life pathways, to be refurbished and reused in new products without shredding. Existing products have many reusable components that do not need the additional energy use of recycling. Additionally, this analysis can be used by product designers to develop future products that have higher disassembly potential.



Figure 2. Integrating Design for Disassembly (DfD) into the facade system product life cycle.

Product ordering, fabrication, and assembly are often well-documented across facade systems. However, there is a large gap in documentation of deconstruction and deinstallation of products. From lifecycle stages A1-A5, extensive product information is accessible through internal digital infrastructure, BIM tools, and product documentation such as EPDs and BOMs. This methodology limits the necessary input from fabricators, manufacturers, and designers by utilizing available data. Fabrication processes vary, even across facilities at the same company. Necessary inputs for disassembly calculation are thus split into two lists: existing company data and data to collect from users (see Table 1). User input is necessary to map a disassembly process to a product system once. Then, it can be stored for future applications using the same products.

Table 1. Data Sourcing

Existing Internal Data	User-Input Needed
System Type	Connection Type
Element Type	Disassembly Sequence
Component Type	Disassembly Action
Material	Disassembly Equipment
Weight	Component Recoverability
Assembly Action	

One of the key outputs of the application is the breakdown of materials based on their End-of-Life potential. All components in a BOM are sorted into three categories: reusable, recyclable (but not reusable), and waste. This breakdown is shown in item count as well as item weight. An additional valuable materials End-of-Life breakdown documents materials that must be recovered in the material flow; ideally none which go to landfill/incineration. Unaltered and slightly altered reusable components can be salvaged and reused with minimal alteration. For example, un-crimped cleats cannot be re-crimped but can be screwed on a second configuration. The list of unrecoverable elements guides product designers to identify weak points of design in relation to disassembly. The disassembly sequence output is the written list of directions to dismantle the configuration. First, the objects in the BOM are sorted into element groups, such as an operable vent, fixed light, door, or framing. Within each element group, each item is placed in descending order based on Sequence ID. Lastly, all identical items within each field group are

grouped to condense the disassembly sequence. In the case of complex façade systems, this is necessary to avoid tediously lengthy lists that spell out the directions for every individual screw and other small components, where there may be dozens, if not hundreds. The disassembly sequence thus lists the number of components in each element to dismantle for each step, with the time estimate. For components that cannot be disassembled, that is reflected instead of an instruction and time estimate. The total number of steps and the total time estimate are provided along with the sequence. An output for sustainability experts includes the Material Circularity Indicator (MCI), developed by the Ellen McArthur Foundation; quantifies the minimization of linear material flow and maximization of restorative flow for all component materials in a product [15]. A score of 100% implies total repurposing upon completing a lifecycle. Typically, the MCI of a facade product is calculated using an EPD generated for life cycle stages A1-A3; in this methodology, the MCI is calculated considering the realistic end-of-life of the product, including potential disassembly failures. Similarly, The Water Circularity Indicator (WCI) is measured as the ratio, in cubic meters, of circularly sourced water over the total amount of water used. The Energy Circularity Indicator (ECI) is measured as the ratio of renewable energy used over the total energy used [16]. An output geared towards product designers involves a summary of all limitations to disassembly identified in the product. This includes conditions such as rolled thermal breaks, permanently sealed joints, structural sealant, and petroleum-derived materials. This section explains the reasons these decisions are typically made and provides suggestions for improvement. In the Case Study section, this methodology is implemented on a specific façade product, Schüco's AWS 75.SI+.

4 Integration and Operability

Building Information Modeling (BIM) facilitates collaboration across all project stakeholders. Disassembly information affects design, engineering, LCA, and construction means and methods; therefore, it should be accessible to all stakeholders. Single platform hosting all performance metrics allows for easy comparison and assessment, weighing the trade-offs of disassembly capacity with structural, thermal, or acoustic performance and cost. As the disassembly information may change across the supply chain, having a single source of truth for reference is essential.

Users of a disassembly application must be able to add all related data points for their model. Product designers can initially specify the BIM content of an architectural component, documenting this initial disassembly data in a standardized "Information Delivery Specification" (IDS) format. Developed by buildingSMART, IDS is a standard for specifying alphanumeric information following the Industry Foundation Classes (IFC) schema [17]. This format is appropriate for applying the outlined disassembly specifications and can be supplemented with additional specifications, such as recommended assessment criteria in ISO 20887(EN) [18]. This documentation can then be shared with engineers for simulation and calculation, shared with sales and other contributors, and collaboratively finalized in an interoperable format such as IFC. For the disassembly information to be open source and publicly accessible, this IFC compliance is critical [19].

Once the product model is complete, all complex system logic and disassembly documentation are available for the project architects, engineers, and contractors. These models can be used in all IFC-compliant BIM tools, such as Autodesk Revit® and Tekla Structures®. In addition, disassembly outputs can be generated and accessed in addition to existing outputs within the BIM tools. As the project develops, these stakeholders can also update the architectural component. For disassembly, this is necessary when, for example, a component is designed as fully deconstructable, but the structural engineer makes a call to use a permanent connection for a site-specific wind load; this disassembly information would be updated in the BIM model to reflect the change for the deconstruction team's disassembly instructions.

5 Case Study

The case study implementing the methodology outlined in this research uses Schüco's AWS 75.SI+ window system (see Figure 3), which shares many manufacturing and fabrication processes with similar window and facade systems, within Schüco and across the facade industry. The disassembly planning was mapped for an inoperable fixed light as well as for an operable turn-tilt window, both with the standard dimensions of 1480 x 1230 mm. A simplified overview of Schüco's workflow begins with product designers and engineers creating profiles and systems; next, manufacturers extrude, cut, and assemble composite profiles; next, fabricators assemble configurations based on project specifications. Many locations provide maintenance services during the products' use phase, and several locations and product types provide recovery services for materials after use. All Schüco profiles and accessories have unique identification numbers that contain profile data, including material, weight, price, description, and more, and are accessible through various internal portals. The digital infrastructure contains documentation platforms, ordering, planning, modeling,

and calculation software, and a material passport system. Although the profile data is comprehensive for many applications, disassembly data is not included. The case study interfaces with configuration tool SchüCal, which is utilized across the value chain. Order, fabrication, and physics data can be exported, including BOMs and EPDs. These two exports are mapped to additional disassembly data, used to derive disassembly metrics in a disassembly application.

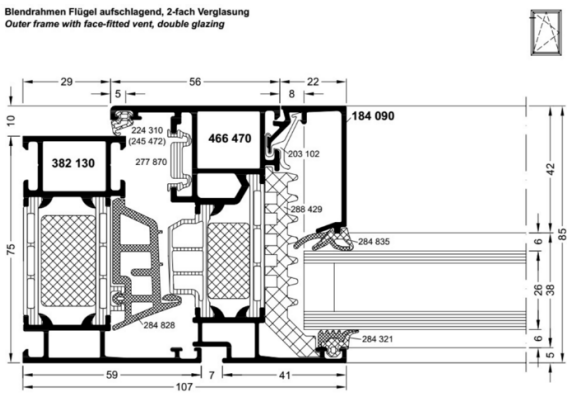


Figure 3. AWS 75.SI+ outer frame with face-fitted vent

The system contains aluminum profiles that are thermally broken with polyamide (PA) or proprietary material Polythermid (PT) isolators, which are permanently rolled into the aluminum profiles with insulation in between each thermal break component. Gaskets are made of EPDM and PE foam is used for insulation. Corners and profile joints are fixed with aluminum cleats and sealed with epoxy in addition to steel screws. Locking bars are a combination of plastic and metal, and most fittings are metal and fully removable.

5.1 Network Analysis

Integrating disassembly to each system type starts with system network analysis. Following the process mapped in Durmisevic's Knowledge Model, all system components are plotted as nodes and all components and connections are labeled. This mapping highlights the hierarchies and dependencies and the order in which they can safely be separated from the total configuration. Network analysis also highlights which components are inaccessible for removal due to sequencing and permanent connections. This process establishes a disassembly sequence for each system type and provides the necessary framework for disassembly metrics via a component database. This mapping requires coordination with fabricators familiar with the products' assembly processes. Fabricators are familiar with the physical limitations of product components and the adjustments necessary during imperfect assembly, knowledge that is not documented in Schüco internal resources, making disassembly mapping, product design, and onboarding incoming fabricators more challenging. For the case study of Schüco's AWS 75.SI+ window system, there was coordination with fabricators. AWS 75.SI+ is a high-performance window system with a depth of 75 mm. Increased thermal insulation for efficiency and performance is included in the center gasket and glazing rebate insulation in addition to standard gaskets and insulation foams [20].

In Figure 4, the left image shows an exploded axonometric view of a corner condition of an AWS 75.SI+ frame, and the top right image shows the frame cross-section. The bottom right diagram shows the network of these components.

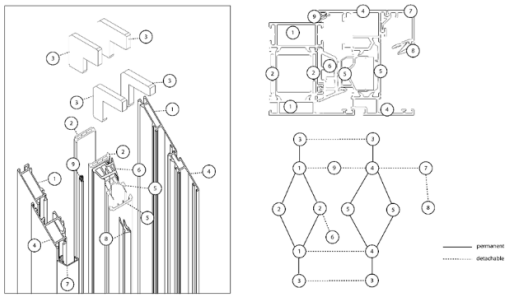


Figure 4. AWS 75.SI+ Network Analysis

Maintaining the connection network, Figure 5 details all names, materials, and article numbers. Connections are distinguished between permanent, detachable, and easily detachable, with connection-type descriptions, and arranged from left to right in order of assembly steps. The aluminum profiles are first rolled with thermal breaks in a permanent connection. After cutting and processing the profiles with end cuts, holes, notches, and slots, the corner cleats are nailed, screwed, or crimped into position, and then sealed with epoxy to secure mitered frame corners. In coordination with fabricators, these two permanent connections were highlighted as assembly challenges, as mistakes during these steps cannot be reversed. Finally, glazing and gaskets are snapped into place, starting with the bottom glazing gasket, after which the aluminum glazing bead is snapped in, followed by the other gaskets. In configurations with operable vents, all fittings are attached after the gaskets and glazing. Schüco's proprietary fitting kits are designed to be easily attached and removed in no specific order.

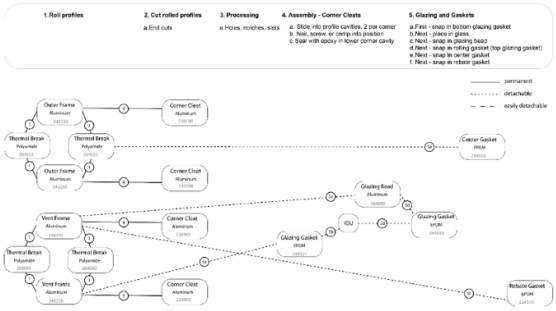


Figure 5. Assembly Sequence Mapping

Using the assembly sequence mapping as a reference and relying more on fabricator knowledge, the disassembly sequence is mapped in Figure 6, showing the components in the order in which they are removed. The remaining components are shaded to indicate that disassembly is no longer possible at these stages.

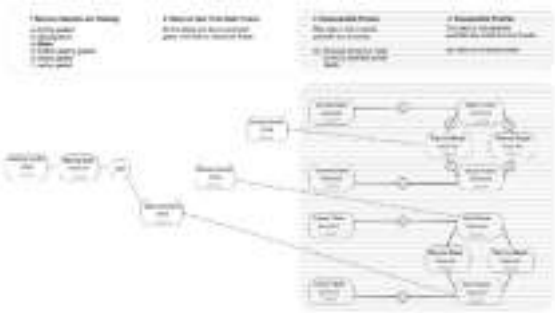


Figure 6. Disassembly Sequence Mapping

5.2 Disassembly Mapping

With system analysis and sequencing complete, disassembly mapping is expanded following portions of the BIMDAS, PAC and MOST® methods. Each separate component from the network analysis is assigned a unique identifier from e^1 to e^n, and each associated connection is assigned a unique identifier from c^1 to c^n. A table is created using these unique identifiers, placing all disconnection actions sequentially, assigning a Sequence ID for every element. Following the principles documented in the BIMDAS method, each component type is assigned several necessary attributions for calculating the Deconstruction Score, the Recovery Score, and DAS. This process establishes a disassembly database for the system. This documentation stage is flexible to adjust for workflow and output goals. Several data points can be mapped from SchüCal, while the rest are mapped from a disassembly database developed for this application to fill the gaps of disassembly documentation. Existing system information taken from SchüCal is the BOM, EPD, system type, and glazing information. Component-specific information from SchüCal is article number, material, secondary finishing, weight, and field number.

Table 2 contains all mapped data of the disassembly database. Column T refers to the component type (ex., frame, gasket, etc.). These tags can be set as defaults for each element type in a system, assuming they are the same for each type. Weight is kilograms, and Material is the material description or associated material code. Fixed (Cf), Bolted (Cb), Nailed (Cn), Prefab (P), Reusable (Ru), Recyclable (Rc), Secondary Finishing (Sf), and Nontoxic (N) are all filled as either true or false. Here, several additions and alterations are made to the table framework detailed in the BIM-DAS method to incorporate inputs for the PAC and MOST methods as well Schüco practices: fitted connections (Cft), demount-ability (D), the disassembly action (DA) as a description of how to remove the component, the sequence ID (SID) integer placing the component along the disassembly sequence, and disassembly time (DT) as the measure of time it takes for that action based on the MOST methodology, measured in TMUs, converted into minutes at later stages.

Table 2. Mapping Data

T	Cf	Cb	Cn	Cft	P	Ru	Rc	Sf	N	D	DA	SID	DT

Database rows are mapped to each component in the BOM fetched by API from SchüCal. A separate API fetches the system's EPDs to calculate circularity indicators following the MCI [15, 17]. In the case study, this database is used to generate disassembly outputs via an external application, but it can also be used for material passport data in Internet of Facades (IoF), the material passport system. Connection-types for each component can be found in the Fabrication Manuals of Schüco products; this is not a guarantee for other companies' workflows. The case study application is designed as a desktop application. With SchüCal open, a user may select any configured project. The application fetches available configuration data and maps disassembly data to each component in the BOM. Figure 7 shows the application workflow with two input streams, internal software SchüCal (for BOM and EPD) and disassembly database (compiled per component, developed with fabricators), and the four disassembly output categories: disassembly instructions, recovery material lists, disassembly and circularity metrics, and design improvement recommendations.

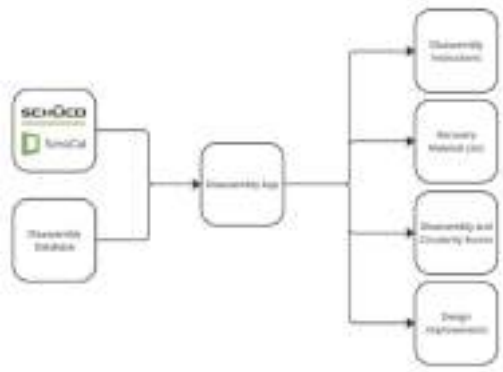


Figure 7. Application Workflow Diagram

The application consists of multiple services, each containing equations and processes relevant for their outputs. The main service maps BOM components to that system's disassembly database and calls calculation engines. One service calculates all ratios necessary for the Deconstruction Score, Recovery Score, and aggregated Deconstructability Assessment Score, and sorts components for material breakdowns and recovery lists. A second service groups and sequences components for the disassembly sequence, calculating time and noting the limitations. The total disassembly time is estimated using the MOST® process. A third service calculates the Material Circularity Indicator, Water Circularity Indicator, and Energy Circularity Indicator. Lastly, a Document Service synthesizes the results from each service into PDF output. Although not included in the case study, this information can also be entered into a material passport to provide stakeholders with up-to-date data.

6 Results

The disassembly sequence for AWS 75.SI+ is an almost direct reversal of the assembly sequence. For configurations with operable vents, disassembly begins with the removal of all fittings. After gaskets and glazing components are removed, the corner cleats and thermal breaks cannot be separated from the aluminum profiles, leaving the outer and vent frame assemblies without any more recoverable components, as thermal breaks and corner cleats cannot be removed without damaging the elements around them. Some assumptions were made to calculate the time required for the complete window disassembly, including that only one skilled worker is required to disassemble the window, and the second assumption is that only one workstation is required. The AWS 75.SI+ fixed light contains 25 components and 6 materials. 5 items are reusable with a total weight of 1.32 kg, 14 items are recyclable at 11.40kg, and 6 items totaling 0.12 kg are designated for waste, none of which contain valuable materials. It has an MCI of 45%, a Deconstruction Score of 79%, a Recovery Score of 57% and an aggregated score of 68%. It has a realistic disassembly time of 42 minutes, with 9 total steps, 3 of which are not possible. Limitations identified are rolled thermal breaks, permanent sealants, and petroleum-derived materials. The AWS 75.SI+ with a Turn-Tilt vent contains 50 components and 7 materials. 11 items are reusable with a total weight of 3.43kg, 32 items are recyclable at 23.48kg, and 7 items totaling 0.21kg are designated for waste, none of which contain valuable materials. It has an MCI of 43%, a Deconstruction Score of 80%, a Recovery Score of 55% and an aggregated score of 67%. It has a realistic disassembly time of 57.85 minutes, with 23 total steps, 6 of which are not possible. There are no additional limitations identified. The most time-intensive process is removing all the hardware from the frames.

7 Conclusion

Design for Disassembly (DfD) for high-performance façade systems is critical to minimizing the extraction of raw materials, lowering recycling energy demands through material reuse, and lowering emissions in the construction sector. Standardization of disassembly modeling and documentation is currently lacking across AEC. As the objectives of circularity regulations and certifications solidify, comprehensive and accessible disassembly documentation future-proofs the framework. In this research, existing disassembly methodologies are compared for relevance to façade system application; components of BIMDAS, Durmisevic's Knowledge Model, PAC, and MOST® are utilized, with network analysis, sequencing, instruction, time estimation, and end-of-life modeling identified as critical to a comprehensive framework. The proposed framework is not exceedingly complex and can integrate with existing digital workflows seamlessly. Interdisciplinary coordination and data collection are front-loaded in the process, consolidating the inputs necessary for multiple disassembly methodologies. Thus, disassembly calculation and presentation processes remain highly modifiable, depending on the use case, generating outputs for various user groups. Fabricator-specific outputs include material recovery lists and disassembly sequencing and time estimation. Designer-specific outputs include disassembly metrics and design limitations and improvement suggestions. Sustainability expert-specific outputs include circularity metrics, recovery lists, and Global Warming Potential. Accessible internal databases are supplemented with input from experienced fabricators. A proof of concept for a

disassembly application was developed to interface with Schüco's existing digital platform SchüCal, documenting the disassembly of the AWS 75.SI+ system for a case study.

In future development, this research's integration of disassembly documentation with BIM tools has the potential to be incorporated directly into the modeling workflow, with a library of parametric components available for connections in new products. Similarly, disassembly modelling can lead to a more dynamic design workflow and results. Digital twin systems can be utilized in combination with BIM and enable real-time monitoring, predictive analysis, and automated decision-making. After construction, BIM data to be continuously updated with sensor feedback and IoT data, ensuring designers and engineers can assess structural conditions, material degradation and disassembly feasibility over time. This research can expand towards the creation of a library of generic and manufacturer-specific products, publicly accessible for utilization in BIM tools.

The methodology outlined in this research and demonstrated in the case study can be adapted to a range of complex building systems, applicable to existing building stock, existing system designs, and future design development. It is particularly pertinent to modular, prefabricated, and high-rise construction. This research demonstrates how disassembly modeling can be integrated into the existing digital workflows of AEC and implemented to map selective deconstruction and Design for Disassembly across lifecycles of building systems.

References

[1] S. Ji, et al. "Building life-span prediction for life cycle assessment and life cycle cost using Machine Learning: A big data approach." *Building and Environment, vol. 205*, 2021.

[2] European Critical Raw Materials Act, Publications Office of the European Union, 2023.

[3] O. Deniz, et al. "Building façade system for deconstruction." *Journal of Sustainable Architecture and Civil Engineering, vol. 8*, 2014.

[4] Homepage | Urbanmining. On-line: www.urbanmining.org, Accessed: 12/12/2024.

[5] Durmisevic E. "Transformable building structures; Design for disassembly as a way to introduce sustainable engineering to building design & construction." Delft , 2006.

[6] O. Akinade, et al. "Waste minimisation through deconstruction: A BIM based Deconstructability Assessment Score (BIM-DAS)", *Resources, Conservation and Recycling*, Volume 105, 2015.

[7] B. Sanchez, et al. "A selective disassembly multi-objective optimization approach for adaptive reuse of building components" *Resources, Conservation and Recycling*, Volume 154, 2020.

[8] D. François, et al. "Using network analysis and BIM to quantify the impact of design for disassembly." *Buildings, vol. 8*, 2018.

[9] MTM (Methods-Time Measurement). The Process Language Methods-Time Measurement (MTM). On-line:mtm.org/en/about-mtm/mtm, Accessed: 12/12/2024.The MOST® Technique, Maynard Operation Sequence Technique. scott-grant.co.uk/terms-and-techniques/most-technique, Accessed 12 Dec. 2024.

[10] C. Langston, et al. "The application of ARP modelling to adaptive reuse projects in Hong Kong." *Habitat International*, vol. 40, 2013.

[11] S. Conejos, et al. "AdaptSTAR model: A climate-friendly strategy to promote built environment sustainability." *Habitat International*, vol. 37, 2013.

[12] G. Formentini, et al. "Design for circular disassembly: Evaluating the impacts of product end-of-life status on circularity through the parent-action-child model." *Journal of Cleaner Production*, vol. 405, 2023.

[13] International Organization for Standardization. ISO 14040: Environmental management—Life cycle assessment—Principles and framework, 2006.

[14] S. Jones, et al., *Methodology an Approach to Measure Circularity*. On-line:https://emf.thirdlight.com/link/3jtevhlkbukz-9of4s4/@/preview/1?o, Accessed: 12/12/2024.

[15] A. González et al. "Methodology to assess the circularity in building construction and refurbishment activities" *Journal of Resources, Conservation & Recycling Advances*, vol.12 ,2021.

[16] buildingSMART International. IDS: Computer interpretable (XML) standard to define Information Delivery Specifications for BIM. On-line: https://github.com/buildingSMART/IDS, Accessed: 12/12/2024.

[17] International Organization for Standardization. ISO 20887 Sustainability in buildings and civil engineering works — Design for disassembly and adaptability, 2020.

[18] Tomczak, et al., Requiring Circularity Data in Bim With Information Delivery Specification, *Journal of Circular Economy*, 2024.

[19] Environmental Product Declaration. On-line: www.byggfaktadocu.se/miljovarudeklaration-1299181/fil-files/schuco_EPD.pdf, Accessed: 12/12/2024.

Enhancing IFC Models for Circular Design: Validation of Required Information through an IDS Framework

Lukas Guntermann[1], Philipp Hagedorn[1] and Markus König[1]

[1]Chair of Computing in Engineering, Ruhr University Bochum, Germany

lukas.guntermann@rub.de, philipp.hagedorn-n6v@rub.de, koenig@inf.bi.rub.de

Abstract -

Given the significant impact of the construction industry on climate change, efforts must focus on reducing energy consumption, emissions, and raw material extraction. Through improved planning and continuous design optimization, building sustainability and circularity can be enhanced. Significant progress has been made in the fields of Building Information Modeling (BIM) and Life Cycle Assessment (LCA). However, it is still necessary to meet the diverse BIM model requirements defined by LCA software vendors, whose proprietary Property Sets (PSets) must be enriched with information in the BIM model for accurate assessments. However, there are also open developments that have agreed on a common information requirement for sustainability assessment and circularity assessment available as datasets in the buildingSMART Data Dictionary (bSDD). This paper presents the development of an Information Delivery Specification (IDS) that combines the information requirements of two LCA software and two bSDDs. The IDS is successfully validated on two BIM models and will be further developed in an ongoing research project.

Keywords -

Circularity, Construction domain, LCA, BIM, IDS, bSDD

1 Introduction

The construction industry is crucial for efficient energy and resource use, because it accounts for an estimated 37 % of global operational energy and process-related CO_2 emissions [1]. The circular economy paradigm emphasizes the importance of reuse over recycling, advocating for the reuse of complete elements as a more sustainable approach than simply recycling materials. To achieve circularity in the construction domain, it is crucial to have a detailed knowledge of the installed elements.

Building Information Modeling (BIM) is a digital planning method that captures extensive planning information, but its application in terms of circular value creation and energy optimization is still limited [2]. To strengthen the connection between sustainable and digital engineering, it is necessary to ensure a consistent data flow between design and assessment, including the circularity potential of elements. Currently, the circularity assessment is based on proprietary information requirements that are custom-tailored for certain vendors.

To facilitate the annotation of BIM models in early design phases with information about reusability and allow for optimizing the design, it is important that this information can be defined in vendor-neutral schemes with open standards.

This paper explores the potential of the Information Delivery Specification (IDS) standard for enhancing information requirement management in circularity assessments. To this end, current requirements from software vendors and publicly available specifications from the buildingSMART Data Dictionary (bSDD) are analyzed and integrated into a unified IDS for validating models in the Industry Foundation Classes (IFC).

Given that various software solutions for Life Cycle Assesment (LCA) are already available, it must be ensured in construction planning that the BIM models used are also suitable for use as input files. The software manufacturers agree on the established IFC standard. For purely geometric and technical planning, the model does not need to be enriched with all the information required for a sustainability and circularity assessment. In particular, the reuse of components is becoming increasingly important. Thus, it must be stored in the data if components originate from reuse or are designed for reuse. Incomplete input data would lead to an incorrect LCA. Assuming that the respective software manufacturers have drawn up modeling guidelines and have either used IFC Property Sets (PSets) for all the required information or have developed their own custom PSets, the following research questions arise:

1. *How can it be ensured that all required information is available in order to carry out an accurate life cycle assessment for buildings?*
2. *How can different proprietary PSets be unified for a vendor-neutral LCA using open standards?*

Answering these research questions, a resulting unified IDS demonstrates how an open standard for LCA can be established and existing approaches harmonized. Its applicability is validated using two IFC models, with further

refinements planned as part of an ongoing research project.

This general objective is addressed in the 'CLUE' research project by the procedure described. This research project aims to develop digital tools that provide feedback on resource-efficient, energy-efficient and circular construction during the planning phase. These tools will evaluate building designs based on energy and resource efficiency and the potential for implementing a circular economy. CLUE is an acronym for the project title '**C**ircu**L**ar b**U**ilding d**E**sign'. The project involves stakeholders from all phases of a building's life cycle, ensuring comprehensive analysis and optimization [3].

2 Background and Related Work

2.1 BIM-based LCA and Circularity Assessment

Life Cycle Assessment (LCA) is pivotal for achieving more sustainable construction practices. Through life cycle analysis, it is possible to evaluate the environmental impacts of an entire building as well as individual components and construction materials, providing guidelines on how planning can be modified to achieve greater sustainability.

According to Feng et al. [4], the adaptive reuse of precast concrete and timber can significantly reduce carbon emissions and concludes that secondary materials should replace virgin materials wherever possible. Reuse of materials leads to a greater reduction in carbon emissions than recycling. Minunno et al. [5] have demonstrated that the positive environmental impacts of reuse far exceed those of recycling, as evidenced by a modular building designed for dismantling and reuse. Specifically, a reduction in greenhouse gas emissions of 88% was achieved with reuse compared to recycling. Prefabricated elements have a higher reuse potential than in-situ and ready-mixed concrete [4]. The comparison shows that wood and metals have high recyclability because they can be easily recycled or reused, unlike concrete which usually has to be crushed. According to Feng et al. [4], maximizing the reuse of materials should be a key strategy for planners when selecting material resources.

Based on the growing demand for sustainable buildings, Azhar et al. [6] reviewed three building performance software solutions in 2009 and described how they work when embedded in BIM. However, the topic of BIM lacks any reference to the IFC standard that has been developed in recent years and this study is concerned with the operation of buildings. Nevertheless, it is found that the greatest influence on the subsequent environmental impact of a building takes place in the early planning phases. Therefore, the environmental impact of the construction project can be analyzed as early as possible during the planning stage. Backes and Traverso [7] include the social and economic impact of buildings in addition to their ecological impact and make a proposal for such an evaluation system. These two aspects should also not be neglected during construction, but it is questionable to what extent they can be analyzed using BIM data or whether it makes more sense to consider them outside of BIM planning. This aspect is, therefore, not considered in this paper. Carvalho et al. [8] have developed a workflow of BIM-based LCA for the portuguese market. They further parameters that are based on the portuguese SBTool^{PT}-H method and could be used in Revit to optimize building sustainability assessment [9]. Based on the fact that circularity research has identified decisive parameters for the LCA of buildings, these must be annotated in the BIM model. Fenz et al. [10] show how PSets can be developed for the sustainability context, compare them with third-party sources and integrate the results into an ontology. This shows how the sustainability parameters can be processed in the BIM context. In addition to the question of how the information required for the LCA gets into the BIM model, it is also crucial that the information in the BIM model is correctly compared with the material databases in order to determine the environmental impact. There is already research in this area, for example, by Forth et al. [11].

There are several software solutions on the market for implementing an LCA (see overview by [12]): The Madaster platform is a well-known LCA software that uses IFC data to analyze the circularity of a building and is also used in scientific publications (see Rieke et al. [13] and Theißen et al. [14]). Madaster can also be used to compare variants of different construction methods. Concular has already been successfully used by Gordon et al. [15] in the field of stocktaking for circularity assessment.

The concept of reuse is inherently connected to urban mining, which emphasizes the importance of establishing a comprehensive knowledge base on potential sources of reusable components and building materials. Additionally, considerations such as minimizing transportation distances and optimizing interim storage durations are critical. Akbarieh et al. [16] present relevant approaches addressing these aspects.

2.2 Open BIM Standards

BIM as a central planning method is established in the construction sector. Open BIM has an open orientation that enables various manufacturers to use the IFC data standard to foster interoperability.

IFC enables the exchange of construction data between different software in the planning process. IFC4 is the current standard of the IFC format as an 'open international standard'[17] and offers, among other things, broader support for infrastructure projects. Version 4 is registered as an ISO standard [18] and has existed since 2014 [19].

In the current IFC4x3 format, geometric, semantic, and alphanumeric information about buildings can be stored. For instance, this can be data that is used for a sustainability assessment. However, the IFC standard only contains the types of information that are industry standard. Other data used by specific software manufacturers can be stored in additional custom property sets (custom PSets). These can be developed and made available by the manufacturers so that full compatibility of the manufacturer's software with a file in IFC format is achieved.

Furthermore, IDS (Information Delivery Specification) defines information requirements in a machine-readable XML format closely related to the IFC format. It is a newly introduced buildingSMART standard that enables the automatic checking of IFC models to verify that all requirements are met [20]. Thus, the IDS file can be transmitted to any stakeholder, that prepares an IFC model for any following purposes. In this context, IDS thus also replaces Model View Definitions (MVDs), which were previously used in part to define information requirements, although the filtering of information is their actual benefit. The IDS format is now recommended by buildingSMART for the definition of information exchange requirements [21].The complete documentation of the IDS schema and the user manual can be found in the buildingSMART Github repository [1].

To facilitate the sharing of standardized definitions for describing the built environment in IFC models, the buildingSMART Data Dictionary (bSDD) is a service provided by buildingSMART [22]. This service aims to ensure the use of consistent and agreed-upon terminology across various software solutions. bSDD comprises a collection of interconnected data dictionaries, each containing definitions of classes and properties relevant to the built environment. These dictionaries are published by independent organizations and researchers and encompass international classifications, national standards, and company-specific agreements. Additionally, bSDD can store relationships between definitions, map between classifications, and provide translations, so that different interpretations of terms by different actors or software are prevented. bSDD focuses on the metadata-definitions of terms used to describe products and projects. This makes bSDD a valuable vendor-neutral reference library for shared definitions in IFC and IDS. The content within bSDD is organized into classes, properties, and their interrelations, adhering to the bSDD data model based on international standards. In the bSDD, the properties listed are divided into groups of properties. However, these groups of properties do not necessarily represent a property set as defined in ISO 16739-1 [23]. The prefix 'PSet_' is reserved for the official IFC and not the bSDDs [23].

[1]https://github.com/buildingSMART/IDS/tree/development/Documentation/UserManual. Accessed: 10.03.2025

3 Methodology and Framework

In order to force more sustainable construction methods and increased circularity, the approach shown in Figure 1 is pursued. First, the information requirements for BIM models are investigated and a modeling guideline is developed on this basis. This incorporates the existing requirements of LCA software manufacturers as well as open approaches. This is the aim of this work. With the help of these modeling requirements, BIM models can be enriched with the necessary information. The result is a collection of existing planning alternatives that can be stored as expert knowledge. If this expert knowledge is incorporated into building planning in the form of a software tool, greater sustainability and circularity can be achieved by considering suitable planning alternatives.

The proposed solution is the creation of an IDS with which the information requirements for the IFC model for an LCA can be checked at any point in the planning process. This is a rule-based approach. The proposed workflow with the IDS is shown in Figure 2:

The IFC model can be checked at any time during planning using the IDS to ensure that it contains all the information required by the LCA software (1). The IFC model is then transferred to the LCA software (2). This could also use the IDS to check whether all the required information has been transferred (3). However, the LCA software analysed already has its own functions for checking whether all the required information is available. Depending on the

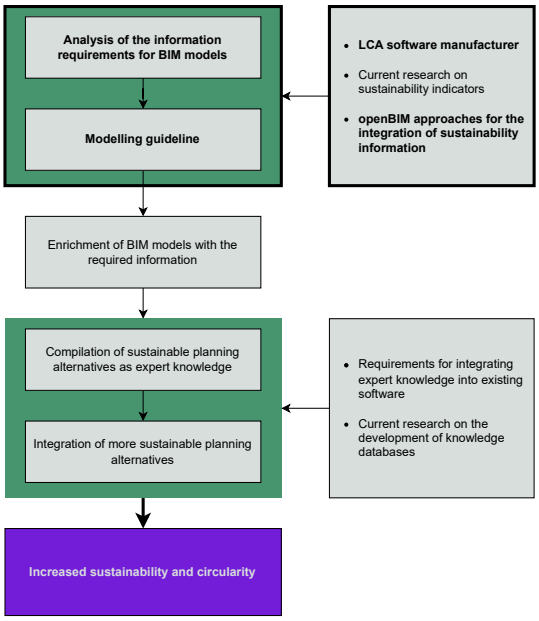


Figure 1. Schematic representation of the overarching research method

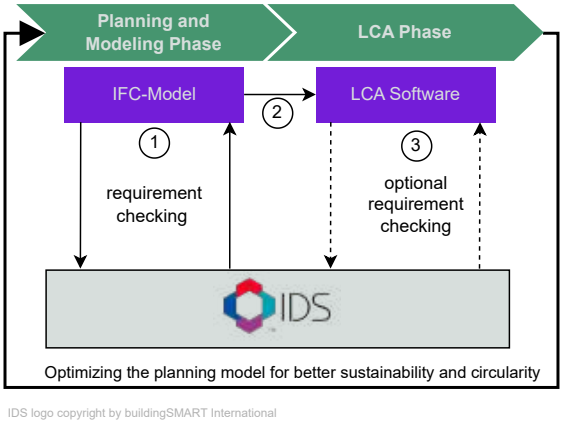


Figure 2. Schematic representation of the proposed workflow for sustainable building planning

assessment by the LCA tool, the design can then be improved in terms of sustainability and circularity. Although Concular and Madaster support both IFC2x3 and IFC4, the IDS created was designed for the current IFC4x3 version, as the different versions each have alternative requirements and no redundant IDS should be set up.

4 Implementation

We implemented a generalized IDS, which combines two proprietary solutions (custom PSets) with open approaches (bSDDs). As described in the methodology and related work above, there are various approaches to carrying out LCA. This paper analyzes the requirements of two distinct software solutions which, according to the DGNB's assessment [12], offer a wide range of functions compared to alternatives. However, despite functioning proprietary solutions, the aim is to map all functions using an open solution.

In Figure 3, the idea of the standardization procedure is shown: The two software solutions of Madaster and Concular and two open bSDDs are combined in an IDS. The bSDD dataset *Decommissioning and Reuse (DOR)* by Akbarieh [24] and the dataset *Circularity* by Al-Qazzaz [25] were used as open solutions. It should be noted that the latter is currently a preview version (as of March 2025).

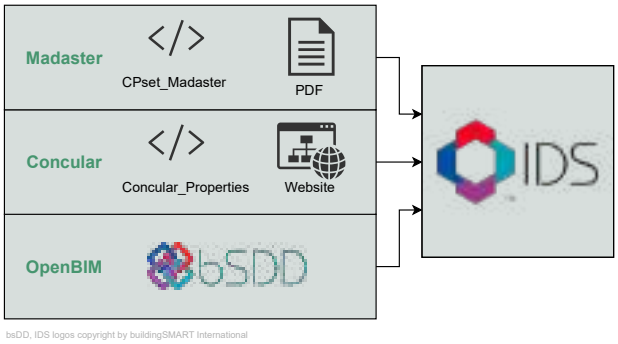


Figure 3. Schematic representation of merging the custom PSets with the *DOR* and the *Circularity* bSDD

Madaster offers a modeling guide in PDF format for preparing IFC models for their LCA platform [26]. The modeling guide contains detailed tables of all property names of their custom PSet. The main disadvantage of a formatted PDF modeling guideline to an IDS is that it is not machine-readable, so there could be some typos included, for example, a false space or a line break. Concular offers their modeling guide on their website [27]. There could be mistakes caused to display errors in the browser or similar. Therefore, all of the property names must be reviewed before transferring them to the IDS.

The respective PSets are also available for download for various BIM software, such as Revit and ArchiCAD [28, 29]. However, the verification of the model to ensure that all information required for the LCA, depending on the specific LCA solution used, is present, takes place within the respective LCA software. Using IDS, this verification can occur without access to the LCA software, allowing it to be integrated into the workflow during the modeling phase and supporting the open BIM approach. This ensures that the planning model being transferred directly meets all requirements, eliminating the need for multiple iterative model adjustments due to missing information.

4.1 Implemented Properties and PSets

For this IDS, parameters of different origins were initially selected to illustrate the standardization approach. The custom PSets of the LCA solutions from Concular and Madaster were selected as well as the mentioned two bSDDs. From these PSets and the bSDDs, parameters were selected that are suitable for describing the reuse case of solid building elements. It is clear from Table 1 that the four approaches partly describe similar element properties, although the property names differ.

4.2 Procedure for Creating the IDS

The IDS was implemented using the IDS web editor by LastBIM [30] as shown in Figure 4.

The main problem in the formulation of a generally valid IDS, in which several approaches are harmonized and which is to check whether the necessary information is available at all elements, is the 'applicability' of the IDS. This must be defined for each requirement. The first significant issue is that only one entity can be employed as applicability (see IDS standard according to buildingSMART [31]). As a result, alternative elements

Table 1. Comparison of PSets [26, 28] and equivalent bSDD [24, 25] - selection of possible properties that could be helpful for the reuse of components and which were saved in the prototype IDS

	Madaster **CPSet_Madaster**	**Concular** **PSet_Concular**	**DOR** **bSDD**	**Circularity** **bSDD (preview)**
1	DetachabilityConnectionType		Connection Type	Type of Connection
2		CpxReusability	Circular Potential	
3	DetachabilityConnectionTypeDetail	CpxDeconstructability		
4	DetachabilityAccessibility			Accessibility to Connection
5	DetachabilityIntersection	CpxSeparability		
6	DetachabilityProductEdge			Geometry of Product Edge
7		CpxPreuse	Current Reuse Count	
8				Probable Outlet
9			Total Reuse Count	



Figure 4. Setting up the IDS with the LastBIM IDS editor [30]

cannot be evaluated, and it must be explicitly clear which elements are present in the IFC model under examination to ensure that the requirement rules are applied precisely to these elements. Another challenge is that abstract IFC classes, which do not have instances themselves but only contain other classes, cannot be checked by the current IDS standard v1.0. This limitation makes it difficult to design the IDS to be as universally applicable as possible.

Therefore, the following approach is used, as shown in Figure 5: First, the IDS was created as shown in Figure 4. The requirements were formulated in general terms and no applicabilities were defined yet. The rules should be applied to all components. The entities for which these rules are to be applied were taken from the IFC documentation as a list [18]. A Python script was then used to manipulate the IDS document in XML format in order to duplicate the requirements, which were only available in one specification and only applied to the placeholder entity, into several new specifications and assign an entity from the list of instantiable entities in the IFC schema to each specification. The newly created IDS is, therefore, the basis for the model check and contains the properties shown in Table 1, which originate from Concular, Madaster, and the bSDDs.

5 Validation

Two existing IFC building models were used to validate the developed IDS. One is a residential building provided by the Institute for Automation and Applied Informatics (IAI) at Karlsruhe Institute of Technology (KIT) [32] as an IFC4 test file, and the other is an office building belonging to the Ruhr University Bochum (RUB). The IDS is validated in the open-source software Blender with these two models, as shown in Figure 6. Blender [33] is known as a 3D-modeling software with roots in video production and special effects. In Blender, the BIM-add-on Bonsai [34]

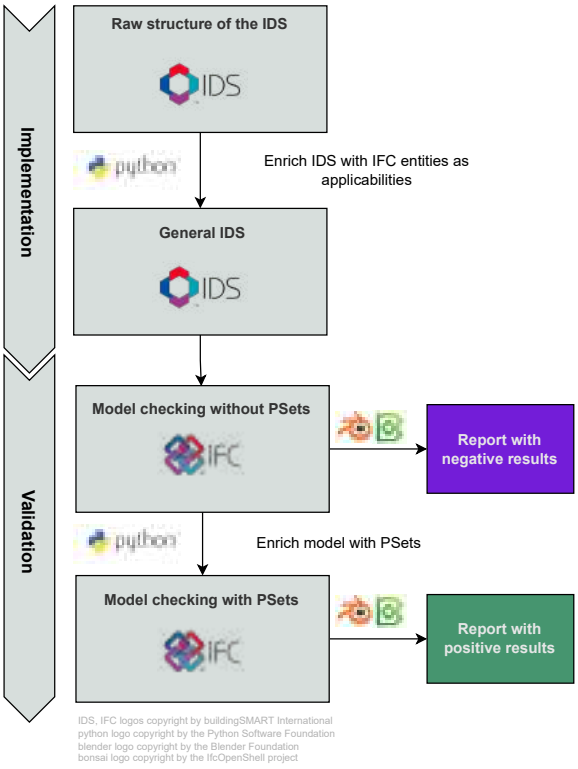


Figure 5. IDS creation and validation workflow

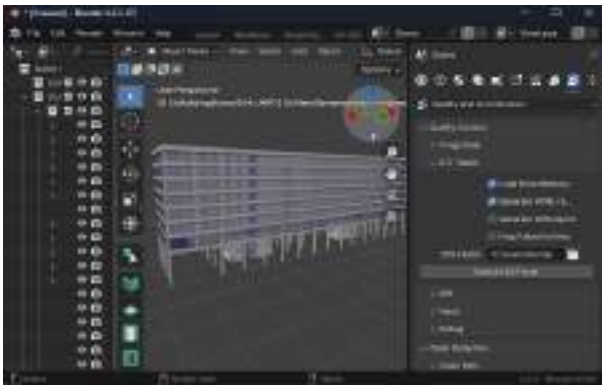


Figure 6. Validation of the IFC model of the RUB office building via Blender and the add-on Bonsai; screenshot of [33], [34]

is used to ensure the open BIM approach and check IDS requirements. An advantage of Bonsai is, that it is based on IfcOpenShell, so it works with native IFC and does not convert it to a proprietary format. The software generates a report after validation, which outlines which requirements are met or violated. This makes it possible to validate that the IDS is working properly. First, the models are checked in which the required PSets and bSDDs are not annotated. This control check, therefore, only resulted in unfulfilled requirements and thus in a negative report (see Figure 5).

Second, both IFC models were edited, and the properties were filled in accordingly with IfcOpenShell [35]. The focus was not on plausible values from an engineering perspective but only on ensuring that all properties were assigned a value. During the audit of the model enriched with sustainability information, the IDS audit rules were fulfilled, and positive results were reported by the IDS checker in Blender. This indicates that the PSets entered were accurately reported as present by the IDS. The IDS model checker displays the results in the program window and generates an HTML report, as well as a detailed ODS report. Based on the entities employed as applicabilities and the selection of PSets implemented (see Table 1), the IDS generated 889 specifications in which a total of 4,318 requirements were checked. Subsequent to this, 2,788 checks were conducted for the IAI building and 497,896 checks for the RUB office building due to the higher number of applicable building elements.

Overall, the IDS standard, as a recently developed standard, necessitates further refinement to ensure its more robust establishment. In this paper, the absence of reference from abstract classes to subordinate classes was of particular significance, as numerous component properties pertaining to reuse must be queried for nearly all solid components. Fischer et al. [36] have previously identified the absence of a reference to related elements, and Kremer and Beetz [37] have demonstrated the integration of geometric and topological checks with semantic information checks, for example, in the simulation of building energy models. The solution provided by the Python script for generalizing the information requirements in this paper could be replaced by a more advanced IDS standard in the long term.

6 Conclusion and Outlook

This paper has reviewed current practices in information requirements management for LCA and circularity assessment and proposed a unified method using IDS. A prototype IDS has been developed and demonstrated to meet the specified requirements for an accurate life cycle assessment for buildings. The proposed method can be utilized in an open BIM environment to ascertain the presence of the requisite LCA data within the IFC model.

The proposed workflow, which involves the implementation of an IDS to verify the requirements stemming from LCA for an IFC model, is thus proven to be a rational approach. However, the current IDS standard v1.0 has applicability limitations that required the use of the Python script to create it. Given the highly diverse requirements and the varying designations of the same model properties across different software, the creation of a universal IDS for LCA necessitates either the standardization of the LCA software solutions to function with the same PSets or an extensive compilation of all prevalent PSets. The latter option is less practical in application, as software manufacturers can continuously evolve their PSets, resulting in the IDS becoming obsolete rapidly. In this regard, software manufacturers could consider the option of automated creation or updating of an IDS. Achieving the greatest conformity with the open BIM working method could be achieved by several participants agreeing on a common bSDD, which is further developed and could be based, among other things, on the four originating sources analyzed. A potential avenue for advancement could be the generalization of the requirements, accompanied by the incorporation of informative reuse data, which could be a by-product in the form of an updated IDS of the CLUE research project. For further validation, a larger study with additional IFC models would be necessary so that all IFC entities used as applicability are validated. The next step is to develop a design-for-reuse, that can be validated using the IDS and assessed for optimizing the reusability potential of a structure.

Acknowledgements

The project ‘CLUE’ is funded by the European Union and the state of North Rhine-Westphalia as part of the EFRE/JTF program NRW 2021-2027. Further information can be found at www.efre.nrw.

References

[1] United Nations Environment Programme. *2022 Global Status Report for Buildings and Construction: Towards a Zero-emission, Efficient and Resilient Buildings and Construction Sector*. Nairobi, 2022. ISBN 978-92-807-3984-8.

[2] Saeed Banihashemi, Senada Meskin, Moslem Sheikhkhoshkar, Saeed Reza Mohandes, Aso Hajirasouli, and Khuong LeNguyen. Circular economy in construction: The digital transformation perspective. *Cleaner Engineering and Technology*, 18:100715, 2024. doi:10.1016/j.clet.2023.100715.

[3] Ruhr University Bochum. CLUE – CircuLar bUilding dEsign – Kreislaufgerechte Planung von Gebäuden, 2024. URL https://clue.blogs.ruhr-uni-bochum.de/. Accessed: 20.11.2024.

[4] Haibo Feng, Qian Chen, Borja García de Soto, and Mehrdad Arashpour. Using BIM and LCA to evaluate material circularity: Contributions to building design improvements. In *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, volume 4, pages 9–16. IAARC Publications, 2022. doi:10.22260/ISARC2022/0004.

[5] Roberto Minunno, Timothy O'Grady, Gregory M. Morrison, and Richard L. Gruner. Exploring environmental benefits of reuse and recycle practices: A circular economy case study of a modular building. *Resources, Conservation and Recycling*, 160:104855, 2020. ISSN 09213449. doi:10.1016/j.resconrec.2020.104855.

[6] Salman Azhar, Justin Brown, and Rizwan Farooqui. BIM-based Sustainability Analysis: An Evaluation of Building Performance Analysis Software. *International Proceedings of the 45th Annual Conference of the Associated Schools of Construction (ASC) at University of Florida, Gainesville, Florida*, 2009. URL http://ascpro0.ascweb.org/archives/cd/2009/paper/CPRT125002009.pdf.

[7] Jana Gerta Backes and Marzia Traverso. Application of Life Cycle Sustainability Assessment in the Construction Sector: A Systematic Literature Review. *Processes*, 9(7):1248, 2021. doi:10.3390/pr9071248.

[8] José Pedro Carvalho, Ismael Alecrim, Luís Bragança, and Ricardo Mateus. Integrating BIM-Based LCA and Building Sustainability Assessment. *Sustainability*, 12(18):7468, 2020. doi:10.3390/su12187468.

[9] José Pedro Carvalho, Luís Bragança, and Ricardo Mateus. Sustainable building design: Analysing the feasibility of BIM platforms to support practical building sustainability assessment. *Computers in Industry*, 127:103400, 2021. ISSN 01663615. doi:10.1016/j.compind.2021.103400.

[10] Stefan Fenz, Julia Bergmayr, Nico Plattner, Serge Chávez-Feria, María Poveda-Villalón, and Giorgos Giannakis. Integration of building material databases for IFC-based building performance analysis. In *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, volume 115, pages 182–189. IAARC Publications, 2021. doi:10.22260/ISARC2021/0027.

[11] Kasimir Forth, Jimmy Abualdenien, and André Borrmann. NLP-based semantic model healing for calculating LCA in early building design stages. In Raimar J. Scherer, Sujesh F. Sujan, and Eilif Hjelseth, editors, *ECPPM 2022 - eWork and eBusiness in Architecture, Engineering and Construction 2022*, pages 77–84, London, 2023. CRC Press. ISBN 9781003354222. doi:10.1201/9781003354222-10.

[12] German Sustainable Building Council (DGNB). Overview of the life cycle assessment tools with details of the tool manufacturers: alphabetically by tool name [in German]. URL https://www.dgnb.de/filestorages/Downloads_unprotected/dokumente/uebersicht-der-oekobilanztools-mit-angaben-der-herstellenden.pdf. Accessed: 20.11.2024.

[13] Leonard Rieke, Jonas Kerstgens, Maximilian Joe Kaufmann, Christian Blanke, Jürgen Reichardt, and Matthias Schmidt. Bewertung der Zirkularität und Minimierung von CO_2-Emissionen im Gesamtsystem Industriebau – Case Study einer Gebäudebewertung mit madaster. *Journal of Production Systems and Logistics*, (4), 2024. doi:10.15488/17979.

[14] Sebastian Theißen, Björn Pelzer, Jannick Höper, Jan Drzymalla, Claudius Frank, Patrick Bergmann, and Anica Meins-Becker. Open BIM-based LCA of HVAC and circularity assessment using the Madaster platform. In *Clima 2022*. Rotterdam, 2022. doi:10.34641/clima.2022.184.

[15] Matthew Gordon, Luise von Zimmerman, Oushesh Haradhun, Dominik Campanella, Milena Bräutigam, and Catherine de Wolf. Digitising Building Materials for Reuse with Reality Capture and Scan-to-BIM Technologies. In Catherine de Wolf, Sultan

Çetin, and Nancy M. P. Bocken, editors, *A Circular Built Environment in the Digital Age*, pages 41–55. Springer International Publishing, Cham, 2024. doi:10.1007/978-3-031-39675-5_3.

[16] Arghavan Akbarieh, Norman Teferle, and James O'Donnell. Semantic Material Bank: A web-based linked data approach for building decommissioning and material reuse. In Raimar J. Scherer, Sujesh F. Sujan, and Eilif Hjelseth, editors, *ECPPM 2022 - eWork and eBusiness in Architecture, Engineering and Construction 2022*, pages 69–76, London, 2023. CRC Press. ISBN 9781003354222. doi:10.1201/9781003354222-9.

[17] buildingSMART International. IFC 4.3.2 Documentation, 2024. URL https://ifc43-docs.standards.buildingsmart.org/. Accessed: 11.11.2024.

[18] ISO 16739-1. *Industry Foundation Classes (IFC) for data sharing in the construction and facility management industries: Part 1: Data schema*. International Organization for Standardization, Geneva, Switzerland, 2 edition, 2024. URL https://www.iso.org/standard/84123.html.

[19] André Borrmann, Markus König, Christian Koch, and Jakob Beetz, editors. *Building Information Modeling: Technology Foundations and Industry Practice*. Springer International Publishing, Cham, 1 edition, 2018. doi:10.1007/978-3-319-92862-3.

[20] buildingSMART International. Information Delivery Specification, 2024. URL https://www.buildingsmart.org/standards/bsi-standards/information-delivery-specification-ids/. Accessed: 14.11.2024.

[21] buildingSMART International. Model View Definitions (MVD), 2024. URL https://technical.buildingsmart.org/standards/ifc/mvd/. Accessed: 15.11.2024.

[22] buildingSMART International. buildingSMART Data Dictionary (bSDD), 2024. URL https://www.buildingsmart.org/users/services/buildingsmart-data-dictionary/. Accessed: 18.11.2024.

[23] buildingSMART International. Data structure of bSDD - buildingSMART Technical, 2024. URL https://technical.buildingsmart.org/services/bsdd/data-structure/. Accessed: 20.11.2024.

[24] Arghavan Akbarieh. Decommissioning and Reuse, 24.04.2024. URL https://identifier.buildingsmart.org/uri/TUe/DOR/0.0.2. Published in the bSDD by Eindhoven University of Technology.

[25] Ihab Al-Qazzaz. Circularity bSDD: Version 0.0.1 (preview), 19.12.2024. URL https://identifier.buildingsmart.org/uri/Al-Qazzaz/Circularity/0.0.1. Published in the bSDD by University of Nottingham.

[26] Madaster Austria. BIM manual for Revit & Archicad: Technical description of the processing of IFC files in Madaster. URL https://docs.madaster.com/files/at/Madaster%20BIM%20Anleitung_20240303.pdf. Accessed: 15.11.2024.

[27] Concular GmbH. Circular LCA, 2024. URL https://circularlca.de/. Accessed: 12.11.2024.

[28] Concular GmbH. IFC model requirements, 15.11.2024. URL https://concular.notion.site/IFC-Modellanforderungen-27ac3a4b5a1d469c8b9657d84ffced68. Accessed: 15.11.2024.

[29] Madaster Documentation. Downloads, 18.12.2024. URL https://docs.madaster.com/us/en/resources/downloads.html. Accessed: 18.12.2024.

[30] LastBIM. IDS Editor — A quick and easy way to create IDS files. URL https://ids-editor.com/. Accessed: 20.11.2024.

[31] buildingSMART International. IDS 1.0 standard. URL https://standards.buildingsmart.org/IDS/1.0/ids.xsd. Accessed: 18.12.2024.

[32] Institute for Automation and Applied Informatics (IAI) / Karlsruhe Institute of Technology. *IFC 4 Examples*. 2020. URL https://www.ifcwiki.org/index.php?title=KIT_IFC_Examples#IFC_4. Accessed: 19.12.2024.

[33] Blender: Open 3D Creation Software. URL https://www.blender.org/. Accessed: 19.12.2024.

[34] Bonsai: beautiful, detailed, and data-rich OpenBIM. URL https://bonsaibim.org/. Accessed: 19.12.2024.

[35] IfcOpenShell. IfcOpenShell v0.8.0, 2024. URL https://github.com/IfcOpenShell/IfcOpenShell. Accessed: 18.12.2024.

[36] Simon Fischer, Harald Urban, Christian Schranz, Patrick Loibl, and Léon van Berlo. Extending Information Delivery Specifications for digital building permit requirements. *Developments in the Built Environment*, 20:100560, 2024. ISSN 26661659. doi:10.1016/j.dibe.2024.100560.

[37] Noemi Christine Kremer and Jakob Beetz. Extending information delivery specification for linking distributed model checking services. In *EC3 & CIB W78 Conference 2023*, 2023. doi:10.35490/EC3.2023.266.

Design Optimization of Steel Structures with Reused Components Using Generative Design

Golnaz Mirzaei[1], Amin Hammad[1] and Faraeen Homayounfard[2]

[1] Gina Cody School of Engineering and Computer Science, Concordia University, Canada
[2]School of Architecture and Urban Planning, University of Science & Technology, Iran
golnaz.mirzaei@mail.concordia.ca, amin.hammad@concordia.ca, faraeenhomayounfard@gmail.com

Abstract –
In recent years, there has been an increasing focus on adopting circular approaches within the construction sector, particularly reusing building components in new projects. Facilitating design for reuse can significantly expand the scope of deconstruction and reuse efforts. Unlike conventional structural design, which assumes an unlimited supply of standardized components and follows a straightforward process of geometric design, structural analysis, and member sizing, reusing structural components introduces unique challenges. With limited component availability, conventional design methods often fall short, leading to time-consuming trial-and-error processes or non-viable solutions. Despite growing interest, research in this area remains limited, focusing primarily on isolated criteria like minimizing material waste or embodied energy, without offering a comprehensive framework that balances environmental and economic considerations. Additionally, conventional approaches rely on traditional mathematical optimization methods, restricting the exploration of complex design spaces. To address these challenges, this research proposes a novel optimization framework that leverages generative design using genetic algorithms to streamline the design of steel structures with reused components. The framework uses the Non-Dominated Sorting Genetic Algorithm II to refine the geometry and topology of steel structures, minimizing environmental impacts and costs. A case study on optimizing a steel truss design using reused components demonstrates the framework's effectiveness.

Keywords –
Optimization; Steel Structures; Reuse; Generative Design; Circular Construction

1 Introduction

In recent decades, global population growth, urbanization, and industrialization have triggered significant environmental challenges, including global warming, resource depletion, and pollution in air, water, and soil. The construction, renovation, and demolition sector has emerged as a major contributor to these issues. This sector is a major energy consumer and greenhouse gas emitter, contributing nearly 30% to global energy usage and one-third of greenhouse gas emissions [1]. Consequently, construction industry has become a focal point in global sustainability initiatives, pivoting towards innovative approaches grounded in the principles of the circular economy.

Circular construction aims to minimize environmental impacts through the principles of reducing, reusing, recycling, and recovering building components throughout the entire lifecycle [2]. The focus is on strategies that keep materials in use, reducing the need for new resource extraction, energy consumption, and CO_2 emissions [3, 4]. This can occur either through extending the service life of a building, or the reuse and recycling of building components at the end of their service life. Reuse is a more valuable option than recycling because it preserves existing components, giving them a new life and adding value. Reusing building components also reduces energy and embodied carbon more effectively than recycling, as reuse typically requires only minor modifications for the components to be incorporated into new structures. This approach aligns with circular economy practices in construction by emphasizing the importance of retaining and repurposing components to minimize waste and maximize resource efficiency [5–7]. In contrast, recycling typically involves breaking down the product or component into secondary materials, resulting in a loss of quality, value, and energy [3, 8].

Among the four stages of the building life cycle—production, construction process, use, and end of life—the first stage (i.e., production), which includes raw

material supply, transport, and manufacturing, can account for up to 50% of total embodied impacts [9]. Therefore, structural engineers can make the most significant reductions in embodied impacts by focusing on designing with reclaimed components [10]. Also, according to Webster [11], load-bearing structures contribute 20-70% of a building's embodied impacts, depending on factors like building type, size, and used components. By incorporating reused primary components into the design of load-bearing structures, structural designers can further reduce embodied impacts [10, 12]. This is particularly important for steel components, where reuse significantly lowers embodied carbon by avoiding the energy-intensive production of steel components made of virgin or recycled steel. Brütting et al. [13] demonstrated the efficacy of this approach in a real-world application, designing the truss for the main station roof in Lausanne using reclaimed components from electric pylons, resulting in a 63% reduction in embodied energy compared to newly fabricated components.

However, reuse is not yet widely considered in the architecture, engineering and construction field. For instance, stakeholders in construction projects often avoid reuse due to economic uncertainties [14, 15]. Designers are also concerned that the use of reusable components, which are often limited in shape and quantity, could complicate the design process [14, 16]. Additionally, the lack of accessible information on available reusable components poses a significant challenge [14, 15, 17]. The primary goal of this research is to develop a comprehensive optimization framework that uses advanced computational methods to streamline the design of steel structures with reused components. By utilizing generative design, the framework will address challenges related to designing with components with limited availability while minimizing cost and environmental impacts.

2 Literature Review

2.1 Optimization of Structures Made from Reused Components

Brütting et al. [13] introduced a novel framework for designing truss systems using reused components. Their approach formulates the problem as a mixed-integer linear programming (MILP) model to minimize mass and cut-off waste, while meeting structural safety and serviceability requirements. The framework begins by defining a stock of reusable components characterized by their material properties, geometric attributes, lengths, and available quantities. Through discrete structural optimization, the proposed framework iteratively optimizes the topology and geometry of the truss to align with the available stock. Structural constraints, such as equilibrium, stress, and serviceability, are integrated to ensure compliance with performance standards. A two-step process is employed, first solving the assignment problem to allocate stock components optimally and then adjusting the geometry to reduce cut-off waste. Brütting et al. [13] expanded their work to propose a *kit-of-parts* approach, enabling the reuse of components across multiple structural configurations. This method integrates two optimization problems: one for assigning stock components to predefined layouts and another for designing a shared stock inventory. The iterative process optimizes both individual structures and the shared inventory to reduce waste and minimize unique component requirements. De Boer [10] further advanced the optimization of reusable structures by incorporating connection design into the process. Recognizing the importance of connections for enabling reuse, the study developed a parametric design tool using Grasshopper for Rhino to automate the design process. By integrating connection length as a variable, the methodology optimizes both structural layouts and bolted steel connections.

2.2 Using Generative Design for Structural Design and Reuse

The complexity of optimizing structures for reuse, especially when dealing with diverse components, has made traditional design and optimization methods impractical. In other words, if the types of reusable components vary or large quantities of components are needed, the large number of alternatives will make it impossible to find the optimal solution using traditional design methods. Therefore, a metaheuristic approach is necessary to tackle such complex problems [18]. One effective metaheuristic approach is generative design, which generates, evaluates, and selects multiple potential solutions based on their effectiveness. This approach often relies on genetic algorithms (GAs) for optimization. GAs are powerful tools that have gained attention in structural design and other engineering domains due to their ability to solve complex optimization problems [18]. As defined by Koza [19], a GA is a sequence of mathematical operations that transforms individual objects within a population into a new population by selecting a subset based on fitness criteria. This process mimics natural selection, where the fittest individuals are chosen to pass on their traits to the next generation. GAs are particularly effective in solving problems with discrete variables, such as the selection of structural components, and are valuable for exploring optimal combinations of structural components under specific constraints in vast solution spaces [20].

Building on the capabilities of generative design and GAs, Kim & Kim [18] explored the structural reuse in modular

construction with a novel method that leverages GAs to minimize both CO_2 emissions and costs. Their study focuses on noise barrier tunnel (NBT) construction. NBTs are modular structures installed along highways or railways to reduce noise pollution. The research integrates Building Information Modelling (BIM) with a multi-objective optimization framework using Non-Dominated Sorting Genetic Algorithm II (NSGA-II). The process begins with a BIM model providing detailed attributes of reusable components, such as shape, length, cross-sectional properties, curvature, and remaining life. This data forms the foundation for optimization, aligning the reuse strategy with structural and environmental constraints. Also, Van Marcke et al. [21] introduced a framework that leverages GAs to optimize the reuse of partially disassembled triangular components from steel trusses. Unlike previous approaches, which emphasized full disassembly and reuse at the component level, this method minimizes the labour-intensive and costly aspects of complete disassembly by leveraging partially intact structural components. The methodology consists of two main components: an aggregation engine for generating initial truss designs and a GA for optimizing these designs. The aggregation engine arranges reclaimed triangular subassemblies row by row to fit the desired truss geometry, allowing for cutting or the addition of new components when necessary. Once an initial design is created, the GA optimizes it by minimizing the total length of new components and reducing the amount of cutting required for reclaimed components.

2.3 Research Gaps

Existing research on structural design optimization with reused components has made valuable contributions but remains limited in scope. Most studies focus on specific criteria, such as minimizing material waste, reducing embodied energy, or optimizing connections, without providing a comprehensive framework that integrates environmental sustainability, economic efficiency, and the maximization of reused components. Moreover, many approaches rely on traditional optimization methods, which restrict the exploration of complex design spaces.

To address these gaps, this research introduces a framework that leverages generative design to explore a broader range of criteria. By considering material waste, energy consumption, CO_2 emission, and cost as a weighted multi-objective optimization problem, the framework achieves a balance between environmental and economic considerations.

3 Proposed Framework

As shown in Figure 1, the proposed framework employs a seven-step computational approach to optimize the geometry and topology of steel structures by reusing predefined stock components, with a focus on minimizing cut-off waste, cost, energy and CO_2 emission. This framework systematically integrates reusable components while ensuring structural integrity and economic feasibility. The process begins with designing an initial structure (topology and geometry) using new components to ensure feasibility, followed by an iterative process of replacing them with suitable reusable components. Each design alternative is generated and evaluated using GA. The key steps are outlined as follows:

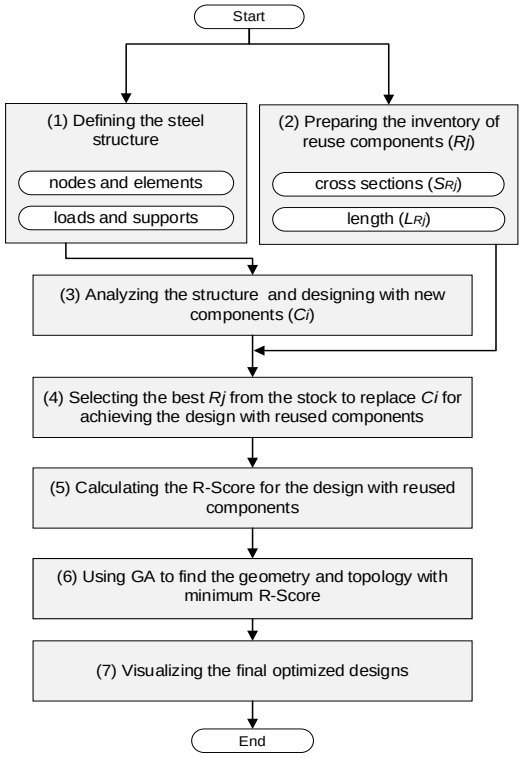


Figure 1. Proposed Framework

(1) Defining the steel structure by specifying its geometry and topology, including nodes and components, along with boundary conditions such as supports and applied loads, to ensure the design accounts for all relevant factors. This step provides a clear starting point for the optimization process.

(2) Preparing an inventory of reusable components ($\boldsymbol{R_j}$) based on their dimensions and properties, including lengths L_{R_j} and cross-sections S_{R_j}, which provides the required data to match components

(3) Analysing the structure to evaluate its performance under specified conditions from Step 1 and **designing the**

initial structure by assigning new components (C_i), which provides a baseline for comparison when incorporating reused components.

(4) Selecting the best match from reused components. This step involves replacing each initial design component C_i with a reusable component R_j from the inventory. The process compares the initial assigned components C_i with inventory components based on two key properties: length and cross section. For R_j to be suitable for C_i, it must meet the following conditions:

$$Match\,(C_i,\,R_j) = \begin{cases} 1 \;\; if\; L_{R_j} \geq L_{C_i}\; and\; S_{R_j} \geq S_{C_i} \\ 0 \;\; otherwise \end{cases} \quad (1)$$

where L_{R_j} and L_{C_i} are the lengths of the inventory component R_j and the design component C_i, respectively; S_{R_j} and S_{C_i} are the cross-sections of the inventory component R_j and the design component C_i, respectively.

After identifying all suitable matches, each initial design component C_i is replaced with the best match from the inventory. The best match, R_j, is the component that is closest in length (L_{R_j}) and cross-section (S_{R_j}) to C_i. If no suitable R_j is available in the inventory, a new component N_i is used to fulfil the requirements of the initial design component C_i.

This approach ensures that the final design uses the lightest and most appropriate reusable components, maximizing material efficiency, while ensuring feasibility and structural performance are not compromised.

(5) Evaluating each design by calculating a Reuse Score (R-Score), which integrates energy consumption (E), CO_2 emissions (G), cost (C), and material waste (W). This provides a quantitative method for comparing different designs, not only minimizing the cost but also environmental impact indicators.

Each parameter is derived from the specific characteristics of *n* new and *m* reused components within the design. As part of this process, a Life Cycle Assessment (LCA) is conducted to evaluate CO_2 emissions and energy consumption. The LCA focuses on a part of the lifecycle of components, beginning with their supply and ending with their transport to the site. Construction-related energy consumptions, CO_2 emissions, and costs are excluded, as they are identical for all reuse-based designs. The detailed steps for these calculations are outlined below:

- **Energy Consumption (E):** The energy consumption for *n* new components (E_N) includes production and transportation:

$$E_N = \sum_{i=1}^{n} \left(E_{product(N_i)} + E_{transport(N_i)} \right) \quad (2)$$

The energy consumption for *m* reused components (E_R) includes deconstruction, modification, and transportation:

$$E_R = \sum_{j=1}^{m} \left(E_{decon.(R_j)} + E_{modify(R_j)} + E_{transport(R_j)} \right) \quad (3)$$

The total energy consumption (E_T) for the design is:

$$E_T = E_N + E_R \quad (4)$$

- **CO_2 Emissions (G):** The total CO_2 emissions are calculated similarly to energy consumption:

$$G_N = \sum_{i=1}^{n} \left(G_{product(N_i)} + G_{transport(N_i)} \right) \quad (5)$$

$$G_R = \sum_{j=1}^{m} \left(G_{decon.(R_j)} + G_{modify(R_j)} + G_{transport(R_j)} \right)$$

$$G_T = G_N + G_R \quad (7)$$

- **Cost (C):** The total cost is calculated by summing the production and transportation costs of new components, as well as the deconstruction, modification, and transportation costs of reused components.

$$C_N = \sum_{i=1}^{n} \left(C_{product(N_i)} + C_{transport(N_i)} \right) \quad (8)$$

$$C_R = \sum_{j=1}^{m} \left(C_{decon.(R_j)} + C_{modify(R_j)} + C_{transport(R_j)} \right) \quad (9)$$

$$C_T = C_N + C_R \quad (10)$$

- **Waste (W):** Material cut-off waste is calculated as the total mass (M) of excess material removed from reused components:

$$W = \sum_{j=1}^{m} \left(M_{R_j} - M_{C_j} \right) \quad (11)$$

- **R-Score:** The R-Score is computed as a weighted sum of these four parameters:

$$RS = \alpha \times E_T + \beta \times G_T + \gamma \times C_T + \delta \times W \quad (12)$$

where $\alpha, \beta, \gamma, \delta$ are weights assigned based on project priorities, satisfying:

$$\alpha + \beta + \gamma + \delta = 1 \quad (13)$$

(6) Implementing GA for minimizing the R-Score, balancing components cut-off waste, environmental impact and cost by exploring a wide range of design possibilities. The optimization process utilizes NSGA-II to explore various design alternatives, applying genetic operators such as crossover and mutation to iteratively refine solutions. The key steps include **(a) Generating initial designs** to form the population of alternatives; **(b) Calculating fitness value** by evaluating the R-Score, where designs with lower waste, energy use, CO_2

emissions, and cost achieve a higher fitness value; **(c) Applying genetic operations**, such as crossover and mutation, to create new design alternatives; and **(d) Iteratively refining the geometry and topology** to minimize environmental impact and cost.

For each design alternative, steps 1, 3, 4, and 5 of the proposed framework are repeated to ensure that the optimization process considers all relevant factors and continuously improves the design until the most efficient and sustainable solution is identified.

(7) Visualizing the final optimized design with the highest fitness value, which allows designers to better understand the structural configurations, as well as their cost and environmental impact, leading to better decision-making.

4 Implementation and Case Study

To demonstrate the practical application of the proposed framework, a case study was conducted using Grasshopper [23] and Rhinoceros [24] to optimize the topology and geometry of a 10-meter-long two-dimensional Warren steel truss.

Grasshopper, a visual programming tool within Rhinoceros3D CAD, was selected for its parametric modelling and generative design capabilities. This tool is widely used by architects and engineers for algorithm-driven design through graphical elements. Additionally, the study integrates plugins such as Karamba [25] for structural analysis and Galapagos [26] for GA optimization. The implementation process consists of the following steps of the proposed framework:

(1) Defining the steel structure: The first step involved modelling the truss geometry and topology (Figure 2) in Grasshopper by defining its nodes, elements, and structural boundary conditions, including loads and supports. A Warren truss configuration with verticals was selected, with a fixed overall span of 10 meters and three bays, with a 5 kN load applied along the lower chord nodes. To explore design flexibility, the allowable movement of the upper chord nodes (0, 2, 4, 6) was constrained to ±0.80 meters in both horizontal and vertical directions as shown in Figure 2. Additionally, to ensure the generated design alternatives remain practical, symmetry constraints were applied to upper nodes so that nodes 0 and 6, as well as nodes 2 and 4, move together. The truss height was set at 2 meters, adjustable within a range of 1.2 meters to a maximum of 2.8 meters. Furthermore, the horizontal positions of nodes 0 and 6 were further restricted to +0.8 meters and -0.8 meters, respectively, so that the length of the upper cord will not exceed 10 meters. In addition to geometry, six different topology configurations, defining how nodes are connected, were incorporated to explore alternative structural layouts. These constraints were integrated as adjustable sliders to explore different design alternatives.

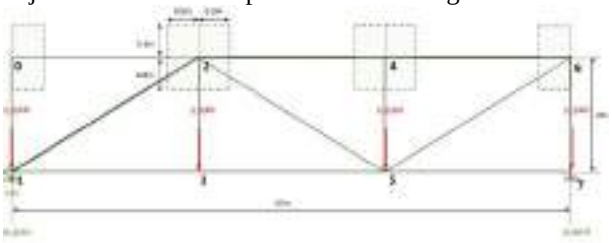


Figure 2. Modelling the geometry of the truss

(2) Preparing the inventory of reusable components: An inventory of reusable steel components (Figure 3) was prepared, consisting of circular hollow sections (CHS) selected from the EN 10219 standards. The cross-sections and lengths of the inventory components were chosen to ensure they are not too large, which would waste material without structural benefits, or too small, which would not properly replace the initial design components. All steel bars are assumed to have a yield strength of 235 MPa, a Young's modulus of 210 GPa, and a density of 7,850 kg/m³. The inventory specifies the types, quantities, and lengths of the available components, ensuring compatibility with the design optimization process.

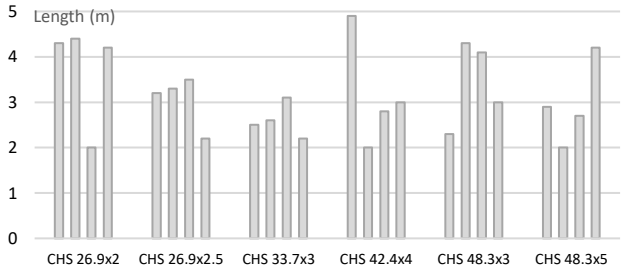


Figure 3. Inventory of reusable components

(3) Analysing the structure and designing with new components: The *Assemble Model* component in the *Karamba* plugin was used to prepare the truss model for structural analysis. This process included defining the cross-section library, loads, joints, and supports. The structural performance of the initial design was then assessed to ensure it met the criteria for stress, displacement, and buckling constraints in accordance with Eurocode 3 (EN 1993-1-1). To achieve this, the *Karamba CroSec Optimizer* was used to determine the lightest cross-sections that satisfied the structural performance requirements while considering stress limits, displacement constraints, and local buckling checks. Specifically, the maximum axial stress in any truss element was limited to 0.9 × yield strength (235 MPa) to maintain a safety margin, while the maximum vertical deflection of any upper chord node was restricted to $L/250$ (e.g., 40 mm for a 10 m span) to comply with serviceability criteria. Additionally, local buckling

checks were incorporated to prevent instability in slender elements.

(4) Selecting the best match from inventory: To incorporate reusable components into the design, each truss components of the design were compared to the available inventory based on two key properties: length and cross-section area. For a component from the inventory to be a suitable replacement, its length had to be equal to or greater than the required length of the design component, and its cross-section area had to meet or exceed the design requirements. Among the suitable options, the component closest in length and cross-section was chosen. If no matching component was found in the inventory, a new steel component was introduced to meet the design specifications.

(5) Calculating the R-Score for the design with reused components: The evaluation of each truss design alternative is conducted by calculating the R-Score, as described in Section 3.

To simplify the analysis, the environmental impacts from modifications are considered negligible. Only components incorporated into the final structure contribute to environmental impacts, as leftover materials are assumed to be reused elsewhere. For new components, it is assumed that they are produced to exact lengths, resulting in no additional waste.

The energy consumption in MJ/Kg and CO_2 emissions in $kgCO2\ eq/Kg$ for a new component, with a transport distance of 70 km are:

$$E_N = 13.227\ M_{N_i} \quad (16)$$

$$G_N = 0.925\ M_{N_i} \quad (17)$$

where M_{N_i} is the mass of the new component. The energy consumption and CO_2 emissions of a reused component are calculated using the following equations, based on transport distances of 200 km [12]:

$$E_R = 3.245\ M_{R_j} + 3.235\ \Delta M_{R_j} \quad (14)$$

$$G_R = 0.277\ M_{R_j} + 0.276\ \Delta M_{R_j} \quad (15)$$

where M_{R_j} and ΔM_{R_j} are the masses of the reused component and the material waste from cutting it, respectively.

The cost estimation for a new component of the truss is calculated by multiplying the unit costs, as shown in Table 1, by its mass (Eq. 18). The cost estimation for the reused components is calculated in a similar way, but it considers both the mass of the uncut component and the cut-off part, as shown in Eq. 19. It is assumed that components are transported between locations at a speed of 40 km/h using a 25-ton truck. The unit cost for purchasing new CHS components is assumed to be 1.25 USD/kg [27], and for this study, the unit cost of reused components is assumed to be half that of new components (i.e., 0.625 USD/kg), and the cut-off waste is assumed to be sold as scrap steel. Consequently, the total costs for new and reused components are estimated using the following equations:

$$C_N = 1.256\ M_{N_i} \quad (18)$$

$$C_R = 1.313\ (M_{R_j} + \Delta M_{R_j}) - 0.325\ \Delta M_{R_j} \quad (19)$$

Table 1. Cost estimation assumptions [18, 22, 27]

Component type	Work	Unit price (US$/kg)
Reuse	Deconstruction	0.383
	Purchase	0.625
	Modification	0.289
	Transportation (200 km)	0.016
	Scrap steel sales	0.325
New	Purchase	1.250
	Transportation (70 km)	0.006

Also, material waste for each design alternative in the case study is calculated based on the material cut-off during the modification process, as mention in Section 3.

For calculating the R-Score, since all factors in this case study are considered equally important, a uniform weighting of 0.25 is applied to each factor (i.e. $\alpha = \beta = \gamma = \delta = 0.25$).

(6) Using GA to find the geometry with minimum R-Score: The Galapagos solver in Grasshopper was then used for generative design to optimize the truss's geometry and topology. Galapagos iteratively refined the design by exploring different configurations of variables (nodes' position and connection) to minimize the R-Score while ensuring structural performance. The GA was tested with two different population sizes of 20 and 50 individuals, both using an elitism ratio of 5%, where the top-performing individuals were carried forward to the next generation. The search space consisted of both discrete and continuous variables, including continuous adjustments in node positions, six discrete topological configurations, and discrete inventory component replacements using a combination of new and reused components. The termination criterion was set to stop the optimization when the improvement in the R-Score is below 1% over 5 consecutive generations. As shown in Figure 4, the GA with a population size of 50 (P50) achieved a lower final R-Score compared to the population size of 20 (P20). The process for the P50 case stopped after 32 generations and took a total of 68 minutes and 17 seconds, including 5 minutes and 28 seconds for the initial design phase.

(7) Visualizing the final optimized truss and results: The final optimized design was visualized in Rhinoceros3D, with detailed structural analysis performed using *Karamba*. Figure 5 shows the final optimized design, which includes 13 reused and 0 new

components. Table 2 compares the criteria considered in this design with those of the initial design, which has the same geometry but uses only new components. The final design reduced energy consumption and CO_2 emissions by approximately 64% and 56%, respectively, while increasing the cost by about 47%. These results show that incorporating reused components in the design can have significant positive environmental impacts that could justify the increase in the cost.

Table 2. Comparing the initial and final designs

Criteria	Initial design	Final design	Change (%)
Energy (MJ)	958.61	589.31	63.73 ↓
CO_2 ($kgCO2\ eq$)	66.78	45.08	55.56 ↓
Cost ($US\$$)	90.68	128.99	-47.24 ↑
Weight (kg)	55.03	62.93	-14.35↑

5 Conclusions and Future Work

This paper proposed a framework using generative design to streamline the design process of steel structures with reused components. It introduces a systematic seven-step framework, leveraging the NSGA-II to iteratively refine structural geometry and topology, enabling efficient matching of limited sets of stock components to the desired design. This framework is able to minimize cut-off waste, energy consumption, and CO2 emissions, highlighting its potential for sustainable design practices aligned with circular economy principles and evolving sustainability regulations. Additionally, it enhances the financial viability of structural reuse by reducing costs in two key areas: (1) material procurement by utilizing existing components instead of new ones and (2) waste management by lowering disposal and handling costs. By demonstrating that reuse can be both environmentally responsible and economically feasible, this study contributes to the broader adoption of sustainable structural design practices.

While the proposed framework offers significant advancements, the case study introduced several simplified assumptions when calculating the R-Score. Future research should aim to improve the framework to enhance the R-Score by incorporating additional criteria and utilizing Pareto optimization. Furthermore, future work would benefit from integrating material passports and Environmental Product Declarations (EPDs) to provide more comprehensive data on the sustainability and lifecycle impacts of reuse. To further validate the framework's robustness, future research will also extend its application to seismic and wind load considerations, ensuring its suitability for a broader range of structural conditions. Additionally, the framework will be applied to three-dimensional space trusses, allowing for a more complex and realistic evaluation of its effectiveness in optimizing material reuse.

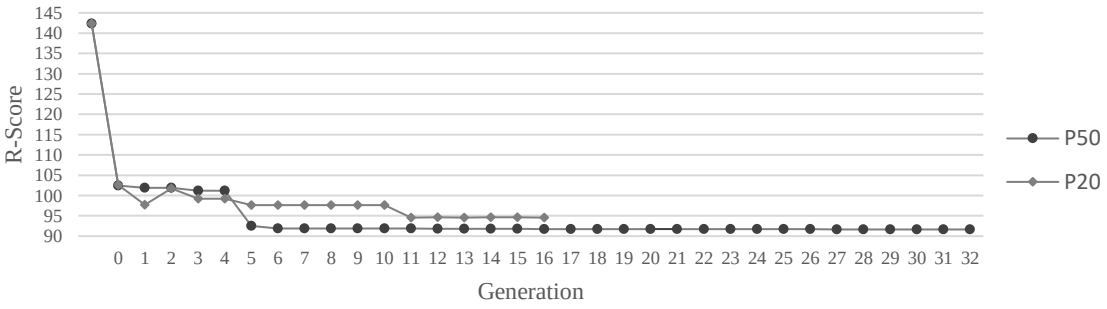


Figure 4. R-Score minimization during the optimization process

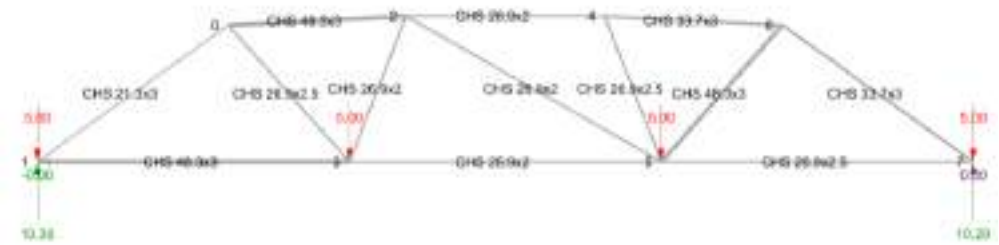


Figure 5. Final optimized design with reused components

References

[1] M. Menegaki and D. Damigos, "A review on current situation and challenges of construction and demolition waste management," *Current Opinion in Green and Sustainable Chemistry*, vol. 13, pp. 8–15, Oct. 2018

[2] O. O. Akinade, L. O. Oyedele, A. Oyedele, J. M. Davila Delgado, M. Bilal, L. Akanbi, A. Ajayi, and H. Owolabi, "Design for deconstruction using a circular economy approach: barriers and strategies for improvement," *Production Planning & Control*, vol. 31, no. 10, pp. 829–840, Jul. 2020

[3] S. Huuhka, L. Aarikka-Stenroos, J. Lahdensivu, P. Jonker-Hoffrén, V. Arnold, E. Stenberg, R. Blok, K. Gudmundsson, P. Teuffel, and A. Mettke, "Recreating the Construction Sector for Circularity," In *The Routledge Handbook of Catalysts for a Sustainable Circular Economy*, 1st ed., London: Routledge, 2023, pp. 42–66.

[4] J. Park and R. Tucker, "Overcoming barriers to the reuse of construction waste material in Australia: a review of the literature," *International Journal of Construction Management*, vol. 17, no. 3, pp. 228–237, Jul. 2017

[5] N. M. P. Bocken, I. De Pauw, C. Bakker, and B. Van Der Grinten, "Product design and business model strategies for a circular economy," *Journal of Industrial and Production Engineering*, vol. 33, no. 5, pp. 308–320, Jul. 2016,

[6] M. C. Den Hollander, C. A. Bakker, and E. J. Hultink, "Product Design in a Circular Economy: Development of a Typology of Key Concepts and Terms," *J of Industrial Ecology*, vol. 21, no. 3, pp. 517–525, Jun. 2017

[7] F. Lüdeke-Freund, S. Gold, and N. M. P. Bocken, "A Review and Typology of Circular Economy Business Model Patterns," *J of Industrial Ecology*, vol. 23, no. 1, pp. 36–61, Feb. 2019

[8] T. Ding, J. Xiao, Q. Zhang, and A. Akbarnezhad, "Experimental and numerical studies on design for deconstruction concrete connections: An overview," *Advances in Structural Engineering*, vol. 21, no. 14, pp. 2198–2214, Oct. 2018,

[9] J. Orr, O. Gibbons, and W. Arnold, "A brief guide to calculating embodied carbon," *STRUCT ENG*, vol. 98, no. 7, pp. 22–27, Jul. 2020

[10] M. B. G. de Boer, "Design and optimization of structures with re-used primary elements and new robotic manufactured connections." Master Thesis. University of Eindhoven, 2021.

[11] M. D. Webster, "Relevance of Structural Engineers to Sustainable Design of Buildings," *Structural Engineering International*, vol. 14, no. 3, pp. 181–185, Aug. 2004

[12] J. Brütting, J. Desruelle, G. Senatore, and C. Fivet, "Design of Truss Structures Through Reuse," *Structures*, vol. 18, pp. 128–137, Apr. 2019

[13] J. Brütting, G. Senatore, and C. Fivet, "Form Follows Availability – Designing Structures Through Reuse," *Journal of the International Association for Shell and Spatial Structures*, vol. 60, no. 4, pp. 257–265, Dec. 2019

[14] Rose, C., "Systems for Reuse, Repurposing and Upcycling of Existing Building Components", Doctoral thesis. University College London, London, UK, 2019.

[15] D. D. Tingley, S. Cooper, and J. Cullen, "Understanding and overcoming the barriers to structural steel reuse, a UK perspective." Unpublished, 2016

[16] E. Iacovidou and P. Purnell, "Mining the physical infrastructure: Opportunities, barriers and interventions in promoting structural components reuse," *Science of The Total Environment*, vol. 557–558, pp. 791–807, Jul. 2016

[17] P. Sassi, "Defining closed-loop material cycle construction," *Building Research & Information*, vol. 36, no. 5, pp. 509–519, Oct. 2008

[18] S. Kim and S.-A. Kim, "Design optimization of noise barrier tunnels through component reuse: Minimization of costs and CO2 emissions using multi-objective genetic algorithm," *Journal of Cleaner Production*, May 2021

[19] JohnR. Koza, "Genetic programming as a means for programming computers by natural selection," *Stat Comput*, vol. 4, no. 2, Jun. 1994

[20] H. Ohmori, "Computational Morphogenesis: Its Current State and Possibility for the Future," *International Journal of Space Structures*, vol. 26, no. 3, pp. 269–276, Sep. 2011

[21] A. Van Marcke, V. Laghi, and J. V. Carstensen, "Automated planar truss design with reclaimed partially disassembled steel truss components," *Journal of Building Engineering*, vol. 84, p. 108458, May 2024

[22] S. Kim and S.-A. Kim, "Framework for Designing Sustainable Structures through Steel Beam Reuse," *Sustainability*, vol. 12, no. 22, Nov. 2020

[23] *Rhino - Rhinoceros 3D*. [Online]. Available: https://www.rhino3d.com/. [Accessed: Jan. 14, 2025].

[24] *Grasshopper - Algorithmic Modelling for Rhino*. [Online]. Available: https://www.grasshopper3d.com/. [Accessed: Jan. 14, 2025].

[25] *Karamba3D - Parametric Structural Engineering*. [Online]. Available: https://karamba3d.com/. [Accessed: Jan. 14, 2025]

[26] *Galapagos - Evolutionary Solver for Grasshopper*. Included with Grasshopper - Algorithmic Modelling for Rhino. [Online]. Available: https://www.grasshopper3d.com/. [Accessed: Jan. 14, 2025].

[27] *Buy A Beam*. [Online]. Available: https://shop.buyabeam.com/suppliers/surrey/tube/steel-tube-S235/products/1397-x-6-chs/. [Accessed: Jan. 14, 2025].

Advancing Circular Economy in the Built Environment: A Framework for User-Centered Software Tool Development

Sai Karthik Vakkanthula[1,2] and Mazdak Nik-Bakht[1,3]

[1]Dept. of Building, Civil, & Environmental Engineering, Concordia University, Canada
[2]s_vakkan@live.concordia.ca, [3]mazdak.nikbakht@concordia.ca

Abstract –

The construction industry (CI) is widely recognized for its inherent reluctance to adopt new digital tools, largely due to less than optimum user interface and user-centered functionalities compared to other sectors. Poorly designed software tools that fail to consider users' needs, such as ease of use, adaptability, and intuitive learning experiences face limited acceptance. Despite this understanding, few studies have reviewed user-oriented capabilities in software tools for the construction industry, particularly those designed for circular economy applications, resulting in tools with fragmented functionalities. This qualitative study addresses this gap by identifying key user-centered requirements for circular economy (CE) software tools (STs), focusing on tools for material inventorying. A literature-based review is performed to develop an evaluation framework adapting the framework of life-cycle assessment (LCA) tools for early-design stages four themes: Input, Analysis, Output, and User Interface (UI) or User Experience (UX). Using this framework, four commercially available STs for material inventorying of built facilities were evaluated to understand their relevance to support CE practices. Results indicate that essential user requirements such as integration, decision-making support, adaptability, and interoperability with other CE technologies are largely unmet. Additionally, no comprehensive tool currently addresses performing financial analyses of various deconstruction techniques, and disposal choices. The proposed framework with user-centered criteria provides a valuable roadmap for developing user-friendly software tools, to advance circular economy practices. While the user-centered needs are defined based on literature, there is a necessity to assign an order of importance to these criteria from the perspective of industry stakeholders.

Keywords –

Deconstruction, Circular Economy, User-Centered Design, Evaluation Framework, Tool Review

1 Introduction

The construction industry significantly contributes to global solid waste. In Canada, the construction, renovation, and demolition (CRD) waste accounts for 12% (4 million tonnes) of Canada's total solid waste, with only about 16% diverted from landfills in 2010 [1]. Circular economy (CE) offers an effective solution to manage CRD waste by eliminating the end-of-life (EOL) phase of materials. CE aims to extend the materials and components lifespan in the built environment, maximizing their economic, environmental, and social value [2]. This is achieved through the direct reuse of materials and components from old structures as secondary resources in new construction [3].

Maximizing recovery of secondary resources requires understanding of material and component stock (MCS) composition of built facilities nearing EOL. This knowledge guides the selection of deconstruction techniques and their sequence, and in turn, influences the type of resources generated and second-life use decisions [4]. However, in legacy buildings, this data in the format of as-built documentation is often unavailable or incomplete [5], necessitating labour-intensive and non-scalable manual pre-demolition audits.

Demonstrating the potential secondary resources is necessary to shift stakeholders' perception from conventional demolition to selective demolition, if not deconstruction. However, most of the current research to simplify estimation of MCS composition and reduce manual efforts focuses on urban-scale and above, helping governments assess potential and formulate regulations for urban mining. They fall short in supporting other industry stakeholders, especially contractors and developers, who require MCS information at smaller spatial scales, preferably building level or lower [6].

With increasing recognition of the value of inventorying existing facilities [7], there is a need for

reliable, quick, accurate, and affordable software tools (STs) that estimates MCS of built facilities, effectively transferring frameworks devised in research to industry. Despite this pressing need, few STs exist and most lack practical functionalities, limiting their use to inform decision making.

Yet, a tool that only estimates the material and component stock may not be able to provide enough value to advance the idea of circular economy practices or get widely adopted by the industry users. If a tool does not follow needs-driven development or imposes labor-intensive workflows, adoption is unlikely [7]. For instance, a Dutch social housing organization generating material passports (MPs) for its building stock as a pilot project preferred a service provider that automated MP creation via artificial intelligence (AI) vision techniques over a provider's tool that made use of building information modelling (BIM) data and required manual input. The users in the latter case perceived MPs as just another task, rather than recognizing its underlying value [8]. This user perception reflects that even with identical outputs, tools that address user needs and minimize workload are favored.

Evaluating how well existing STs are designed to meet the needs of stakeholders is crucial, considering the wide recognition of the construction industry's (CI) inherent reluctance to adopt new digital tools [8]. This reluctance is often paired by less than optimum user interface (UI) and user-centered functionalities in CI compared to other sectors [9]. This understanding is essential for developing STs for fast quantity take-off (QTO) of built facilities, which are still in their infancy, that go beyond material stock estimation and account for other user-centered needs.

What are these user-centered design requirements that must be incorporated into software tools for fast QTO of built facilities to advance CE? This paper seeks to evaluate the technical features, and user-centered design requirements for such a tool by addressing the following specific objectives:

1. Understand user-centered requirements to increase a software tool's acceptance; and
2. Identify the capabilities and limitations of commercially available software that support automated QTO by assessing their alignment with the defined user-centered requirements.

2 Methods

This paper achieves its objectives in two steps. In first step, the criteria from existing studies on Life Cycle Assessment (LCA) tools for early design stages are utilised to establish user-centered design requirements of STs for automated QTO in the CE space. In second step, commercial software tools are identified and evaluated based on their features, shortcomings, and alignment with the defined criteria, providing insights for future ST development. The two steps are discussed in detail below.

To define user-centered requirements, criteria formulated by [10] -[12] to assess LCA tools for early design stages were taken as a starting point, see Figure 1. This is because these tools share the similar context: limited data availability on building geometry and materials during early design stages, and they aid users in decision-making of design choices. This parallels the requirements for STs for fast QTO of built facilities.

These criteria were refined and adapted to fit the specific nature of fast QTO for built facilities, through a review of articles on digital technologies for the circular economy, resulting in a revised framework with four themes: Input, Analysis, Output, and User Interface (UI) or User Experience (UX).

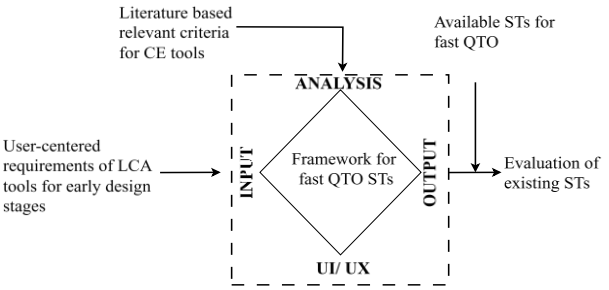


Figure 1. Methodology for framework development

The list of STs for automated QTO was compiled through a review of published and grey literature as well as an informal Google search. A keyword search using ("Software Tools" OR "Digital Tools") AND ("existing buildings" OR "building stock" OR "legacy buildings" OR "built environment") was conducted in "Scopus" and "Google Scholar" yielding many results. The search identified tools that are either proposed in research, existed as proprietary software, or function primarily as material inventory management software for recovered materials at a building's EoL. While these STs come under the broader CE context, they are either underdeveloped as research projects or unavailable for purchase or subscription, or lack core CE functionalities, making them unsuitable for this study's scope. To ensure relevance, the STs presented in this study are selected based on the following criteria:

- focus on built facilities (as opposed to new construction);
- generate material composition as a primary function or as part of other functionality, such as LCA;
- provide granularity at the building level or lower (i.e., building assemblies and subsystems);
- do not require detailed as-built drawings or 3D

digital models; and
- are not BIM software plugins or addons.

The selected STs were evaluated based on their software characteristics, purpose, data requirements, analysis mechanisms, output generation, and limitations, gathered from software websites, documentation, and literature. Subsequently, their alignment with the user-centered requirements was analysed.

3 Framework Development

Few studies analysed the essential functionalities required for software tools to promote a circular economy but have done so in a fragmented way. Rexfelt and Selvefors [7] provided a comprehensive review of circular design tools and their characteristics. Çetin et al. [8] and Yu et al. [9] identified barriers to adopting digital technologies for circularity in the built environment. Akinade et al. [13] outlined the essential technical functionalities for a BIM-based deconstruction tool to ensure wider acceptance. Relevant recommendations from these studies were used to revise and adapt original framework, LCA tools for early design stages, to suit the context of STs for fast quantity take-off of built facilities. These user-centered requirements are categorized into four key themes: Input, Analysis, Output, and UI/UX, and are summarized in Figure 2, with each theme represented by a distinct colour. These criteria are discussed in detail in the following subsections.

3.1 Input theme

In terms of inputs, the primary requirement is the integration of software tools into existing systems of users' organizations, with minimal need for new infrastructure or adjustments, and the ability to read information from existing databases. Additional requirements include limited data input, indicating the use of easy-to-access building characteristics or properties, and support for scalability.

3.2 Analysis theme

In the analysis theme, users favour transparency concerning the assumptions, equations, and thumb rules used to derive the material quantities [12], for a better understanding of a software tool's performance. Another key requirement is fast calculation, which in the current context implies minimal run time to generate results through reduced computational complexity and resource intensity. The tool's acceptance is also affected by its ability to support decision-making through the evaluation of alternative disposal choices, respective financial costs and returns [8], making the potential profits tangible.

Apart from the above, Kobylinska et al. [14] suggested that STs using machine learning to predict material quantities should include a pipeline in the system architecture for continuous training of machine learning (ML) models on the expanded dataset when certain preset criteria are met.

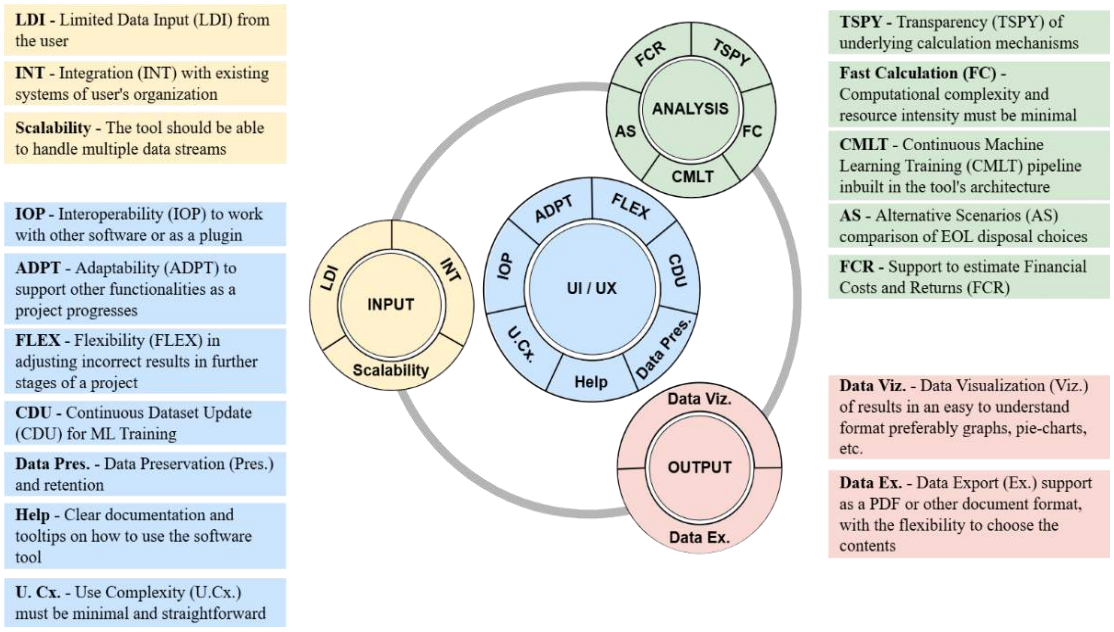


Figure 2. User-centered requirements for software tools

3.3 Output theme

The qualities favoured in the output theme include results presentation in visually appealing formats such as bar charts, pie-charts, etc., and the ability to generate reports in commonly used file formats for data sharing among stakeholders [12]. The tool should also allow users to query material quantities aggregated at the whole building level as well as by building systems and assemblies. Additionally, it should enable exporting the results in an open file format for use in other circular digital tools.

3.4 UI/ UX theme

Under this theme user-centered requirements, which are general, include ease of use and intuitive learning, floating tooltips, data preservation to prevent loss, and documentation for tool use and troubleshooting. Other characteristics preferred include interoperability, representing communicability with other software such as geographical information systems (GIS) to query geometrical information of a building, circularity indices, LCA, MPs, and digital marketplaces for secondary resources. These tools are being increasingly used in the construction industry to drive CE initiatives. The adaptability of a tool as the project advances into the next stages is another important requirement, which for the circular economy means support for deconstruction scheduling, in-built material passport generation, end-of-life LCA, etc. Since the material quantities generated by the STs are not always accurate, the tools should be flexible enough to allow users to modify quantities and specifications when inaccuracies are identified in later project stages.

4 Evaluation of Commercially Available Software Tools

To assess the extent to which the user-centered criteria were incorporated into commercial software, STs were identified and evaluated for alignment, with results provided below.

4.1 Existing Software Tools

Four software tools, with an overview provided in Table 1, were identified and selected after performing the search outlined in section 2: Demolition Waste Generation (DWG) Calculator for Metro Vancouver, EVALMETAB, Carbon Designer 3D add-on for One Click LCA, and the LCAbyg tool. Of the four tools, two tools focus on demolition waste, and the other two are designed for the whole building LCA.

DWG Calculator, Metro Vancouver is a web-based developed for the Vancouver Metropolitan Area, British Columbia to allow denizens to estimate the waste generated from residential demolitions [15]. The tool estimates the quantities of five materials, in kg (or) lb, using MICs based on gross floor area (GFA) and the municipality where the house is located.

EVALMETAB is a web-based tool, specific to France, that quantifies fourteen materials, in tons, based on GFA and covers 14 typologies of existing buildings and 12 typologies of new buildings. The tool allows users to visualize the materials generated from the demolition of old buildings and those required for new construction. It also supports the evaluation of various recycling scenarios, transportation cost estimation for waste disposal, and the associated greenhouse gas emissions [16], [17].

Carbon Designer 3D is a web-based add-on for the One Click LCA tool, originally designed for early design stages, generates predefined assemblies and corresponding material quantities of existing buildings to perform whole building LCA. The areas of predefined assemblies and material quantities are calculated based on parametric equations and thumb rules. Users start by selecting country-specific representative buildings from predefined types and input information regarding building characteristics to estimate the areas of assemblies aggregated by building systems [18], [19]. The tool allows users to modify the estimated quantities, assembly types, and material composition of the assemblies, and inventory is presented to the user in tabular format and carbon footprint as a heatmap on a 3D model of the building.

LCAbyg is a desktop-based building LCA tool for Danish buildings and features an automated material inventory functionality [20], [21]. The tool requires users to input information for eight variables, which are then used to calculate the quantities of predefined component layers of assemblies using predefined geometrical constants and parametric equations. At present, the tool provides automated inventorying of only one typology, masonry apartment – 1850 to 1920, and has the flexibility of allowing users to modify the generated results.

4.2 Discussion

The few software tools available in the market are not comprehensive and focus on demolition waste prediction at the whole building level for limited typologies or support whole building LCA rather than building material inventorying, falling short in supporting the circular economy for built facilities. Apart from technical functionalities, the available STs fall short in their lack of alignment with user-centered requirements.

Key user requirements, as outlined in Figure 3, missing from available tools include integration, decision-making support, adaptability, interoperability with other CE technologies, and enhanced data output

Table 1. Overview of software tools for automated inventorying of built facilities

	Tool name	DWGR Calculator	EVALMETAB	Carbon Designer 3D	LCAbyg
CHARACTERISTICS	Type	WebApp	WebApp	WebApp – add-on for OneClick LCA	Desktop Application
	Country	Canada	France	Europe, USA, Canada, and International	Denmark
	Release Date	July 2018[1]	June 2021	May 2022	October 2020
	Access	Open-Access	Open-Access	Paid	Open-Access
	Purpose	Demolition Waste	Material Flows and Demolition Waste	WBLCA	WBLCA
	Calculation Algorithm	MIC*	N.S. *	Geometrical and parametric equations, average ratios, and thumb rules	Geometrical and parametric equations
CAPABILITIES	Number of Typologies	1	14 - Existing and 12 - New	15	1
	Number of Building Systems	0	0	6	7
	Lifecycle Phase	End-of-Life	End-of-Life and Early-Design	Early-Design and Renovation	Renovation and End-of-Life
	Unit of Measurement	kg (or) lb	Tons	m^2 or m	m^2 or m
	Level of Granularity	Whole Building	Whole Building	Assemblies	Assemblies
	Level of Recovery	Materials	Materials	N.A. [2]	N.A. [2]
	Calculable Materials	5	12	N.A. [2]	N.A. [2]
INPUT	Input Features (Total/ Categorical/ Numerical)	2 / 1 / 1	1 / 0 / 1	17 / 3 / 14	8 / 5 / 3
OUTPUTS	Special Features	None	Recycling Costs	3D building model with carbon heatmap	None
	Data Export	None	Excel	PDF, Excel, CSV, XML	PDF, Excel, JSON
	Data Visualization	None	Bar charts	Predefined charts, Sankey, Spidergrams	Predefined charts

* MIC – Material Intensity Coefficient (material weight in kg/ m2 of floor area), WBLCA – Whole Building LCA,
N.S. – Not Specified, N.A. – Not Applicable
1. Reflects the first known reference to the tool. However, exact release date remains unavailable
2. The software tools provide a detailed material composition of each assembly; however, material quantities must be added manually if required.

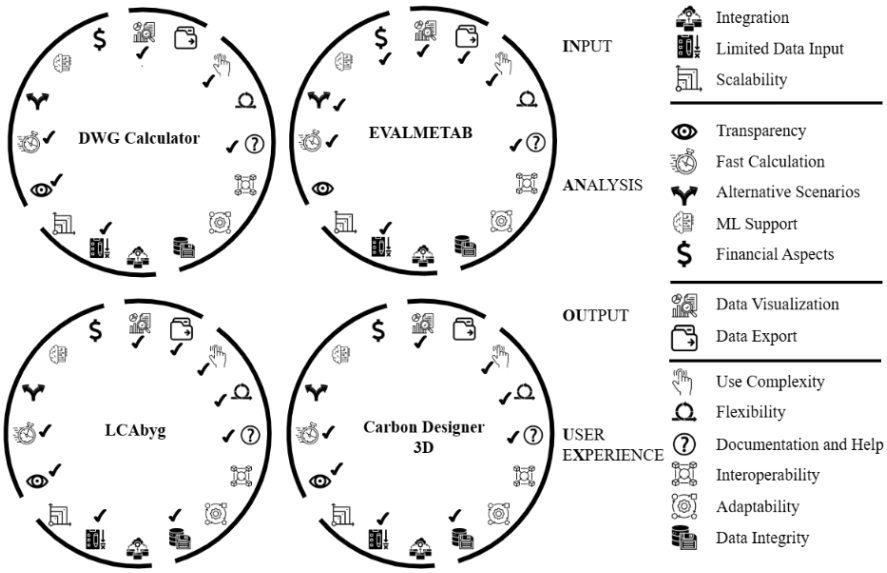


Figure 3. Alignment of identified software tools with user-centered requirements

capabilities. Although the reviewed tools provided some knowledge of the underlying calculation mechanisms, the information is neither comprehensive nor complete. This lack of transparency hinders user acceptance, as potential users may not fully understand the results [22]. The reviewed tools also have limited capabilities for performing financial analyses of alternative demolition techniques and disposal choices, and the absence of these affects their usefulness in decision-making. The lack of these functionalities in Carbon Designer 3D and LCAbyg tools is understandable as they are not primarily designed for material inventorying.

Currently, the STs for circular economy possess stand-alone functionalities requiring the use of several tools to accomplish diverse tasks as a project advances. As the data requirements for circularity applications are not yet standardized [8], the variety of digital technologies used by stakeholders can result in interoperability challenges. Accommodating different functionalities under the umbrella of a single tool or establishing data standards for data sharing among different tools can solve adaptability and interoperability challenges.

Most digital tools in the CE environment such as OneClick LCA and GaBi Circularity Toolkit, focus on LCA component during the early design stages, as the decision-making strategy [18],[23]. The research on software tools for CE has focused on developing digital marketplaces (DMP), data management, and material passport generation tools such as Madaster [24], Cirdax [25], and Loopfront **Error! Reference source not found.** require material information to be entered manually and are useful only once the material inventorying is completed.

As an industry known for the slow adoption and the workforce's inherent resistance to modern technologies, the shift to circularity is further complicated by high implementation costs, lack of trained personnel, and difficulty in identifying available tools [27]. To address these issues STs should be robust, effective, comprehensive, and meet the actual needs of CE stakeholders. Future tools developed for the circular economy should consider these requirements and aim to bridge the gaps in the existing STs to improve their quality.

5 Conclusion

The results indicate a consensus among the industry stakeholders on the potential of mining buildings for secondary resources and the vital role of digital technologies in driving circularity. However, a gap remains in the availability of STs for rapid, accurate, and affordable material inventorying of existing buildings.

This paper derived a literature-based framework on user-centered design requirements to guide the development of software tool that performs fast quantity take-off of built facilities to support a circular economy. Based on this review, an evaluation framework with criteria for user-friendly software tools in circular economy applications was developed and used to determine the adequacy of four identified commercial tools for automated inventorying. The development of this key-criteria is a valuable step toward designing new STs for CE in built environment.

This study is limited to the user-centered requirements derived from published literature, and the software tools were evaluated based on these criteria,

highlighting only theoretical requirements. Consequently, further research will focus on identifying user-centered design requirements and technical functionalities considered critical and feasible by industry stakeholders, through surveys and interviews. Based on these insights, the feasibility of implementing the prioritized criteria in a software tool for fast QTO of built facilities will be investigated.

References

[1] CCME, 2019. Guide for Identifying, Evaluating and Selecting Policies for Influencing Construction, Renovation and Demolition Waste Management. Online: https://ccme.ca/en/res/crdguidance-secured.pdf, Accessed: 15/12/2024

[2] F. Ossio, C. Salinas, and H. Hernández. Circular economy in the built environment: A systematic literature review and definition of the circular construction concept. *Journal of Cleaner Production*, 414(1): 137738, 2023.

[3] G. Cai and D. Waldmann. A material and component bank to facilitate material recycling and component reuse for a sustainable construction: concept and preliminary study. *Clean Technologies and Environmental Policy*, 21(10): 2015–2032, 2019.

[4] A. S. Allam, R. O. Panizza, and M. Nik-Bakht. A SWOT Analysis for Deconstruction of the Canadian Built Environment. *In Proceedings of the Canadian Society for Civil Engineering Annual Conference*, Vol.4, pages 287–299, Moncton, Canada, 2023.

[5] C. M. Rose and J. A. Stegemann. Characterising existing buildings as material banks (E-BAMB) to enable component reuse. *Proceedings of the Institution of Civil Engineers - Engineering Sustainability*, 172(3):129–140, 2019.

[6] N. Francart, S. R. B. Gummidi, E. Hoxha, and H. Birgisdottir. A Danish model of building macro-components to promote circularity. *Journal of Physics: Conference Series*, 2600(19): 192001, 2023.

[7] O. Rexfelt and A. Selvefors. Mapping the landscape of circular design tools. *Resources, Conservation & Recycling*, 209:107783, 2024.

[8] S. Çetin, V. Gruis, and A. Straub. Digitalization for a circular economy in the building industry: Multiple-case study of Dutch social housing organizations. *Resources, Conservation & Recycling Advances*, 15:200110, 2022.

[9] Y. Yu, D. M. Yazan, V. Junjan, and M.-E. Iacob. Circular economy in the construction industry: A review of decision support tools based on Information & Communication Technologies. *Journal of Cleaner Production*, 349:131335, 2022.

[10] E. Meex, E. Knapen, and G. Verbeeck. A framework to evaluate the architect-friendliness of environmental impact assessment tools for buildings. *In 35th International Conference on Education and Research in Computer Aided Architectural Design*, pages 425-434, Rome, Italy, 2017.

[11] E. Meex, A. Hollberg, E. Knapen, L. Hildebrand, and G. Verbeeck. Requirements for applying LCA-based environmental impact assessment tools in the early stages of building design. *Building and Environment*, 133:228–236, 2018.

[12] A. Hollberg, M. Tjäder, G. Ingelhag, and H. Wallbaum. A Framework for User Centric LCA Tool Development for Early Planning Stages of Buildings. *Frontiers in Built Environment*, 8: 744946, 2022.

[13] O. O. Akinade et al. BIM-based deconstruction tool: Towards essential functionalities. International *Journal of Sustainable Built Environment*, 6(1): 260–271, 2017.

[14] N. E. Kobylinska, D. Raghu, M. Gordon, J. Hunhevicz, and C. D. Wolf. Predicting recoverable material stock in buildings: using machine learning with pre-demolition audit data as a case study. *In Proceedings of 2023 European Conference on Computing in Construction - 40th International CIB W78 Conference*, Heraklion, Crete, Greece, 2023.

[15] DWGR, Metro Vancouver. Demolition Waste Generation Rates Calculator | Metro Vancouver. Online: https://metrovancouver.org:443/services/solid-waste/demolition-waste-generation-rates-calculator, Accessed: 31/08/2024.

[16] Chaire Economic circulaire et metabolisme urbain. EVALMETAB. Online: https://www.evalmetab. fr/, Accessed: Sep. 27, 2024.

[17] G. Suciu, I. Petre, and C. Stalidi. Urban transformation processes and material stock techniques to facilitate circular economy and urban resilience. *In Proceedings of 15th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, pages 1-5, Bucharest, Romania, 2023.

[18] A. Dervishaj and K. Gudmundsson. From LCA to circular design: A comparative study of digital tools for the built environment. *Resources, Conservation & Recycling*, 200:107291, 2024.

[19] Carbon Designer 3D. Early embodied carbon & whole life carbon optimization | One Click LCA. Online: https://oneclicklca.com/software/design-construction/early-design-decarbonization, Accessed: Aug. 31, 2024.

[20] K. Kanafani, A. Garnow, R. Zimmermann, C. Sørensen, E. Stapel, and H. Birgisdottir. Automated

Life cycle inventories for existing buildings – a parametric reference model approach. *In Proceedings of sbe22- Built Environment within Planetary Boundaries*, 1078(1):012097, Berlin, Germany, 2022.

[21] LCAbyg. LCAbyg.dk | Home. Online: https://www.lcabyg.dk/en/, Accessed: Aug. 31, 2024.

[22] D. Dharmarathna, V. Bunster, and P. Graham. A review of building life cycle assessment software tools: Challenges and future directions. *In Proceedings of World Sustainable Built Environment Conference 2024*, 1363(1): 012063, virtual, 2024.

[23] S. Muñoz, M. R. Hosseini, and R. H. Crawford, "Circular economy software tools at the material and product level: A scoping review," *In Proceedings of the 55th International Conference of Architectural Science Association*, pages 169-178, Perth, Australia, 2022.

[24] Madaster Global. Our platform. Online: https://madaster.com/platform/, Accessed: Oct. 01, 2024.

[25] Cirdax. Cirdax Real Estate Digitalization. Online: https://www.cirdax.com/, Accessed: Nov. 18, 2024.

[26] Loopfront. Product - Loopfront. Online: https://www.loopfront.com/product, Accessed: Nov. 18, 2024

[27] S. Thirumal, N. Udawatta, G. Karunasena, and R. Al-Ameri. Barriers to Adopting Digital Technologies to Implement Circular Economy Practices in the Construction Industry: A Systematic Literature Review. *Sustainability*, vol. 16(8): 3185, 2024.

Feasibility of Reusing Prefabricated Panels from an Embodied Energy and Economics Perspective

Mikael Gilbert[1], Khalegh Barati[1] and Xuesong Shen[1]

[1]School of Civil and Environmental Engineering, The University of New South Wales (UNSW), Sydney, NSW 2052, Australia
mikael.gilbert@student.unsw.edu, khalegh.barati@unsw.edu.au, x.shen@unsw.edu.au

Abstract -

The prevalence of prefabrication in the modern construction industry has highlighted the importance of sustainable management of these new construction materials. Despite reduced emissions through minimizing construction time and production waste, at the end of the building's life, the amount of construction waste produced remains constant. This paper aims to investigate the environmental and economical sustainability impacts from reusing prefabricated concrete and cross laminated timber panels sourced from buildings reaching the end of its service life. The main framework used focuses on applying a lifecycle assessment in terms of embodied energy to a unit of the selected prefabricated material. The results from this analysis are then further expanded to include varying distances between the donor and host building, allowing for a comparison between the embodied energy saved and additional energy consumption from transporting the construction materials. Furthermore, the economic sustainability of this construction practice are calculated. A case study was conducted using a 3D BIM model of a residential structure whereby the materials are quantified and paired with data obtained from the LCA. The results have shown that there are energy savings of 36,628.6 MJ/m^3 and 66.6 MJ/m^3 for concrete and cross laminated timber panels, respectively. Additionally, a trendline showing the correlation between distance and embodied energy savings for both materials were observed. The research contributes to providing alternatives for sustainable waste management and material selection policy in the construction industry.

Keywords –
Embodied Energy, Life Cycle Assessment, Modular Construction, Prefabrication, Sustainability

1 Introduction

As climate change becomes a more significant challenge, it is in the best interest of humanity to combat this phenomenon to minimize the impact of climate change on the global ecosystem. The World Commission on Environment and Development (WCED) defines sustainability as the preservation of the current world for future generations [1]. The construction industry is named one of the major contributors to climate change because of its large greenhouse gas emissions. This is largely caused by an increase in demand for infrastructure, housing and resources as the human population continues to grow and modernize, which ultimately results in the expansion of the construction industry. One of the construction methods that has been implemented to keep up with demand for infrastructure is through prefabrication, which has been proven to be faster to build and more environmentally friendly [2].

A building's lifecycle starts in the pre-building stage where the building is planned and designed, this is the best stage to include any sustainable design principles into the project as nothing physical has been built and can be easily altered as needed. This stage is then followed by the construction stage, where the building is constructed and consequently by the operational stage where the building is used for its intended purpose. Once the building has reached the end of its service life, the building will be demolished or deconstructed to make way for a new building to better suit the needs of society at the time [3]. A major issue of this linear lifecycle is that there will be a large amount of waste produced. Improper demolition waste management reduces reuse potential as the waste will head straight to landfill. Reusing and recycling waste obtained from all the stages of a building's lifecycle ensures the minimization of waste that goes to landfill and thus reduces the harvest of new raw material.

Studies on prefabrication have shown that with prefabricated products, construction waste and time can both be reduced, leading to more sustainable practices [2,4]. Despite these established benefits, there seems to be limited knowledge of methods in which the prefabricated products can be harvested and reused as part of a new building's structure. A major downside to

current industrial practices is that significant additional processing energy will be needed, such as when steel is melted and reused [5]. However, reusing already made prefabricated products circumnavigates this problem as these products are relatively ready to use. It appears that a gap in the research exists in finding the benefits of reusing prefabricated materials salvaged from a building's end-of-life stages. Surveys have shown that despite poor public perception of second-hand structural products, this reusing technique may be necessary to ensure responsible use of the Earth's finite resources and promote a sustainable construction industry moving forward.

This paper aims to investigate the environmental and economic feasibility of reusing prefabricated concrete and cross laminated timber panels sourced from buildings reaching the end of its service life. The paper focuses on the comparison of the embodied energy and cost savings from reusing these materials against common transportation methods. In this research, the benefits as well as the limitations of reusing prefabricated concrete and CLT panels were quantified. The benefits to be analyzed will revolve around the two of the three pillars of sustainability, namely environmental, and economic sustainability.

2 Literature Review

Sustainable development is "a development that meets the needs of the present generation without compromising the ability of future generations to meet their own needs" [1]. This concept is commonly separated into three main categories namely, economic, environmental, and social sustainability [6]. When all three sustainability categories are achieved, then total sustainability has been attained.

There are three basic principles that can serve as a basis for sustainable construction such as resource management, life-cycle design, and human design [7]. Resource management involves the efficient use of the Earth's finite resources. This can be done by implementing renewable energy sources, reducing construction wastes, etc. For example, the design of a building's façade can be optimized to reduce its operational energy consumption [8]. Additionally, human design involves both biophysical and social part of the project in which the requirements of the building does not exceed the capacity of the environment to fulfil those needs [7]. This design is usually implemented in the design stage of the building's lifecycle [9]. In life-cycle design, all aspects from the design up to the demolition is to be considered in the planning stage [3].

The promotion of sustainability in the construction industry can also be achieved through optimal material selection. Being one of the most used materials in the construction industry, concrete plays a key role which greatly affects the sustainability of the industry. Another method in which sustainability of concrete can be improved is through the incorporation of waste in less demanding work [10]. For example, an Egyptian case study employed rice stalks can be used as cement fillers for low-cost housing [11]. Another approach in material selection is with the use timber in the form of Cross Laminated Timber (CLT) as a renewable source of material, which has been growing in popularity. The structure of CLTs involves planks which are sawn, glued, and layered wood where the layers are placed perpendicular to one another [12]. The main advantage of CLT is that it is made from timber, a renewable source of material, unlike concrete and steel which are both non-renewable [13].

Prefabrication has also been closely linked to a more sustainable construction practice. In general, there are two types of prefabrication namely 2D and 3D prefabrication [14]. There are both advantages and drawbacks of prefabrication in the construction industry. Advantages of prefabrication includes a reduction in construction time leading to cost savings, lower margin of error and that the modules can be reused or altered as needed [2,4]. When compared to conventional construction, prefabrication has also been shown to reduce the amount of waste by 81.3% and 83.2% for small and large structures, respectively [15]. A case study conducted in Hong Kong has shown that prefabrication can reduce construction waste by 65%, labour requirements by 16%, and construction time by 15% [16]. However, there are also drawbacks such as the need for specialized skill and equipment during assembly, the lack of product variety, etc. Another key drawback is that when an error occurs either during production or assembly, there may be major delays to the project [2,4].

Despite of these advantages during construction, prefabricated buildings are still being decommissioned in a linear manner. Identifying and prioritising the end-of-life solutions for prefabricated buildings are essential to reduce carbon footprint and promote circularity in construction [17]. Research has shown that prefabricated buildings have higher recyclability rates compared to traditional buildings, which drives the possibility of reusing these components in future projects with minimal processing [18]. With regards to these recycled components, BIM has been shown to aid in the deconstruction process by analysing existing conditions, labelling reusable elements and simulating the deconstruction process [19].

The concept of reusing prefabricated components has been applied to multiple projects across the world. One of the concepts is the Pilot Houses in Berlin developed by TU Berlin. In this project, prestressed concrete slabs of type WBS70 were modified by sawing them into the

required sizes and then connected to form the external walls of the houses [20]. The façade is then covered with an External Thermal Insulation Composite (ETIC) system. A case study house in Schidlow, Berlin has achieved 245m^3 of concrete saved with a significant reduction in construction time, which has been shown to reduce greenhouse gas emissions such as CO_2, N_2O and CH_4 [20,21]. Other examples can be seen in the BedZED development in London and the UN17 Village in Copenhagen where it successfully implements the basic principles of sustainability. These projects also adopted the practice of reusing second-hand materials as part of an efficient resource management policy and design for human strategy.

As the use of prefabricated components become more widely accepted, steps should be taken to consider the end-of-life stage of these parts [17]. Improper disposal strategies may lead to unsustainable construction techniques which may accelerate the deplet ion of natural resources. Reusing these components may be one of the ways for the industry to process and control the amount of waste being produced. Additionally, it may lower construction costs and preserve finite resources.

3 Research Methodology

In this research, a method known as Life Cycle Assessment (LCA) will be used to measure the benefits brought on by reusing prefabricated concrete and CLT panels to construct new buildings. There are three different methods for LCA. The first of these methods is called Process Analysis where a product is identified and the inputs that go into production of the product is analysed. The second method is called the Input-Output (I/O) analysis where the energy usage of a particular industry in a country is obtained from national survey data. These two methods can be combined to form a "hybrid" LCA to have the simplicity of I/O analysis and the accuracy of process analysis.

A detailed framework of the method used in this paper can be seen in Figure 1. It highlights the major components starting from the LCA, then followed by both environmental and economical sustainability studies. Finally, to show how this method can be applied, a case study involving a 3D BIM model will be conducted. The framework aims to provide a generic plan in which all materials can be analysed and then applied to various case studies.

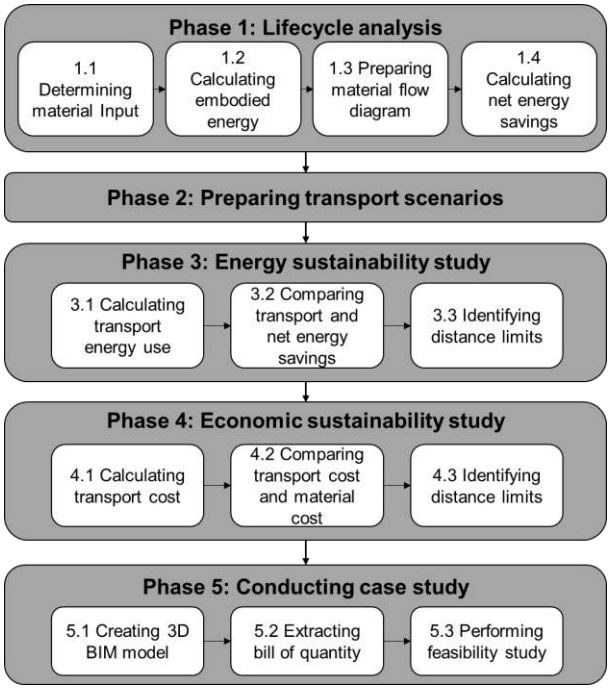


Figure 1. Research framework to quantify feasibility of reusing imported prefabricated components

Firstly, the unit of system processes will be obtained from sources such as research papers, industry examples, databases, etc. The data collection process is followed by the data calculation stage. In this stage, the data obtained is calculated according to the amount of material used. Then, these values are arranged according to the appropriate flow of processes when reusing prefabricated components. The flow will then be based on the law of conservation of energy where the sum of the energy entering a process must be equal to the sum of the energy leaving the process. This stage in the LCA allows for the verification of data energy used throughout the process. This iterative process ensures that the inputs and outputs can be maximally captured to ensure a high-quality and accurate assessment.

In this LCA assessment, data such as the materials being used and their respective processes are first gathered. The respective processes such as transport, assembly and panel cutting will be added onto the analysis. The framework used in this LCA will begin from the material harvesting up to the assembly of the model houses. The maintenance and operation as well as the end-of-life stage of the buildings will not be included in the analysis as it is assumed that the digital model will have similar operational conditions as its real-life counterpart if it is built using new raw materials.

In order to evaluate the framework, a sensitivity analysis is conducted. The aim of the sensitivity analysis is to establish which processes will impart the largest effect on the overall LCA flow diagram. This is done by

altering parts of the processes and assessing which of these changes results in a large effect on the overall embodied energy. Sensitivity analysis can then be used to determine which processes to include and exclude from the system boundary. In this study, the sensitivity analysis will revolve around the transportation aspect of the reuse process. This is done by changing the location of the donor building and changing the method of transport accordingly. These scenarios can help determine the limit in which the proposed construction technique is no longer sustainable. By changing the location of the donor building, changes in the feasibility of the project can be seen. Using these changes, trends can be inferred and thus provide a quantitative value to measure the sustainability of reusing these prefabricated panels. This value allows for this technique to be applied to real-life projects and have their sustainability measured.

In Phase 5, a real-life building will be modelled using Revit and then a bill of quantities for the materials used in the house is gathered. The building being modelled draws inspiration from the house mentioned in [20], where the external wall of the house is made from reused concrete panels. In addition to the external walls, the internal walls will also be considered. In this study, it will be assumed that the internal walls will be made from CLT panels. Then, the energy and cost savings value obtained in the LCA can be used to determine the sustainability of the project. This section of the paper is used to verify the framework set out in Figure 1. It shows how the framework can be applied to other projects as the data produced from the LCA is per unit volume.

Finally, an interpretation of the LCA results will be conducted, which includes the verification of the results. This section will highlight the potential limitations of this study and ways in which these limitations can be overcome. It will also show paths in which future research can take to further develop this construction concept and provide a better understanding of material and resource sustainability in the construction industry.

4 Reusing Concrete and CLT Panels

The LCA conducted in this study involves 1 m^3 of reused concrete and CLT panels, where both the donor and host buildings are both located in Sydney. This will include processes starting from the harvesting of these materials up to the assembly of the materials. In this lifecycle analysis, several assumptions will be made. For example, panels harvested from the donor building will be in the exact measurements required by the host building and thus no additional processing of the panels is required. Next, as materials are being reused, the embodied energy of the material itself will no longer be included in the LCA as it has already been included in the LCA of the donor building. Finally, the energy used to assemble 1 m^3 of each material is sourced from [24], where the total energy used for onsite construction is divided by the total volume of material used during construction. This results in the energy consumed for assembly of 294.2 MJ/m^3.

The LCA results suggests that reusing concrete panels requires an additional 1,032.8 MJ/m^3. It is also worth noting that 71.3% of the energy consumed arises from transport which is followed by the energy consumed when assembling of the panels (28.5%). When factoring in the embodied energy of the raw material itself, energy will still be saved. This is because no new material will be required and thus there will be no energy expenditure from the material aspect. The embodied energy of 1 m^3 of tilt up concrete panel which constitutes 95% concrete and 5% steel has an embodied energy of 37,661.4 MJ. Therefore, it can be concluded that the net energy savings for reusing concrete panels is 36,628.6 MJ/m^3.

On the other hand, reusing CLT panels requires an additional energy consumption of 433.4 MJ. In the case of CLT panels, most of the energy consumed comes from the assembly process. It constitutes 67.9% of the additional energy required and then it is followed by transport which contributes to 31.8% of the energy consumed. The embodied energy of 1 m^3 of CLT panels is 5,000 MJ [17]. As there is no new CLT panels being used, the embodied energy of CLT can be considered as energy saved. Thus, it can be inferred that the net energy savings for reusing 1m^3 of CLT panels is 4,566.6 MJ/m^3.

The next section of the journal paper revolves around a cost-benefit analysis of this construction method to fulfil the economical sustainability aspect of this construction technique. The analysis adds eight scenarios in addition to the base scenario previously mentioned. The additional scenarios alter the location of the donor building, leading to increased transportation cost and its environmental impact while keeping the host building in Sydney. The location of the donor building is comprised of cities in Asia, Europe, and North America. The list of scenarios and their respective travel distances can be seen in Table 1. The total distance travelled is the sum of the travel distances of all modes of transport namely, truck, rail, and ships, used during the panel's transportation process. Truck, rail, and ships were selected as the preferred mode of transportation as they are representative of how most construction materials are transported from the factories to the construction site.

Table 1. Scenarios and their respective travel distances

Source	Distance (km)			
	Truck	Rail	Ship	Total
Sydney	120	0	0	120
Perth	182	3,290	0	3,472
Jakarta	170	0	7,119	7,289
Tokyo	150	0	8,043	8,193
Shanghai	270	0	8,579	8,849
San Francisco	70	0	11,957	12,027
London	105	0	21,352	21,457
Stockholm	110	0	23,230	23,340
Moscow	90	631	23,728	24,449

In Table 1, the scenario where Perth is the location of the donor building, Rail is mainly used to transport these panels over the Australian component as it is deemed more efficient than trucks. For the Moscow case, Rail was used to transport the panels from landlocked Moscow to St. Petersburg to be loaded onto ships. This material transfer is assumed to be done by rail. The distances for ships are calculated using distances of typical shipping routes between the locations.

The first part of the result analysis involves comparing the energy consumed during transport of the materials and the energy saved from reusing the material itself. The energy consumption for trucks, rail and ships are 4.6 MJ, 0.9 MJ and 0.4 MJ per ntkm (net-tonne kilometre), respectively [23]. This is done for all scenarios and both concrete and CLT panels and the results are shown in Figure 2. The threshold whereby the transport energy exceeds the saved embodied energy is also included in this figure.

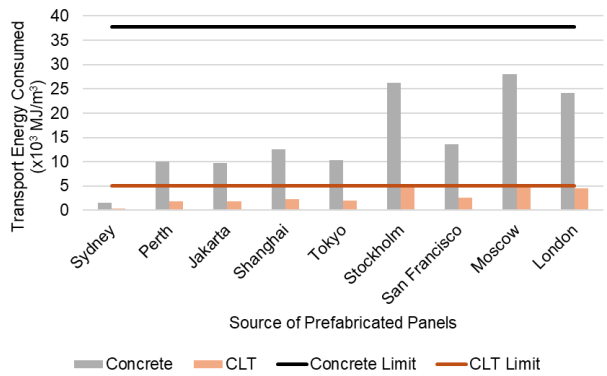


Figure 2. Energy consumed by transport for 1m^3 of (a) concrete panels (b) CLT panels for various locations

When energy consumed by transport (E_t) is plotted against the distance travelled (d) by the material, a linear trendline can be seen. The trendline equations for both concrete panels and CLT panels can be seen in Equations (1) and (2), respectively.

$$E_{t,concrete}\ (MJ/m^3) = 0.989d + 3119.7 \quad (1)$$

$$E_{t,CLT}(MJ/m^3) = 0.185d + 584.2 \quad (2)$$

The maximum distance that a material can be sourced from can be calculated by replacing the energy consumed by transport by the embodied energy of the material being reused including the processes required to reuse them while excluding the energy required for transport. The resulting maximum distance calculated from an environmental perspective (D_{env}) are 34,180.79 km and 22,272.43 km for concrete and CLT, respectively.

The next stage of the cost benefit analysis revolves around comparing the cost of transport and the cost savings brought on by reusing the materials. The price of the material in Australia can be assumed as AU$302 per m^2 of concrete panels and AU$863.32 per m^2 of CLT panels. Assuming the exchange rate for AU$1 is US$0.62, the price of concrete and CLT are US$187.24 per m^2 and $535.26 per m^2, respectively [23]. It is also assumed that the price for material transport is US$0.052, US$0.024, and US$0.018 per ntkm for road, rail, and ships, respectively [25]. Another assumption made in the cost analysis is that the thickness of the concrete panel is 200 mm and the thickness of the CLT panels is 135 mm.

A limit for the distance for the material source from an economic standpoint can be obtained. Table 2 below shows the distance limits for both concrete and CLT panels. From an economical perspective, the maximum distance limit for concrete is 19,180.03 km. This is obtained in the case when the concrete panels are obtained for free or in other words 100% discount. It is also shown that CLT panels can be sourced from all over the world and still be economically sustainable.

Table 2. Distance limit in kilometres depending on material discounts for concrete and CLT panels from an economical perspective

Percentage of Original Cost	Distance Limit (D_{econ}) (km)	
	Concrete Panels	CLT Panels
0%	19,180.03	314,542.25
20%	15,211.30	251,500.00
40%	11,242.84	188,457.75
60%	7,274.11	125,415.49
80%	3,305.39	62,376.06
100%	-663.39	-665.01

The maximum distance between the source of reused materials and the construction site can be obtained from both an environmental and economical perspective. Selecting the larger value from the two distances produces a comprehensive table with the de facto distance limitation. The equation used to decide on which distance to choose is shown below in Equation (3). Where D_{env} and D_{econ} are the distance limit from an

environmental and economical perspective, respectively. i is the material that is being analysed and r is the percentage of original cost for the selected material i.

$$D_{i,r} = \min\left(\frac{D_{env,i} + |D_{env,i}|}{2}, \frac{D_{econ,i,r} + |D_{econ,i,r}|}{2}\right) \quad (3)$$

Table 3 shows the distance limit of both economic and environmental sustainability of this construction technique. It is shown that the sustainability of reusing CLT panels is governed by the environmental limitations of the material. While concrete panels are governed by its economics. It is suggested that as the materials density increase, the main barrier for sustainability shifts from being an embodied energy to an economic limitation.

Table 3. Distance limit in kilometres depending on material discounts for concrete and CLT panels from both environmental and economical perspectives

Percentage of Original Cost (r)	Distance Limit (D) (km)	
	Concrete Panels	CLT Panels
0%	19,180.03	22,272.43
20%	15,211.30	22,272.43
40%	11,242.84	22,272.43
60%	7,274.11	22,272.43
80%	3,305.39	22,272.43
100%	0	0

Although this study shows that CLT panels are not environmentally sustainable when sourced from long distances, common industry practice has disregarded this fact. This is because of the significant cost savings that reusing CLT has, as shown in the cost analysis section of this paper. Due to the slim profit margins in the construction industry, cost savings takes priority over environmentally sustainable practices.

The values obtained from this section are unit values for both $1m^3$ of concrete and CLT panels. These values can aid designers in extrapolating necessary emission and cost saving calculations for various building types and applications.

5 Case Study

In this case study, an apartment building, located in Sydney, Australia, comprising of seven units was analysed. The apartment block is built on a site with an area of $686.95m^2$ and covers five levels ranging from a basement parking level, three levels for apartment units and a rooftop area. the 3D BIM model of the apartment block can be seen in Figures 3. This case study will serve as an example in which the framework developed in this paper is to be implemented in a real-life case study.



Figure 3. 3D BIM model of case apartment

During the analysis of this case study, the external walls are assumed to be made from the reused prefabricated concrete panels, and the internal walls are made of reused CLT panels. The external walls are supposed to be 200 mm thick, and the internal walls are 138 mm thick. The total volume of concrete and CLT panels used in the construction of the apartment are 246.53 m3 and 49.71 m3, respectively. Assuming the material is purchased at 20% of its new price. Table 4 below shows the sustainability of reusing prefabricated panels when implemented into the case study from both an economic and environmental point of view.

Table 4. Sustainability analysis of case building

Material	Concrete	CLT
Building Element	External Walls	Internal Walls
Sydney- Sydney		
Volume of Material (m^3)	246.53	49.71
Distance (km)	120	120
Energy Saved ($x10^6$ MJ)	9.03	0.23
Cost Savings (thousands of AUD)	291.2	177.8
Environmentally Sustainable?	✓	✓
Economically Feasible?	✓	✓
Stockholm – Sydney		
Volume of Material (m^3)	246.53	49.71
Distance Travelled (km)	23,339.64	23,339.64
Energy Saved ($x10^6$ MJ)	2.94	-0.23
Cost Savings (thousands of AUD)	-151.7	161.1
Environmentally Sustainable?	✓	×
Economically Feasible?	×	✓

The results from the case study shown in Table 4 shows that when the materials are sourced locally within Sydney, this reusing construction technique is both economical and environmentally sustainable. This is shown by the positive value for both cost and energy

savings. The sustainability benefits from the Sydney-to-Sydney case can be attributed to the distance between the donor and host building. Lower distance values lower the corresponding GHG emissions created through transport and minimizes the transport costs. However, this may not be the case for all location sources for these prefabricated materials. Some situations may result in a project being economically viable but not environmentally sustainable as shown in the case where CLT panels are sourced from Stockholm.

6 Conclusions

This research is set to answer the question of whether it is in the best interest of the construction industry to reuse prefabricated panels which can be harvested from buildings reaching the end of its useful life. As more new prefabrication parts are used in the construction industry, the significance of efficiently and sustainably handling a building's construction or demolition waste can no longer be ignored. This paper provides an alternative to project developers to adopt reusability into their design mandate. Additionally, it proves that reusing prefabricated materials has both economic and environmental benefits which can be used in favour of the developer. Lastly, a framework is created to quantify the distance limit in which prefabricated materials can be sourced before being either environmentally or economically unsustainable.

Based on the LCI flow diagram, this construction method has shown that significant energy savings can be achieved when prefabricated panels are reused. However, this energy savings decreases the further the material is sourced. Because there will be a rise in the energy consumption brought on from transporting the panels. Thus, from case studies involving different donor building locations, it is concluded that, environmentally, the maximum distance concrete panels can be sourced from is 34,925.85km and for CLT the distance is 23,869.14km. Apart from energy savings, there are also cost savings to be achieved with varying values. These values depend on the transportation method and cost of the reused panels when sourced from other projects. When factoring in the cost of the method, it is found that the distance in which concrete can be sourced is limited by the transportation cost which increases steeply due to the high density of concrete. In the other hand, the distance in which CLT panels can be sourced from is limited by the energy consumption of transport being higher than the energy saved from reusing the panels. This may result in situations where a product is environmentally sustainable but not economically feasible and vice versa.

The proposed framework provides designers and construction managers with a method to calculate the cost and embodied energy savings from reusing prefabricated panels. Reusing prefabricated panels may not be environmentally and economically sustainable in all cases and which highlights the need to quantify these savings to ensure that the project remains sustainable.

One of the main limitations of this paper is that the only items being analysed are the external and internal walls. When the entire bill of materials is considered for the whole building, the sustainability of the project may not be the same. Different materials have different embodied energies, and some materials may not be worth transporting from any location. Additionally, there are also limitations in terms of the regulations that each country may have regarding the import and export of second-hand construction products as well as different quality standards for these products. These obstacles may affect the viability of a project due to difficulties during the material procurement process. This can be overcome by creating more case scenarios where patterns can be recognized to produce an optimum source for these components. Future studies can the social impacts of using this construction technique, which includes the health impacts of the new construction technique towards the population living in the vicinity of the project. It can also include the inclusion of sustainability impact calculations as well as feasibility studies of using state-of-the-art transportation technologies.

References

[1] World Commission on Environment and Development. Our Common Future. World Commission on Environment and Development, Oxford University Press, Oxford, 1987.

[2] Y. Teng, K. Li, W. Pan, and T. Ng. Reducing building life cycle carbon emissions through prefabrication: Evidence from and gaps in empirical studies, *Building and Environment*, 132:125-136, 2018.

[3] A. Ginzburg. Sustainable Building Life Cycle Design. MATEC Web of Conferences, 73: 02018, 2016

[4] P. Behera, Prefabricated Structures & Prefabrication – Concept, Components & Advantages. On-line: https://civildigital.com/prefabricated-structures-prefabrication-concept-components-advantages-ppt/, Accessed 05/11/2024.

[5] T. Echterhof. Review on the Use of Alternative Carbon Sources in EAF Steelmaking. *Metals*, 11(2):222, 2021.

[6] B. Purvis, Y. Mao, and D. Robinson. Three Pillars of Sustainability: in search of conceptual origins. *Sustainability Science*, 14:681-695, 2018.

[7] A. Sev. How can the construction industry contribute to sustainable development? A conceptual

framework. *Sustainable Development*, 17(3):161-173, 2009

[8] M. Gilbert, K. Barati and X. Shen, Investigation of the Insulation Properties' Impact on the Thermal Load of Residential Buildings. In *Proceedings of the 46th Australasian Universities Building Education Association (AUBEA) 2023 Conference*, Auckland, New Zealand, 2023.

[9] H. Demirel and V. Duffy. A Sustainable Human Centered Design Framework Based on Human Factors. In *Proceedings of the 4th International conference on Digital Human Modelling and Applications in Health, Safety, Ergonomics, and Risk Management*, pages 307-315, Las Vegas, United States of America, 2013.

[10] E. Wong and R. Yip. Promoting sustainable construction waste management in Hong Kong. *Construction Management and Economics*, 22(6):563-566, 2004.

[11] A. Mansour, J. Srebric, and B. Burley. Development of Straw-cement Composite: Sustainable Building Material for Low-Cost Housing in Egypt. *Journal of Applied Sciences Research*, 3(11):1571-1580, 2007.

[12] E. Souza. Cross Laminated Timber (CLT): What It Is and How to Use It. On-line: https://www.archdaily.com/893442/cross-laminated-timber-clt-what-it-is-and-how-to-use-it, Accessed: 27/05/2022.

[13] M. Mallo and O. Espinoza. Outlook for Cross-Laminated Timber in the United States. *BioResources*, 9(4):7427-7443, 2014.

[14] PrefabAUS, What Is Prefab — prefabAUS. On-line: https://www.prefabaus.org.au/what-is-prefab#:~:text=Types%20of%20Prefab,conjunction%20with%20traditional%20construction%20approaches, Accessed: 01/03/2021.

[15] L. Loizou, K. Barati, X. Shen and B. Li, Quantifying Advantages of Modular Construction: Waste Generation. *Buildings*, 11(622), 2021.

[16] L. Jaillon and C. Poon, Sustainable construction aspects of using prefabrication in dense urban environment: a Hong Kong case study. *Construction Management and Economics*, 26:953-966, 2008.

[17] S. Samarakoon and R. Ratnayake, End-of-life solution prioritization for pre-cast concrete components aligning with circular economy targets. *Structural Concrete*, 24(2): 1751-1765, 2022.

[18] H. Wang, Y. Zhang, W. Gao, and S. Kuroki, Life Cycle Environmental and Cost Performance of Prefabricated Buildings. *Sustainability*, 12 (2609), 2020.

[19] M. van den Berg, H. Voordijk, A. Adriaanse, BIM uses for deconstruction: an activity-theoretical perspective on reorganising end-of-life practices, *Construction Management and Economics*, 39(4): pages 323-339, 2021.

[20] C. Asam, Recycling Prefabricated Concrete Components – A contribution to Sustainable Construction. 2008.

[21] M. Gilbert, K. Barati and X. Shen, Reusing Prefabricated Concrete Panels: A Sustainability Approach. In *Proceedings of the World Building Congress*, Melbourne, Australia, 2022.

[22] N. Mehrvarz, K. Barati, and X. Shen, Embodied Energy Modelling of Modular Residential Projects Using BIM. In *Proceedings of the 38th International Symposium on Automation and Robotics in Construction*, pages 483-490, 2021.

[23] M. Lenzen, Total requirements of energy and greenhouse gases for Australian transport, Transportation Research Part D4, 1999.

[24] Reserve Bank of Australia, Exchange Rates. On-line: https://www.rba.gov.au/statistics/frequency/exchange-rates.html, Accessed 11/01/2025.

[25] D. Gargett, Freight Rates in Australia: Information Sheet, Australian Government: Department of Infrastructure and Regional Development: Bureau of Infrastructure, Transport and Regional Economics. 2017.

How good are the best practices in carbon reduction? UK highway sector case studies

Jinying Xu[1] and Kristen MacAskill[1]

[1]Department of Engineering, University of Cambridge, United Kingdom.
jx314@cam.ac.uk, kam71@cam.ac.uk

Abstract –

To achieve net-zero targets in the transportation infrastructure sector, cost-effective carbon reduction measures are essential. Despite increasing efforts to explore such measures, their cost-effectiveness remains underreported and insufficiently studied. A database of best practices with 1,046 entries from UK National Highways is extracted to investigate the cost-effectiveness of its carbon reduction strategies. The analysis identifies "Project and programme management" as the primary focus area, with the construction stage being the most emphasized stage. Four "Carbon Quick Wins" approaches and the carbon reduction hierarchy (Avoid-Reduce-Switch) are the main strategies employed. Although "Avoid" has the greatest reduction impact and the highest hierarchy in decision-making, it is less frequently reported in the register of best practices than "Reduce" and "Switch" to low carbon material and energy. Using data mining techniques, the cost and carbon reduction data are extracted to evaluate the cost-effectiveness of carbon reduction using marginal abatement cost (MAC) indicator. The MACs of eighteen cases demonstrate significant variations, complicating the performance evaluation of these best practices. The study reveals a lack of data in most practices to measure the cost-effectiveness of carbon reduction initiatives. Even with available data, its granularity is rough, potentially influencing the selection of appropriate approaches. Standardized methods to measure the cost and wider impact of carbon reduction initiatives are recommended for future research along with data reporting policies.

Keywords –

Highway infrastructure; Carbon reduction; Carbon-cost effectiveness; Best practices; UK

1 Introduction

The transportation sector, particularly highways, is a significant contributor to carbon emissions, a primary driver of climate change [1]. In response, the transportation sector in different countries has implemented various strategies aimed at reducing these emissions to meet both national and international environmental targets. For example, UK National Highways (NH) aims to achieve net-zero corporate emissions by 2030, net-zero maintenance and construction emissions by 2040, and net-zero road user emissions by 2050 [2].Their carbon reduction strategies include avoiding high-emission activities and materials; switching to low-carbon materials and vehicles, and sustainable energy; and improving design and construction approaches, and road infrastructure services to support sustainable transportation. Evaluating the effectiveness of these strategies is crucial for selecting and prioritizing carbon reduction methods, ensuring cost-effectiveness, and accelerating progress toward net-zero targets. Other sectors and countries can also benefit by learning from the most effective practices, contributing to collective efforts in combating climate change.

In the highway sector, numerous efforts have been made to reduce carbon emissions, ranging from quick wins to more comprehensive long-term strategies. Quick wins, such as the adoption of energy-efficient technologies and low-emission vehicles, have shown immediate benefits as evidenced in the NH 2022-2023 annual progress report [3]. However, many initiatives focus solely on carbon mitigation without considering long-term costs, leading to sustainability challenges over time. Additionally, some promising strategies such as building more charging facilities are stalled due to budget constraints and await adequate funding to be implemented. Despite these efforts, overall progress in carbon reduction within the highway sector has been slow, and the effectiveness of these strategies remains largely unassessed. This presents a significant practice problem: without a thorough evaluation of both immediate and long-term impacts, and without proving the value-for-money, it is challenging to prioritize and implement the most effective carbon reduction measures.

Existing research has extensively explored the effectiveness of carbon reduction strategies in transportation infrastructure at various scales, comparing different design alternatives, scenarios, materials,

equipment, and measures within infrastructure projects. Studies have analyzed a wide range of scenarios related to road construction and maintenance, considering factors such as pavements, materials, machinery, transport distance, energy sources, and land-use change, to identify optimal emission reduction alternatives. It has been highlighted that the most effective strategies include using low-carbon and recycled materials during the construction stage and adopting sustainable energy sources for off-road machinery during the maintenance stage [4]. Similarly, a life-cycle analysis of highway reconstruction projects studied the contribution of materials, construction equipment, resource mobilization, traffic disruption, and life-cycle maintenance to whole project carbon emissions, finding that material choice is a major factor in reducing GHG emissions, while equipment contributes minimally [5]. Another study compared the GHG emission reduction potential of ten innovative techniques in road infrastructure construction and maintenance in the Netherlands, recommending the switch to renewable energy for lighting installations, the use of low-energy asphalt, and the application of rejuvenators to achieve significant reductions [6]. However, these studies, focusing on the effectiveness of the strategies, have not provided quantifiable and comparable evaluation of the cost effectiveness, revealing a significant gap in the current literature.

This paper sets out to answer two research questions: (a) What are the latest key carbon reduction strategies in the highway sector? and (b) How efficient are these carbon reduction strategies? To address these questions, a case study database with a total of 1,046 entries from the UK NH is analyzed. Text mining techniques are used to extract cost and carbon data from the documents. Quantitative analyses were conducted where quantitative data are available to summarize the key carbon reduction strategies and the cost effectiveness of carbon reduction.

The rest of the paper is organized as follows. The Methods section provides a description of the dataset, data collection and preprocessing procedures, and the analytical methods used. The Results section reports the findings of the thematic analysis, content analysis, cost-benefit analysis. The Discussion section interprets the results, discusses the implications for policy and practice in the highway sector in the UK and beyond, and provides some limitations of the study and suggestions for future research. Finally, the Conclusions section summarizes the study and its key findings and implications for the transportation sector.

2 Methods

The UK NH operates a "Lessons learnt reporting" scheme, where its staff and supply chain can share their experiences. The "National Highways Carbon Hub" is an internal database for the sharing of lessons learnt regarding net-zero tools and resources (including case studies, webinars, and other learning materials). New case studies and resources are continually added to the database. As of the data extraction date (27 June 2024), there are 1,046 entries, with the first one added on 30 November 2017. These entries include basic information of each case, including its name, main category and subcategory, relevant Project Control Framework (PCF) Stage covered, and the date it was added. Additionally, they include the "Lessons Learned Report" as an appendix. With permission granted to access and download these documents, the researchers can extract data for analysis.

Due to the vast volume of documents, manually extracting data is impractical. Data mining techniques are applied to extract the desired information from the documents using Python. The basic process involves converting documents of different formats (e.g., .pdf, .docx, .pptx, .jpg, .xlsx, and .mp4) to machine-readable text formats, then using natural language processing techniques to identify and extract specific cost and carbon data. By automating and iterating this process for all documents, the cost and carbon data are extracted and stored in a structured format (e.g., .csv) file for further analysis. The "pytesseract" module is used to identify and extract specific cost and carbon data. Out of the 1,046 entries, 252 have cost data and 78 have carbon data, with only 46 entries containing both. A follow-up screening ensures the relevance (whether the cost data and carbon data are about the savings of carbon and the cost associated) of the extracted data, resulting in 18 entries with data regarding cost effectiveness of carbon reduction for further analysis. Figure 1 shows the process of data extraction and analysis.

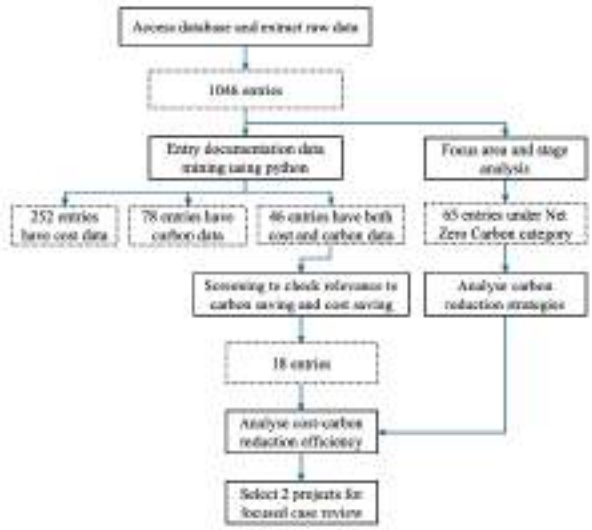


Figure 1. Data extraction and analysis process

diagram

For data analysis, the categories, subcategories, and PCF stage information are first mapped to identify the focus areas and stages of all cases in the database, as displayed in the right side of Figure 1. The strategy types are then summarized and categorized to address the first research question. Following this, the cost effectiveness of carbon reduction of the 18 cases is calculated for comparative analysis to answer the second research question. Finally, two cases are selected for a detailed investigation, examining how different strategies are employed and their respective benefits and challenges. It should be acknowledged that there is no means to assess the data trustworthiness of the data sources, therefore, the analyses are reliant on the quality of data being reported by those supplying them (i.e., NH and its contractors, suppliers). The assumption is that the database well represents the carbon management efforts in NH and its supply chain, although the extent of accuracy and comprehensiveness remains unclear to the researchers.

3 Results

3.1 Focus areas and stages

Carbon management is a very broad topic which is relevant to various project management aspects. The 1,046 entries in the "National Highways Carbon Hub" database are to share tools and resources relevant to carbon, however, some of the entries may mainly focus on other project management dimensions while carbon management was a minor focus. It is valuable to map put which areas the entries are focused while they directly or indirectly contribute to carbon management. The database classified the entries into 14 main categories including Project and programme management; Commercial and procurement; Construction; Customer and stakeholder; Carbon net zero; Consents, orders and planning; Design; Health and safety; Wellbeing and people; Environmental; Governance; Business case and funding; Specifications, standards and surveys; and Handover. Each main category contains multiple subcategories. This classification aligns with the organizational functional units and practices. Each case is classified based on the reporting unit or its closest relevance to a category (as some case may be relevant to several categories), with minor adjustments for similar terms such as merging "knowledge management admin" with "knowledge management," and "well being" with "well-being."

The Sankey diagram in Figure 2 maps out the distribution of categories and PCF stages. It shows that Project and programme management is the largest main category, with 347 entries, including major subcategories like knowledge management which include meetings, forums, workshops and competitions to share knowledge (210), innovation and continuous improvement (73). This focus reflects the critical role of strategic planning and management in achieving carbon reduction goals and underscores the importance of organizational learning and continuous improvement in carbon management. Construction cases are distributed among subcategories like road pavement (17), structures (17), earthworks (14), equipment and materials (10), road restraints (9), and traffic management (9). In the customer and stakeholder category, major subcategories include public liaison (27), roadworks customer experience (26), public consultation (15), and communication planning (14). For the commercial and procurement category, most cases fall under statutory undertakers (46), followed by social value (15 in total), contract management and supplier performance (13), and commercial assurance (11). The relatively lower number of cases in design (33 entries) indicates a potential area for improvement. Given that design and construction stages offer substantial opportunities for carbon reduction through innovative design solutions, material choices, and construction methods, it is crucial to enhance these categories during the construction and design phases for significant long-term carbon reduction benefits. Sixty-five cases fall within the Carbon net zero category, directly reporting the lessons learned about carbon management.

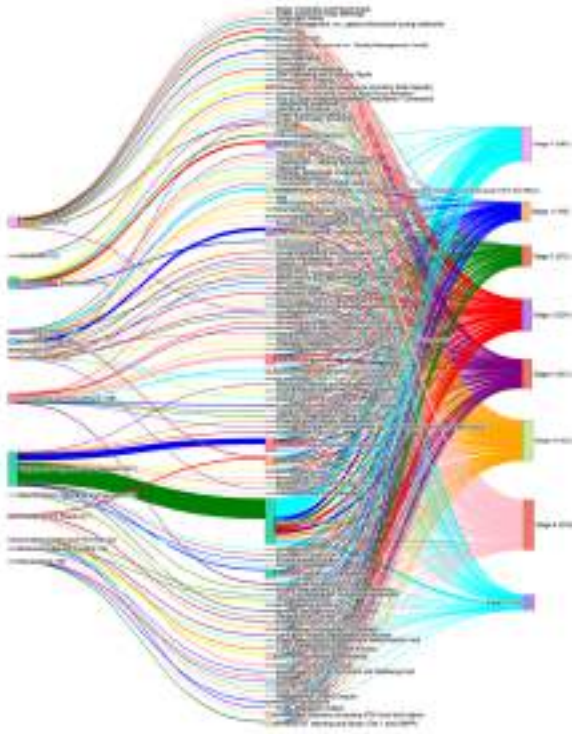


Figure 2. Category distribution of the cases

There are 8 PCF stages in total: Stage 0 to 7 from pre-project, to options, development, and construction

following the project lifecycle, as shown in Figure 3. It is worth noting that with the progress of the PCF stages, the ability to reduce whole life carbon decreases as prior decisions create commitments to particular pathway [7]. Figure 3 also lists the key carbon management activities of a major project across different PCF stages. According to Figure 2, most records in the database have covered Stage 6 (construction, commissioning, and handover) with 518 entries, followed by Stage 5 (construction preparation) with 422 entries, Stage 0 (strategy, shaping, and prioritization) with 365 records, and then Stages 3 and 4, both with over 300 records. Comparatively, Stage 1 and Stage 2 for the options identification and selection, which are crucial for carbon reduction, are not very well covered in the cases.



Figure 3. PCF stages and their abilities to reduce whole life carbon

The distribution of entries favors the development and construction stages, particularly Stage 6 (construction, commissioning, and handover) and Stage 5 (construction preparation). By the time projects reach these later stages, many decisions that significantly influence carbon emissions, such as material selection and design choices, have already been made. Consequently, the potential for reducing whole life carbon is considerably diminished at these stages. The underrepresentation of Stage 1 (options identification) and Stage 2 (options selection) is concerning. These early stages are critical for setting the groundwork for effective carbon reduction strategies. Decisions made during these phases have the greatest potential to influence the overall carbon footprint of a project. For instance, selecting low-carbon materials and technologies, optimizing design for sustainability, and incorporating carbon reduction goals into the project scope are best addressed early in the project lifecycle.

3.2 Carbon reduction strategies

To answer the first research question about what strategies are taken for carbon reduction, we thoroughly investigate the Carbon Net Zero entries in the "National Highways Carbon Hub" database. It is found from the reports that there are two guiding set of strategies: "Carbon Quick Wins" and carbon reduction hierarchy.

3.2.1 Carbon quick wins

To meet their Net Zero targets, NH has implemented a strategy called "Carbon Quick Wins" for major projects. This strategy is supported by a resource known as the "Low Carbon Opportunities Register", which contains a plethora of ideas to reduce carbon emissions throughout the design and construction phases. From the numerous opportunities available, the "Carbon Quick Wins" initiative advocates for the adoption of one or more innovative and mature solutions that offer additional benefits, including cost reduction or neutrality. This initiative targets projects in the design, development, and construction phases to drive carbon reduction efforts while ensuring consistency with current practices and establishing minimum expectations throughout the NH supply chain.

"Carbon Quick Wins" are required to meet five key criteria:

1. High Impact: The innovation should provide additional efficiency, health and safety, performance, and other co-benefits.
2. Cost Reduction/Cost Neutral: While there may be some initial up-front costs, the overall result should be cost-saving.
3. Carbon Reducing: The primary objective is to reduce carbon emissions.
4. No Departures from Standards: The innovation should not require additional approvals and should be supported by technical memos or pilot studies.
5. Mature Level of Innovation: The solution should have a maturity level of 4 or 5 and be applicable with little to no limitations.

Carbon quick win requirements

The "Carbon Quick Wins" strategy emphasizes both cost efficiency and carbon reduction, ensuring that initiatives provide tangible financial benefits while contributing to net-zero targets. This dual focus is crucial in gaining buy-in from stakeholders who may be hesitant to adopt new practices due to perceived costs. Ensuring that the "Carbon Quick Wins" do not require departures from existing standards is bounded by the bureaucratic hurdles and delays often associated with seeking additional approvals for new practices. This approach facilitates smoother implementation and integration of these innovations within current projects, reducing the risk of resistance from stakeholders accustomed to existing standards. Similarly, requiring a high maturity level (i.e., Innovation Capability Maturity Levels Four or Five which indicate innovation activities and resources are integrated and aligned, even synchronised [8] for the innovations ensures that the proposed solutions are well-tested and proven in real-world scenarios. By prioritizing mature innovations, NH can reduce the risk associated

with implementing new technologies and increases the likelihood of successful outcomes. Overall, the "Carbon Quick Wins" strategy represents a proactive, balanced, and pragmatic approach to achieving significant carbon reductions in the short term. By leveraging high-impact, cost-effective, and mature innovations that align with existing practices, NH can drive substantial carbon reductions while maintaining operational efficiency and stakeholder engagement.

Carbon quick win examples Four "Carbon Quick Wins" strategies NH has identified include: using warm mix asphalt, eliminating equipment idling, specifying alternative mixes for ancillary concrete, sourcing renewable energy. Table 1 provides an overview of the four example strategies. They are all applicable to PCF 3, 5, 6. Two of the strategies are about low carbon materials, asphalts and concrete which are two major materials used in road construction. They produce materials with same function using alternative methods while also save costs, which makes them "quick wins" as they are effective, practical, and economically viable for large-scale infrastructure projects. Eliminating equipment idling is a straightforward yet impactful strategy with significant reduction in both carbon emissions and fuel costs, with the ability to be widely adopted with minimal investment and training. The shift to renewable energy sources is one of the most impactful strategies with the potential for up to 90% carbon reduction, highlighting its transformative impact. The cost savings, although significant in the example provided, will vary depending on the scale and duration of projects. However, the long-term benefits of reduced reliance on fossil fuels and enhanced sustainability make this strategy highly desirable.

3.2.2 Carbon reduction hierarchy

There are three types of carbon reduction decision-making strategies: Avoid (e.g., build nothing), Reduce (e.g., build less), and Switch (e.g., build smarter). Avoiding any unnecessary building has the highest ability to reduce carbon and the highest priority in decision-making, reducing the amount of work and material if building is unavoidable comes second, and switching to smarter and cleaner ways when further reduction is impossible comes last. This hierarchy is based on the whole-life carbon reduction potential of each strategy: avoid is the most effective, followed by reduce, and then switch. These strategies are applied across various aspects: low carbon material and technology, design, and construction. As shown in Figure 4, of the 48 cases focused on low carbon material and technology, 37 involve switching and 11 involve reducing. For design, six cases involve switching, six involve reducing, and one involves avoiding. There are only four cases about construction, with one on avoiding, two on reducing, and two on switching. The relatively active learning area of low carbon materials demonstrates a commitment to incorporating sustainable practices in material selection and usage. Also, a big proportion of the cases focused on switching to low carbon energy including electricity, hydrogen, and hydrotreated vegetable oil (HVO). Also, the distribution of the sixty-five cases between switch, reduce, and avoid strategies suggests a need to emphasize "Avoid" and "Reduce" strategies more, as they offer higher potential for carbon savings.

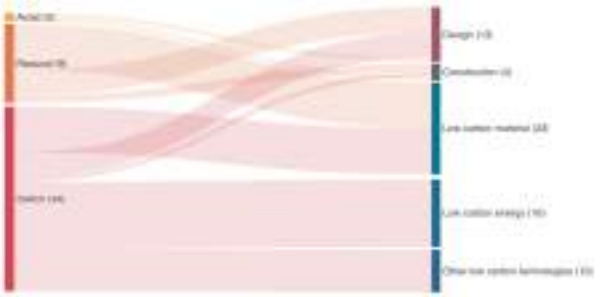


Figure 4. Carbon reduction strategies

3.3 Cost effectiveness of carbon reduction strategies

The Marginal Abatement Cost (MAC) is a key indicator for measuring the cost of carbon reduction [9]. It represents the marginal cost per tonne CO_2e abatement [10]. To assess carbon reduction cost effectiveness, the MAC is calculated by dividing the extra cost/benefit by the total carbon reduction. To answer the second research question, the authors extracted the data from the 18 database entries with both cost change data and carbon reduction data available and valid (which means the scope of cost change data and carbon reduction data aligns, as there are cost saving of the whole project while carbon reduction of a single initiative) and calculated the MACs, which are shown in the last column of Table 2, with the unit £/tCO_2e. Negative MACs indicate cost saving while reducing carbon, whereas positive numbers imply extra costs. Seventeen of the 18 cases have negative MACs ranging from -60£/tCO_2e to -11,111£/tCO_2e, indicating that in these cases, with one tonne of CO_2e reduced, £60 to £11,111 are saved. Only one project required extra investment, with a marginal cost of about £5 per tonne of CO_2e.

The data reveals that significant cost savings can be achieved alongside carbon reduction, as evidenced by the predominantly negative MAC values. However, the significant variance in MAC values among projects indicates that the cost-effectiveness of carbon reduction strategies can differ widely based on project specifics and

Table 1. Summary of Carbon Quick Wins strategy examples

Strategy	Explanation	Carbon reduction potential	Cost saving
Warm mix asphalts (WMAs)	Asphalt mixtures made at lower temperatures—up to 40°C cooler than regular Hot Mix Asphalts (HMAs).	5-10% in production compared to HMAs	Circa 20% shift output increase, saving around £70 million/year across NH
Eliminating equipment idling	Eliminating the operation of equipment, machinery, or vehicles while they are not actively engaged in productive work.	15-18.6% according to cases	£10,000 on fuel savings across 16 machines in one case
Specifying alternative mixes for ancillary concrete	Designated GEN (general) /RC (reinforced) /FND (foundation) concrete mixes are an alternative to Standardized prescribed (ST) concretes, with a more refined mix design compared and require less cement in the mixture to attain equivalent strength levels.	25-50%	Saving £9/m3 for non-structural concrete used
Sourcing renewable energy	Power sourced from naturally replenished resources such as sunlight, wind, water, geothermal heat, and biomass.	up to 90%	Save £3,700 over ten weeks using hydrogen power generation and photovoltaic panels compared to conventional diesel generators

Table 2. Marginal Abatement Cost of 18 cases

Case	Strategy	Category	PCF Stage(s)	Carbon Saving (tCO2e)	MAC (£/tCO2e)
C1	Avoid	Low carbon material	3,4,5,6	2,221.6	-3,632
C2	Switch	Design	3,5,6	5,721	-422
C3	Switch	Low carbon material	3,4,5,6	2,727	-2,180
C4	Switch	Low carbon energy	5,6	0.341	-352
C5	Switch	Low carbon energy	3	99,937	5
C6	Switch	Low carbon energy	4,5,6	164	-504
C7	Switch	Construction	5,6	498.56	-468
C8	Switch	Low carbon material	3, 5,6	8,100	-1,926
C9	Switch	Design	5,6	149.08	-60
C10	Switch	Low carbon energy	3	140	-250
C11	Switch	Low carbon energy	5,6	21.1	-175
C12	Switch	Low carbon material	6	70	-214
C13	Reduce	Low carbon material	3,4,5,6	115.4	-303
C14	Reduce	Design	3,4,5,6	408	-2,941
C15	Reduce	Design	6	130.5	-530
C16	Reduce	Low carbon material	3,4,5,6	339.7	-2,208
C17	Reduce	Construction	6	174	-3,448
C18	Reduce	Low carbon material	5,6	90	-11,111

implementation practices. Eleven of the eighteen cases adopted "Switch" strategy and only one reported "Avoid" as a guiding strategy. The "Avoid" strategy shows more significant cost saving than the average of "Switch" strategy which generally demonstrate cost savings, though with varying extents. Even the "Reduce" strategy (average MAC is -3,423.5 £/tCO_2e) seems more effective than the "Switch" strategy (average MAC is -595.09£/tCO_2e).

Twelve of the projects approached low carbon material, energy, and technology, with seven emphasizing materials, four using better design, and only two adopting better construction methods. The average MAC of low carbon material is -3,096.29 £/tCO_2e, it is -283.67 £/tCO_2e for low carbon energy, -212.5 £/tCO_2e for other low carbon technologies, -988.25 £/tCO_2e for design approaches, and -1,958 £/tCO_2e for construction approaches. While there is not a large dataset and there is a lot of variation in MACs, the adoption of low carbon materials is an area where lessons learnt reporting is indicating particularly cost effective, especially for the C18 case. The two construction cases, particularly C17, also achieved substantial cost-effectiveness, followed by design approaches. In contrast, low carbon energy and other technologies seemed less cost-effective.

The PCF Stages of implementation range from 3 to 6. Most cases introduced carbon reduction measures at Stages 5 and 6, which are relatively late in project life-cycle. Interestingly, the average MAC decreases from Stage 3 (-1,550.17 £/tCO_2e) to Stage 6 (-2,098.40 £/tCO_2e). The absence of early-stage interventions (especially Stages 0, 1 and 2) within the documented cases does not necessarily mean that the early stages were neglected in practice, it only links to the fact that the late

stage case studies were reported in the Carbon Hub dataset more than the earlier stages.

The C18 case, with the lowest MAC, indicates the best cost efficiency in carbon reduction. This £120 million highway construction project reduced emissions by collaborating with another nearby project (less than 6 miles away) for earthworks, reusing 73,350m^3 of spoil. This reduced the use of primary natural resources and material haulage distance by around 50,000 lorry miles, resulting in an estimated net cost saving of about £1 million. The carbon reduction of 90 tonnes CO_2e was primarily due to reduced lorry miles. In summary, this is a special case where the economic benefit is significant while the carbon reduction of 90 tonnes CO_2e may not seem as impactful.

The C5 case required extra investment. It assessed the Renewable Energy Guarantees of Origin (REGO) Certified Electricity at PCF Stage 3. This scheme involves using a tunnel boring machine to drill a tunnel at least two miles long. The machine is very energy-intensive, predicted to consume 197,407,795 kWh of electricity, with the total estimated electricity usage of the project being 202,194.47 MWh, which equated to 99,937 tCO_2e in 2018 coefficients. By using REGO Certified Electricity, grid emissions can be reduced to virtually zero. It was estimated to require £537,837.29 in additional costs in 2018 to buy REGO Certified Electricity, resulting in a MAC of £5.38/tCO_2e. While this is a positive number, it is relatively low, suggesting that the investment in REGOs is a cost-effective way to achieve significant carbon reductions.

Putting the MACs in the wider context, we can evaluate the cost effectiveness of the carbon reduction efforts in UK highway sector. The UK government assumed a traded carbon price of £70/tCO_2e at 2030 [11]. As estimated by Yue et al. [12], the transport sector can achieve full decarbonization in 2050 with MAC at €518/tCO_2e. Comparing to these benchmark numbers, the UK highway sector's carbon reduction strategies are very cost effective.

4 Discussion

The analysis highlights several critical issues that impede the effective evaluation and implementation of the cost effectiveness of carbon reduction strategies in the UK highway sector. Notably, the lack of comprehensive cost and carbon data significantly limits the ability to assess the cost-effectiveness of various measures. Out of 1,046 entries, only 46 contain both cost data and carbon data, and just 18 entries were suitable for comparative analysis after screening. Moreover, the current dataset's inconsistency and rough granularity further complicate the assessment process. The scarcity of detailed, granular data hampers robust evaluation and informed decision-making, underscoring the need for more thorough data collection and reporting to include detailed cost and carbon data for all projects.

The comparative analysis of 18 cases reveals significant variability in the MAC across different initiatives. This variability indicates that the same strategy may have different cost-effectiveness in different cases, depending on the specific context and implementation. The findings suggest that a standardized method for measuring the cost and impact of carbon reduction is necessary for future research.

"Carbon Quick Wins" offer practical, cost-effective measures with significant carbon reduction potential. Adoption and scaling of these low-carbon opportunities across NH projects can maximize their impact, with cost savings reinvested into further initiatives. Building on them, a balanced application of the "Avoid", "Reduce", and "Switch" strategies should be encouraged. Enhancing "Avoid" and "Reduce", integrating various carbon reduction strategies during the option and design phases can maximize carbon savings.

The study's limitations include the lack of comprehensive data of all the carbon management practices apart from the lessons learned and the lack of sources to validate the data trustworthiness [13, 14]. Future research should focus on developing standardized methods for measuring cost and carbon impact and collecting more detailed data [15, 16]. It is also recommended that government should drive the change to require organizations to report betterer quality cost and carbon reduction data with higher granularity. In the meantime, GHG and Environmental, Social, and Governance (ESG) reporting standards should provide clear guidance on how to report the cost and carbon data simultaneously. The technology sector is also encouraged to develop tools to capture carbon reduction and cost information together. Additionally, government policy should explore the long-term impacts and costs of carbon reduction initiatives to provide a more holistic view of their effectiveness and make the most cost-effective carbon reduction policies.

5 Conclusions

The analysis of the 1,046 "Lesson Learned Reports" on carbon management in the UK "National Highways Carbon Hub" found that the primary focus area is "Project and programme management", with a significant emphasis on the construction stage, which occurs late in the lifecycle and has limited impact on whole life carbon reduction. The study highlights the adoption of four "Carbon Quick Wins" approaches and the carbon reduction hierarchy (Avoid-Reduce-Switch), noting that "Avoid" strategies, although most impactful, are less frequently adopted compared to "Reduce" and

"Switch" strategies, as suggested by the number of lessons learnt reports. Filling the gap of cost effectiveness of carbon reduction methods, the analysis of 18 cases using marginal abatement cost (MAC) calculations revealed considerable variability. The results suggest that the cost-effectiveness of carbon reduction strategies can vary significantly depending on the natural of the initiatives and the context they were adopted.

Another key implication from case studies is the significant lack of comprehensive cost and carbon data, with only 46 entries containing both data types and a mere 18 cases suitable for comparative analysis. Even for these limited available data, it is difficult to ensure the data trustworthiness. This scarcity of detailed data impedes the robust evaluation of the cost-effectiveness of various measures, emphasizing the need for standardized and granular data collection methods.

Policy implications for the transportation infrastructure sector include enhancing the focus more on option and design phases, adopting a balanced strategy application. Developing standardized methods for measuring cost and carbon impact, collecting more detailed data, and exploring long-term impacts and costs will provide a more comprehensive understanding of the effectiveness of carbon reduction initiatives. By addressing these areas, the transportation sector can make substantial progress towards net-zero targets and contribute to broader climate change mitigation efforts.

6 Acknowledgement

This project has received funding from the European Commission's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement No 101034337. UK National Highways is acknowledged for sharing the data. The publication reflects only the authors' view, and that the Agency and the Commission are not responsible for any use that may be made of the information it contains.

References

[1] S. Thacker, D. Adshead, C. Fantini, R. Palmer, R. Ghosal, T. Adeoti, G. Morgan, and S. Stratton-Short. Infrastructure for climate action. Technical report, UNOPS, 2021.

[2] National Highways. Net zero highways: Our 2030/2040/2050 plan. Technical report, National Highways, 2021.

[3] National Highways. Net zero highways: Our 2030/2040/2050 plan. progress report 2022-2023. Technical report, National Highways, 2023.

[4] G. Fern´andez-S´anchez, ´A. Berzosa, J. M. Barandica, E. Cornejo, and J. Serrano. Opportunities for GHG emissions reduction in road projects: a comparative evaluation of emissions scenarios using co2nstruct. *Journal of Cleaner Production*, 104:156–167, 2015.

[5] R. B. Noland and C. S Hanson. Life-cycle greenhouse gas emissions associated with a highway reconstruction: A new jersey case study. *Journal of Cleaner Production*, 107:731–740, 2015.

[6] E.E. Keijzer, G.A. Leegwater, S.E. De Vos-Effting, and M.S. De Wit. Carbon footprint comparison of innovative techniques in the construction and maintenance of road infrastructure in the Netherlands. *Environmental Science & Policy*, 54:218–225, 2015.

[7] BSI. PAS 2080: 2023 Carbon management in buildings and infrastructure. Technical report, BSI Standards Limited, 2023.

[8] H. Essmann and N. Du Preez. An innovation capability maturity model–development and initial application. *International Journal of Industrial and Manufacturing Engineering*, 3(5):382–393, 2009.

[9] S. Huang, L. Kuo, and K. Chou. The applicability of marginal abatement cost approach: A comprehensive review. *Journal of Cleaner Production*, 127: 59–71, 2016.

[10] F. Kesicki and N. Strachan. Marginal abatement cost (mac) curves: confronting theory and practice. *Environmental Science & Policy*, 14(8):1195–1204, 2011.

[11] Kesicki, F. (2012). Intertemporal issues and marginal abatement costs in the UK transport sector. *Transportation Research Part D: Transport and Environment*, *17*(5), 418-426.

[12] Yue, X., Deane, J. P., O'Gallachoir, B., & Rogan, F. (2020). Identifying decarbonisation opportunities using marginal abatement cost curves and energy system scenario ensembles. *Applied Energy*, *276*, 115456.

[13] J. Xu and K. MacAskill. A carbon data trustworthiness framework for construction sector. In *Proceedings of the 2023 European Conference on Computing in Construction*, Crete, Greece, 2023.

[14] J. Xu and K. MacAskill. Building trustworthiness in carbon data to achieve net-zero across the life of highway assets. Technical report, University of Cambridge, 2023.

[15] J. Xu, K. MacAskill, M. Yin, J. Zhu, and I. Brilakis. Information requirements over the asset lifecycle to include carbon into digital twin: A UK highway example. In *CIB W78 2024 Conference*. Marrakesh, Morocco, pages 1–3, 2024.

[16] J. Xu and K. MacAskill. Carbon data and its requirements in infrastructure-related GHG standards. *Environmental Science & Policy*, 162:103935, 2024.

Data Collection Challenges in Waste Diversion Assessment: A Building Demolition Case Study

Seungah Suh[1*], Alyssa Roy[1], Sophia Lee[1], Sophia Shelton[1], and Christopher Rausch[1]

[1]Fariborz Maseeh Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin, United States
sasuh@utexas.edu, royalyssa@utexas.edu, ssl2349@utexas.edu, svs838@my.utexas.edu, c.rausch@utexas.edu

Abstract

Optimizing waste diversion in construction and demolition projects is often indeterminate, imposing difficulties on deciding whether to reuse, recycle, or dispose of the materials. While data can be used to help bring objectivity to this process, the data collection process imposes numerous challenges (e.g., missing data, non-standardized data, different reporting requirements, etc.). To address waste diversion data collection difficulties, data requirements are first investigated based on a real-world case study of a building demolition project. Three challenges are then identified: isolated data management, propagation of uncertainty, and lack of standards and guidelines. Additionally, recommendations for data collection strategies including enhancing waste management plans, waste tickets, legal requirements, and incentive programs are provided. This study serves as a guideline for project managers on data definition and collection strategies in practice.

Keywords –

Data collection challenges; Waste diversion; Demolition; Economic assessment; Environmental assessment

1 Introduction

600 million tons of construction and demolition (C&D) debris is generated every year, more than double the annual municipal solid waste production [1]. Accordingly, strategic waste management is crucial to mitigate environmental pollution as well as landfill deficiency and save natural resources [2]. Spurred by a growing consciousness of sustainable C&D waste management and its urgency, waste diversion has garnered significant attention in the Architecture, Engineering, Construction, and Facility Management (AEC-FM) industry.

Waste diversion refers to redirecting materials from landfills through three major strategies: reduction, reuse, and recycling. Reuse and recycling are mainly considered in this study as reduction may not be applicable for some unavoidable waste [3]. To identify the optimal waste diversion solution, assessing the costs and benefits of each method is required [4]. Specifically, economic and environmental impacts, the two primary contributing factors, should be evaluated for effective decision-making.

Previous research using quantitative waste diversion analysis has commonly highlighted the significance of data in the process [5]. Mohareb et al. (2011) compared the methods used to calculate greenhouse gas (GHG) emissions, underlining the influence of onerous data requirements on the method selection [6]. Ratnasabapathy et al. (2020) examined the current status of waste diversion rate applications and illustrated the considerable importance of data quality [7]. Guerra et al. (2024) explored data mining technologies that are relevant to the existing data to predict project-level waste diversion rates [8]. Despite the agreement on the importance of the data, their availability was largely assumed to be present, yet in practice remains a significant challenge.

In contrast, general barriers and opportunities to C&D waste management have been investigated in multiple studies. For example, Yuan et al. (2017) holistically analyzed the drawbacks of current C&D waste management practices [9]. Jin et al. (2017) identified challenges in practitioners' perceptions of C&D waste diversion [10]. Although existing research has provided valuable insights into the practical adoption of waste diversion, they are focused on general challenges, not specifically on data. Therefore, further investigation into the data collection process is needed to address the basis of the waste diversion assessment.

This study aims to identify data collection challenges during the waste diversion assessment based on an actual building demolition case study project. Data requirements, typical data sources, and challenges are identified in order to evaluate the environmental benefits and economic costs of reusing, recycling, and landfilling three building materials in the case study. Data collection

challenges are then categorized with recommended strategies.

2 Case study

The Frank Erwin Center (ERC) demolition project is a case study used to investigate data collection challenges in waste diversion assessment. The Erwin Center, a large multi-purpose arena, was initially built in 1977 in Austin, Texas, United States. Its demolition commenced in July 2023 and was completed in October 2024. Environmental and economic impacts of three building components (i.e., doors, ceiling tiles, and carpet tiles) in the ERC were assessed. A widely adopted unit-based quantification algorithm is used in which total impact is the multiplication of the total quantity by the sum of unit cost, instead of a rough order of magnitude estimation for a more detailed analysis [11]. Cost items for all waste processing methods are identified from previous work but common expenses (e.g., sorting) are excluded for simplicity [12,13]. The authors (with an academic background in civil engineering and life cycle assessment experience in the industry) performed the case study.

3 Data sources and requirements

3.1 Data for waste quantity take-off

The total quantity of the target building components is required to assess both economic and environmental costs using the unit-based approach. As shown in Figure 1, which summarizes the data sources of cost items, the quantity take-off (QTO) can be done manually or retrieved from existing documents (e.g., inventory, bill of materials). Another method is automated QTO using building information modeling (BIM) or QTO software for 2D drawings. Notably, the quantity should be extracted in the unit of the impact factors. Unit conversion formulas would be necessary if the exact match is not available.

For the case study, doors were manually counted from 2D CAD since a BIM model was not provided. In contrast, the number of ceiling and carpet tiles was inferred from the summation of the area of rooms with tile finishing divided by tile size. Relevant rooms were filtered based on the site visits and room types denoted in floor plans. Existing construction documents that summarized the areas of all rooms in the building were used to map their areas. The tile sizes are defined as the standard dimension (carpet tiles 18”x18”; ceiling tiles 2’x2’). The case study also changed the unit. The doors, for instance, were first counted [EA] but converted to weight [kg] based on a rough estimate of the average door weight to follow the landfill cost unit [$/kg].

Item	Data source	Doors	Ceiling Tiles	Carpet Tiles
Quantity	Manual survey	•		
	Existing documents		•	•
	Automated QTO (BIM/CAD)			
Economic: Landfill	Landfill website	•	•	•
	Regional average cost			
Economic: Recycling	Manufacturer's recycling program		•	
	Recycling company	•		•
Economic: Reuse	Secondary market			
	Deconstruction company	•	•	•
Environmental	Environmental Product Declaration (EPD)	•	•	•

Figure 1. Data sources of waste diversion assessment cost items. The data sources for each building component of the case study are marked on the right.

3.2 Data for economic cost assessment

Unit economic costs involved in each waste processing method should be collected for the economic evaluation. Firstly, landfilling requires unit disposal charges and labor cost data. These can be obtained from the official dataset from the Environmental Protection Department [14, 15] or project-specific surveys such as adjacent landfill’s web pages [16].

In terms of the recycling cost, basic recycling charges and extra transportation costs were identified. General recycling costs can be discovered from websites about deconstruction or demolition, considering material type and regional characteristics. More detailed cost data can be scrapped from a product-specific recycling program operated by manufacturers. For example, the case study applied detailed cost data for ceiling tiles from the recycling program of Armstrong, one of the leading ceiling and wall system design and manufacturing companies [17]. However, due to data accessibility, carpet tile data were extracted from the cost analysis report by a carpet tile recycling organization [18].

The economic cost of reuse denotes the estimated resale price subtracted from the sum of transportation and installation costs. This information can be obtained from deconstruction service websites similar to recycling or by reviewing the prices posted on secondary market platforms such as Rheaply [19].

3.3 Data for environmental impact assessment

The environmental impact assessment requires the material’s environmental product declaration (EPD) that covers the unit global warming potential (GWP) of all relevant life cycle stages. The overall concept of the environmental impact assessment complies with a general life cycle assessment (LCA) method. The impact

is calculated by multiplying the total amount of materials and the sum of the applicable life cycle stages' GWP extracted from the corresponding EPD.

The relevant life cycle stages differ by the waste processing method as illustrated in Figure 2. The GWP from the life cycle stages encompasses those from the next life cycles in addition to the current ones. The A4-A5 (Construction) and B (Operation) stages are commonly eliminated from the calculation, assuming all waste management options have identical values. The environmental impact of landfilling, serving as the baseline, indicates the emission from the C (End-of-life) of the current and A1-A3 (Product) phases of the next life cycle. The current D (Beyond) stage is not but the next life cycle's A1-A3 are considered because the material is directly disposed of and it should be remanufactured from scratch. The recycling involves C1-C2 and D of the current and A3 from the next life cycle. C3-C4 stages are not considered because D is incorporated instead, replacing the need to extract raw materials (A1-A2). Similarly, reuse incorporates C1-C2 and D of the current lifecycle but even A3 of the next cycle is not added.

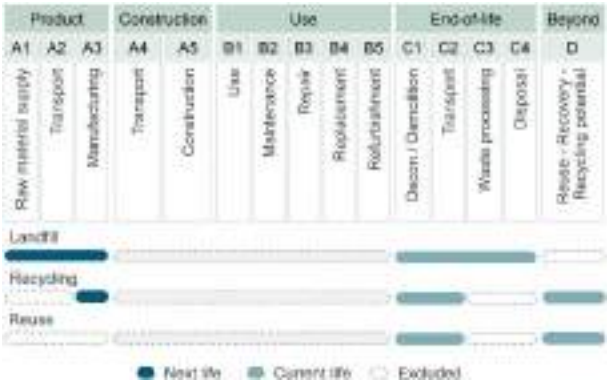


Figure 2. Classification of life cycle stages for the environmental impact assessment by the waste processing method. Different phases of current and next life cycles are applied to calculate the environmental impact of landfilling, recycling, and reuse.

In the case study, the specifications of the building components were unknown and EPDs with detailed information were unavailable. As a result, specific material's lower-level GWP (e.g., A3) was roughly approximated from the upper-level GWP (e.g., A) of a similar product's EPDs provided by its manufacturer.

4 Data collection challenges

Data collection challenges should be clearly identified to prevent them from cascading to future waste management practices which require reference data points to predict and validate the estimations. This section examines the challenges and outlines recommended data collection strategies.

4.1 Isolated data management

A poorly designed documentation plan for waste-related data in the earlier building life cycle phases adds obstacles to information retrieval. A great number of documents have already been generated through the existing practices. However, they cannot be used for the assessment as C&D waste management is not prioritized in the process [13, 20]. Accordingly, readily available documents lack details for waste diversion assessment and are difficult to search.

BIM models, often created during the design and construction phases, are more likely to be unavailable for demolishing buildings. A labor-intensive and error-prone manual QTO is thus prevalent, making the data collection more challenging.

Inventories and specifications generated for building construction and maintenance do not contain the data for waste (e.g., EPD, corresponding quantity, and the construction project manager's comments on future waste management plans). Finding an appropriate EPD that includes information for the end-of-life phase, for example, is challenging without enough description that only the design engineer or installer can provide. Even when the information exists, it is not organized well to search easily. Unstructured pieces of information were often scattered around the hundreds of pages in the case study.

Table 1. Insufficient data fields in the waste tickets from the ERC demolition project

Data Types	Fields
General	Contractor Project Manager
Logistics	Hauler Facility Date
Material Type	Metal Concrete Manifest Trash Trash
Analysis	Total Tons Total Recycling Recycling Rate

Moreover, waste tickets alone rarely support the process although they are required to be submitted. They usually do not designate the work type (e.g., new construction, demolition, renovation) and classify waste by only a few materials (e.g., concrete, metal, brick, wood, cardboard) instead of functional units as illustrated in the waste ticket data of the case study (Table 1).

Furthermore, waste tickets are restricted to the weight of recycled and landfilled waste. This discrepancy in the stakeholders' interests between the earlier phases and waste diversion blocks smoother and integrated data collection.

4.2 Propagation of uncertainty

Vast uncertainties exist in numerous variables of waste diversion, including item quantity, quality, historic value, and changing market needs [21,22]. The target level of detail of the assessment is one of the factors defining the variables. Waste recycling cost, for example, can be calculated at a high level based on the estimated waste type and regional recycling cost. However, it can also be decomposed into details by selecting a specific recycling facility of interest considering the distance from the site, accepted waste type, waste diversion rate, and more accurate waste estimation [14]. As the applicable variables differ by data requirements of the target level of detail, the degree of uncertainty changes accordingly.

Managing the uncertainty becomes more complicated as the uncertainty in one aspect of waste diversion is multiplied by assumptions from other variables. It is difficult to define an allowable range of differences for data collection. For instance, resale price includes uncertainty stemming from a market need for a specific material type; at the same time, material quantity can affect the pricing policy. The Erwin Center case had to consider three major compounded sources of uncertainty to calculate the reuse cost of ceiling tiles: 1) quantity unit conversion from area to weight, which propagates to estimating 2) the labor cost for tile removal that also varied by the project location, which affects 3) the resale price as it differs by the quantity, regional demand, and item condition. Concluding whether the final variance imposed from multiple sources is tolerable is difficult, and these ambiguities can therefore bring verbosity to waste diversion estimation.

4.3 Lack of guidelines and standards

Missing guidelines and standards, that could assist the data gathering on behalf of experienced human experts, impede the collection process. Skilled experts in waste diversion can use their own database or find the most suitable data sources but most experts still need assistance for the best practice. Due to the lack of well-established standards and guidelines in the waste diversion field, data collection efficiency is compromised [10,5].

In the case study, experienced experts' advice and standardized processes were mostly restricted or did not exist. Even the authors studying waste diversion had difficulties in determining a realistic resale price for reuse as the price changes by the fluctuating demand. Furthermore, an unstandardized dataset intensified the problem. There was no well-established guideline or standard listing crucial considerations and possible database to find the relevant data such as EPDs that perfectly match the material of interest. Unit conversion formulas and assessment methods also had to be modified multiple times following different EPD formats. When selecting EPDs of ceiling tiles, carpet tiles, and doors for the case study, the research team decided to prioritize EPDs that separately documented the global warming potential of Product stages (A1-A3) in addition to other basic requirements. These criteria for reasonable decision-making were developed from scratch as readily available resources were difficult to access and the EPDs were all distributed across different manufacturer's websites in varying formats.

It is currently infeasible to locate data sources that match all project attributes and variables. Hence, clear directions defining what is acceptable and how to adapt the data are necessary to ensure the reliability of the assessment. A centralized database and standardized criteria are needed to make the waste diversion assessment more accessible.

4.4 Recommendations

This research introduces four recommendations to enhance the data collection practices. The first strategy is to require contractors to submit a comprehensive waste management plan (WMP) in advance of a construction or demolition project. A WMP allows integrated data generation and management, addressing the limitation of isolated data management. The stakeholders can incorporate waste diversion in their plans before the project begins to minimize missing data and redundancy. Terminologies can also be standardized with WMP, which can contribute to communication efficiency. Furthermore, incorporating quantitative impact analysis and supply-demand simulation of reclaimed materials will enable WMP to play a more strategic and proactive role in waste diversion [23].

Restructuring the C&D waste post-reporting framework is also needed. Explicit descriptions of processing types, such as demolition/new construction and reuse/recycling/landfill, should be added to waste tickets. Categorization of materials in detail (e.g., plastic/furniture/contaminated, link to drawings/BIM elements) is also suggested for further analysis. This enables reducing uncertainty due to the lack of well-documented data.

Advanced legislation is another recommendation to motivate the execution of the strategies, and eventually, promote the better establishment of waste diversion guidelines and standards. Previous studies in waste recovery have demonstrated that financial incentives are

effective but need further support for practical implementation [24,25]/landfill, should be added to waste tickets. Categorization of materials in detail (e.g., plastic/furniture/contaminated, link to drawings/BIM elements) is also suggested for further analysis. Internal motivations of general contractors, demolition, or waste management specialists can drive the efforts but legislative actions can facilitate the change. In practice, the C&D ordinance has already been published or is under development in multiple states and cities to enhance the C&D waste management practices [26-28]. Recycle Colorado, a nonprofit organization to advance circular economies in Colorado, also provided a toolkit for policy development [29].

Similarly, enhancing certification programs such as Energy and Environmental Design (LEED) can provide societal and indirect financial motivations by providing constructive instructions. Indeed, the U.S. Green Building Council has already started a discussion to include Building and Materials Reuse credit in their new LEED v5 rating system to encourage more material reclamation and reuse [30]. Clear legislative and certification processes, accompanied by knowledge exchange and assistance from organizations such as Build Reuse Wiki [31], will accelerate the improvements in data collection.

5 Conclusions

This study analyzed data collection challenges for the environmental and economic assessment of waste diversion methods based on a case study. Calculating the impacts of landfilling, recycling, and reuse of building materials in a sample demolition project revealed data collection challenges (i.e., isolated data management, propagation of uncertainty, and lack of standards and guidelines). Recommendations for improved data collection are also presented according to the analysis.

This research analyzed practical difficulties experts can encounter when gathering data for the quantitative waste diversion assessment, a task often left to individuals to resolve. The findings will serve as a guideline to establish improved data collection strategies. It is also expected to assist the project managers and future studies in implementing them for more waste diversion.

Still, the study has a limitation in that it focused on investigating the case study-specific data collection challenges. The identified problems might not apply to other projects, or different difficulties can be identified in new regions and authorities. Future research is necessary to discover possible solutions for data collection in practice. Their feasibility and effectiveness should also be interpreted theoretically and through real-world applications.

References

[1] US EPA. Advancing Sustainable Materials Management: 2018 Fact Sheet. On-line: https://www.epa.gov/sites/default/files/2021-01/documents/2018_ff_fact_sheet_dec_2020_fnl_508.pdf, Accessed: 29/04/2024.

[2] Arora, M., Raspall, F., Cheah, L., and Silva, A. Buildings and the circular economy: Estimating urban mining, recovery and reuse potential of building components. *Resources, Conservation and Recycling*, 154:104581, 2020.

[3] Kabirifar, K., Mojtahedi, M., Wang, C., and Tam, V. W. Y. Construction and demolition waste management contributing factors coupled with reduce, reuse, and recycle strategies for effective waste management: A review. *Journal of Cleaner Production*, 263:121265. 2020.

[4] Lai, K., Li, L., Mutti, S., Staring, R., Taylor, M., Umali, J., and Pagsuyoin, S. Evaluation of waste reduction and diversion as alternatives to landfill disposal. In *Proceedings of 2014 IEEE Systems and Information Engineering Design Symposium (SIEDS)*, pages 183–187. Charlottesville, USA, 2014.

[5] Knoth, K., Fufa, S. M., and Seilskjær, E. Barriers, success factors, and perspectives for the reuse of construction products in Norway. *Journal of Cleaner Production*, 337:130494. 2022.

[6] Mohareb, E. A., MacLean, H. L., and Kennedy, C. A. Greenhouse Gas Emissions from Waste Management—Assessment of Quantification Methods. *Journal of the Air & Waste Management Association*, 61(5):480–493. 2011.

[7] Ratnasabapathy, S., Alashwal, A., and Perera, S. Investigation of waste diversion rates in the construction and demolition sector in Australia. *Built Environment Project and Asset Management*, 11(3):427–439. 2020.

[8] Guerra, B. C., Koo, H. J., Caldas, C., and Leite, F. Prediction of waste diversion and identification of trends in construction and demolition waste data using data mining. *International Journal of Construction Management*, 24(4):374-383, 2024.

[9] Yuan, H. Barriers and countermeasures for managing construction and demolition waste: A case of Shenzhen in China. *Journal of Cleaner Production*, 157:84–93. 2017.

[10] Jin, R., Li, B., Zhou, T., Wanatowski, D., and Piroozfar, P. An empirical study of perceptions towards construction and demolition waste recycling and reuse in China. *Resources, Conservation and Recycling*, 126:86–98. 2017.

[11] Akbarnezhad, A., Ong, K. C. G., and Chandra, L. R. Economic and environmental assessment of deconstruction strategies using building

information modeling. *Automation in Construction*, 37:131–144, 2014.

[12] Duran, X., Lenihan, H., and O'Regan, B. A model for assessing the economic viability of construction and demolition waste recycling—the case of Ireland. *Resources, Conservation and Recycling*, 46(3):302–320. 2006.

[13] Yuan, H. P., Shen, L. Y., Hao, J. J. L., and Lu, W. S. A model for cost–benefit analysis of construction and demolition waste management throughout the waste chain. *Resources, Conservation and Recycling*, 55(6):604–612. 2011.

[14] Cheng, J. C. P., and Ma, L. Y. H. A BIM-based system for demolition and renovation waste estimation and planning. *Waste Management*, *33*(6):1539–1551. 2013.

[15] US EPA. Versions of the Waste Reduction. On-line: https://www.epa.gov/warm/versions-waste-reduction-model#v16, Accessed: 09/12/2024.

[16] Texas Disposal Systems. Gate Rates & Waste Drop-Off Prices. On-line: https://www.texasdisposal.com/gate-rates/, Accessed: 29/04/2024.

[17] Armstrong Ceiling Solutions. Ceilings Recycling Program. On-line: https://www.armstrongceilings.com/commercial/en/performance/sustainable-building-design/ceiling-recycling-program.html, Accessed: 29/04/2024.

[18] Crowe. CARE Cost Analysis Report. On-line: https://carpetrecovery.org/wp-content/uploads/2019/09/CARE-Cost-Analysis-Report-08-28-2019.pdf, Accessed: 09/12/2024.

[19] Rheaply. Rheaply. On-line: https://rheaply.com/, Accessed: 01/12/2024.

[20] Yuan, H., and Shen, L. Trend of the research on construction and demolition waste management. *Waste Management*, 31:670–679. 2011.

[21] Rašković, M., Ragossnig, A. M., Kondracki, K., and Ragossnig-Angst, M. Clean construction and demolition waste material cycles through optimised pre-demolition waste audit documentation: A review on building material assessment tools. *Waste Management and Research*, 38(9):923–941. 2020.

[22] Rausch, C., Suh, S., Repousis, N., Titterington, N., and Nguyen, H. Assessing the Viability of Robotic Disassembly of Building Components for Resource Recovery. In *Proceedings of the International Symposium on Automation and Robotics in Construction*, pages 1048–1056, Lille, France, 2024.

[23] Rausch, C., Suh, S., Repousis, N., Lee, S., Shelton, S., Roy, A., Nguyen, H., Titterington, N., and Saab, M. Hook'em Building Materials Before the Landfill Does: Diverting Demolition Waste from Frank Erwin Center – Final Project Report. University of Texas at Austin, 2024.

[24] Asare, W., Oduro-Kwarteng, S., Donkor, E. A., and Rockson, M. A. D. Cost-effectiveness of incentive schemes for waste material resource recovery. *Cleaner Waste Systems*, 2:100019. 2022.

[25] Lakhan, C. The relationship between municipal waste diversion incentivization and recycling system performance. *Resources, Conservation and Recycling*, 106:68–77. 2016.

[26] AustinTexas.gov. Construction and Demolition Recycling Ordinance. On-line: https://www.austintexas.gov/cd, Accessed: 29/04/2024.

[27] CalRecycle. Construction and Demolition (C&D) Diversion Informational Guide. On-line: https://calrecycle.ca.gov/lgcentral/library/canddmodel/, Accessed: 09/12/2024.

[28] San Antonio Reuse. Deconstruction Ordinance. On-line: https://www.sareuse.com/deconstruction, Accessed: 29/04/2024.

[29] Recycle Colorado. Colorado Construction, Demolition & Deconstruction Policy Toolkit. On-line: https://www.recyclecolorado.org/c-d-policy-toolkit, Accessed: 09/12/2024.

[30] U.S. Green Building Council. Introduction to LEED v5 for Design and Construction. On-line: https://www.usgbc.org/scorm/introduction-leed-v5-design-and-construction, Accessed: 21/10/2024.

[31] Build Reuse. Build Reuse Wiki. On-line: https://www.buildreuse.org/wiki, Accessed: 09/12/2024.

Evaluating the Carbon Neutrality of Concrete Recycling Using CO_2 Capture Technologies from Cement Production

Svetlana Besklubova[1*], Ekaterina Kravchenko[2], Muhammad Huzaifa Raza[1] and Ray Y. Zhong[1]

[1]Department of Data and Systems Engineering, The University of Hong Kong, Pokfulam, Hong Kong
[2]Southern Federal University, Rostov-on-Don, Russia
lanabess@hku.hk; ekakra@sfedu.ru; huzaifa@connect.hku.hk; zhongzry@hku.hk

Abstract -

The construction industry encounters significant environmental challenges, primarily due to concrete waste from demolitions and high CO_2 emissions from cement production. Concrete waste is recycled into aggregates, generating a significant amount of waste concrete powder (WCP) as a byproduct. However, the application of WCP in recycling processes is limited, leading to landfill disposal and environmental risks. This study explores the potential of combining carbon-neutral technology with large-scale concrete production by using WCP as a partial substitute for both cement and fine aggregates. Active carbonation is highlighted for its ability to capture CO_2 while improving the properties of recycled concrete aggregates (RCA) and WCP. Although technological advancements exist, a thorough evaluation of financial viability is essential. This research focuses on assessing the economic feasibility of integrating carbon-neutral technology into concrete recycling, particularly by carbonating waste concrete using flue CO_2 from cement plants. Thus, the paper proposes a method to incorporate environmental assessments into an activity-based costing (ABC) model, enabling a dual evaluation of financial costs and environmental impacts. By linking environmental impact assessment results to costs, this approach facilitates a comprehensive evaluation for involved stakeholders. An assessment of the Hong Kong market showcases the practicality of the proposed matrices. The comparison of recycling methods indicates significant environmental improvements, with 362% and 440% reductions in CO_2 emissions from cement plants and natural material processing, respectively, underscoring the potential of carbonization technology and the need for government support.

Keywords -

Construction waste management, Concrete recycling, Flue gas, CO_2 sequestration, Recycled concrete aggregate

1 Introduction

The construction sector stands as a foundational component of the economic system, notable for its substantial energy usage and considerable carbon footprint. It has been documented that approximately one-third of worldwide carbon emissions can be attributed to the construction sector. Importantly, the adoption of recycled materials has been recognized as a powerful strategy to cut down on carbon emissions within this industry. Over fifty percent of construction and demolition debris is comprised of materials like waste concrete, as well as fractured bricks and blocks. These materials hold significant promise for lowering carbon emissions through recycling and reuse [1]. Research by Xiao, et al. [2] indicates that using recycled concrete has the potential to reduce carbon emissions by 12.4-52.2%, a percentage that is likely to grow in tandem with the escalation in recycled material usage. Furthermore, the employment of recycled concrete can help to mitigate the global scarcity of raw materials, like sand and aggregates. The adoption level of recycled concrete now serves as a critical measure within the green building evaluation frameworks of various countries.

While the advantages of waste concrete are widely recognized, the actual rates of recycling in numerous regions, including Hong Kong, still remain a concern for stakeholders. Waste concrete can be processed into recycled concrete aggregate (RCA), suitable for use in fresh concrete mixes. This recycling also leads to the creation of waste concrete powder (WCP), which can constitute 30-70% of the utilized concrete [3]. In countries such as China, the yearly output of WCP from recycling activities is estimated to be a staggering 60 million tonnes [4]. Instead of the prevalent method of landfill disposal, WCP could be used as a partial replacement for cement in concrete mixes. This substitution can lower the usage of cement-based materials, improve the rheological and mechanical characteristics of the concrete, and increase its durability.

Thus, incorporating WCP into the recycling process can further improve the recycling rate and sustainability.

Given the considerable promise of waste concrete, a multitude of research has been conducted in recent years. These studies have primarily concentrated on enhancing the attributes of recycled concrete through technical means. For instance, Poon and Chan [5] explored the properties of recycled concrete when applied to concrete pavement, offering various recommendations for enhancement. A method employed to improve the strength and quality of construction materials sourced from waste concrete is carbonation, which also supports strategies aimed at achieving carbon neutrality [6]. Carbonation refers to the reaction between carbon dioxide in the atmosphere and the calcium hydroxide present in cement paste. To achieve concrete carbonation, a high concentration of pure CO_2 gas (up to 99%) is typically employed [7]. As indicated by Ouyang, et al. [8], carbonation curing can enhance the compressive strength of RCA by 13%. Furthermore, carbonation enhances the material's resistance to corrosion and frost, thereby prolonging its lifespan [9]. The carbonation of RCA not only enhances their properties but also helps reduce the greenhouse effect caused by carbon dioxide emissions [10]. WCP, owing to its inferior physical and mechanical properties, is not suitable for direct incorporation into concrete mixes, and similarly, RCA requires a pre-carbonization procedure. Recent studies have demonstrated the promising potential of carbonizing WCP in terms of enhancing the strength of future concrete and its viability as a material for carbon sequestration [11]. Liu, et al. [12] and Wu, et al. [13] research indicates that WCP can absorb CO_2 up to 21% of its own weight.

Recently, anthropogenic gases, such as flue CO_2 from cement plants, have emerged as a promising source for concrete carbonation. Studies have demonstrated the feasibility of sequestering CO_2 from flue gases into materials like air pollution control residues, capturing 8–12% CO_2 by weight [14], while improving concrete properties such as compressive strength [15]. Research has also highlighted the cost-effectiveness and carbon reduction potential of integrating CO_2 capture in concrete production [16]. Pilot projects worldwide, such as those in Canada and the U.S., have validated large-scale mineral carbonation and technologies like Reversa, demonstrating their readiness for commercialization and compliance with industry standards [17, 18].

The carbonation of waste concrete for CO_2 sequestration not only supports carbon neutrality goals but also improves the strength of construction materials derived from such waste. While projects utilizing flue gases for CO_2 sequestration show promise, reliance on commercially captured CO_2 remains a drawback, potentially conflicting with sustainable processing. Additionally, the lack of economic analysis limits the understanding of its cost-effectiveness. Additionally, the integration of WCP into the concrete waste recycling process has yet to be fully explored, despite research suggesting that WCP can substitute for cement and natural fine aggregates in concrete formulations [19]. In response to these challenges, this study investigates the feasibility and environmental impact of active carbonation techniques for RCA and WCP derived from concrete debris, aiming to incorporate them into non-structural concrete mixtures. It also assesses the feasibility of using alternative CO_2 sources, such as flue gases from cement plants, to sustainably process concrete waste and incorporate both elements into the recycling process.

2 Model development

The research methodology comprises four key steps. First, a comprehensive literature review was conducted, providing insights into construction and demolition waste (CDW) management, the carbonization of demolished concrete, and the integration of this technology. Second, a cost estimation matrix was developed to evaluate financial costs and environmental impacts, with parameters selected based on the relevant processes involved. Third, a case study focused on the Hong Kong market was chosen to assess the feasibility of the proposed process with carbon-neutral technology integration, comparing it to traditional methods. Finally, the parameters were entered into a simulation environment to evaluate the model's practicality and analyze the individual cost components.

2.1 Parameters selection

In Hong Kong's traditional waste management system, all concrete wastes are first processed through crushing and separation before being sent to recycling facilities for conversion into RCA [20]. The RCA is then delivered to concrete production plants, while any excess WCP generated during the refinement of demolished concrete is disposed of in landfills [21]. Natural materials such as sand, gravel, and cement are sourced from approved suppliers to prepare the final concrete product. The emphasis on recycling WCP and utilizing flue CO_2 from cement plants has prompted the exploration of an alternative supply chain for concrete recycling. In this approach, demolished concrete from construction sites is transported to the recycling facility for processing into RCA and WCP. Flue CO_2 serves as an alternative gas source for carbonization and is transferred from the cement plant to the recycling plant. The carbonized RCA and WCP are subsequently sent to the concrete plant. Additionally, natural materials (sand, gravel) and cement are sourced from designated suppliers, though the

quantity of these materials is reduced due to their partial substitution with recycled aggregates and WCP. The processes discussed are illustrated in Figure 1, where black and blue arrows represent the traditional methods, while orange arrows indicate the proposed processes.

The activities carried out at various locations—such as suppliers, recycling facilities, cement plants, and landfills—are referred to as facility-based activities, with the associated expenses known as facility-based costs [22]. To achieve an accurate evaluation of all costs, an Activity-Based Costing (ABC) approach is utilized to allocate resources effectively to each process within a specific facility [23]. The energy (E) required for equipment operations at each facility and technology-related costs (AT) are quantified. The technology-related costs include expenses for acquiring additional equipment necessary for new concrete waste recycling processes. This encompasses equipment for capturing flue CO_2 and the enhanced capacity needed for carbonizing RCA and WCP. Additionally, the avoided effects are considered from an environmental perspective. Processes in the alternative scenario that replicate those in the traditional scenario are excluded from the model, as their impacts would remain consistent in both cases.

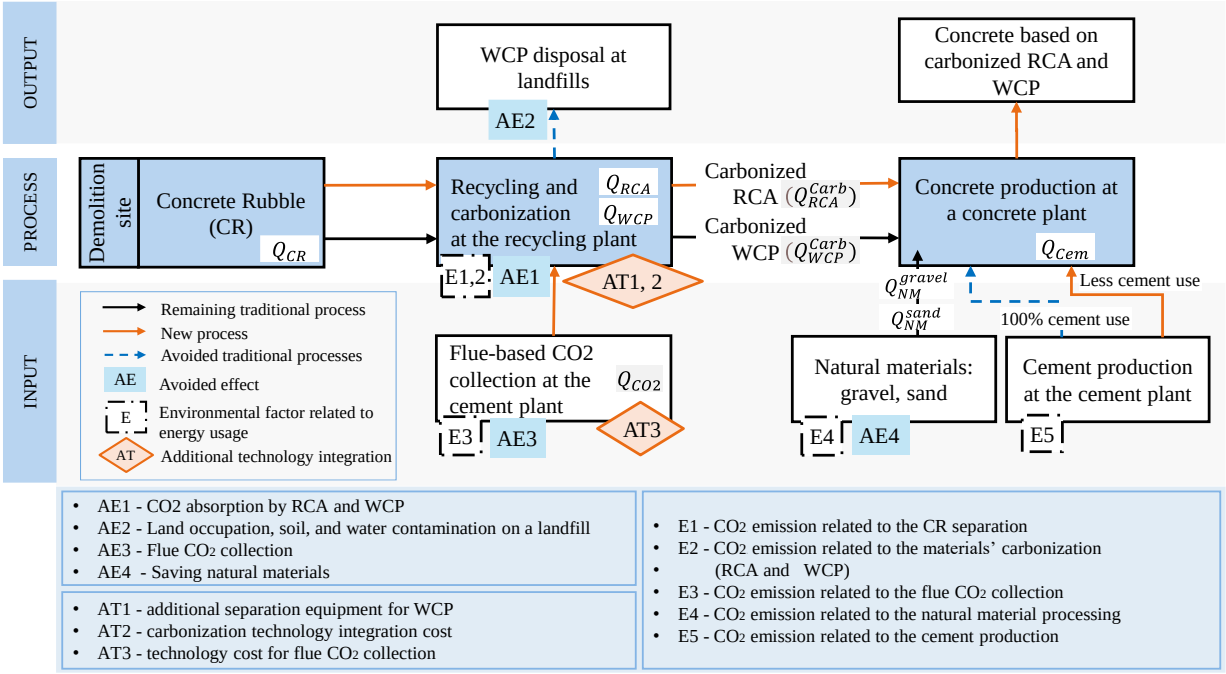


Figure 1. Identification parameters for the traditional and alternative concrete waste recycling processes

2.2 Model assumptions

The model assumes a one-way material flow, moving from demolition sites and suppliers to landfills or recycling centers, and ultimately to concrete plants, reflecting typical industry practices. No limits are imposed on the number of facilities at each stage to ensure scalability. All demolished concrete rubble (CR) delivered to recycling facilities is assumed to be fully processed into RCA or WCP, based on the efficiency of modern recycling technologies. Technology costs, reflected as depreciation, are based on initial purchase and installation expenses and are limited to equipment essential to the recycling process. Environmental benefits, including resource conservation and CO_2 emission reductions, are included to capture the sustainability advantages of substituting cement and natural aggregates with recycled materials. These assumptions are grounded in established practices and the specific context of this study.

2.3 Quantification of parameters

The quantification of parameters begins with the identification of activities within each facility and material flow based on Figure 1. The following material flow and related major facility-based activities can be determined:

a_1: Concrete rubble (CR) processing
a_2: RA processing
a_3: WCP processing
a_4: Anthropogenic CO_2 processing
a_5: Cement production
a_6: Natural materials processing (gravel)

a_7: Natural materials processing (sand)

For each type of activity, resources known as "activity drivers" will be allocated. An activity driver enables the allocation of economic transaction costs to an activity [24]. For instance, when processing concrete rubble (CR, represented as a_1) into RCA and WCP at a recycling facility, the resources utilized include the operational time of sieving and screening machines, as well as the energy consumed per unit of material processed. These resources can be represented as b_1 and b_2 respectively:

b_1: Usage of equipment (hours) per unit of processed material for CR processing into RCA and WCP (sieving and screening machines)

b_2 : Energy usage (kWh) of equipment per unit of processed material for CR processing into RCA and WCP (sieving and screening machines)

b_3: Usage (hours) of carbonization equipment per unit of processed material for RCA and WCP carbonization

b_4: Energy usage (kWh) of carbonization equipment per unit of processed material for RCA and WCP carbonization

b_5: Usage (hours) of CO_2 collection equipment per unit of processed material at the cement plant

b_6: Energy usage (kWh) of CO_2 collect equipment per unit of processed material at the cement plant

Environmental parameters were calculated using life cycle assessment (LCA) to evaluate the sustainability of the proposed materials and processes. An extensive LCA was performed following ISO 14040 and 14044 standards using OpenLCA 2.0.3 software. The Ecoinvent 3.9.1 database was utilized for the assessment due to its reliable and compliant data source. The environmental parameters were determined by considering the electricity required for processing recycled aggregates before their inclusion in the production cycle. Additionally, environmental parameters were calculated for the production of natural aggregates and cement. To ensure consistency and comparability, transportation was not included in the boundary conditions for CO_2 emission assessment. By focusing on the direct energy and material inputs, the study isolates the specific contributions of transformation, providing a clearer insight into the sustainability potential of recycled aggregates.

To create a comprehensive representation, resources can be allocated to each activity and illustrated as a matrix (Eq. 1). In this matrix, the horizontal rows represent the various activities (e.g., CR, RCA, WCP, CO_2 processing), while the vertical columns correspond to the types of resources utilized (e.g., energy, depreciation costs). For instance, 'b_{11}' denotes the equipment usage time for processing one unit of CR, while 'b_{12}' represents the energy consumption of the separation equipment for processing one unit of CR.

$$B = \begin{matrix} CR \\ RCA \\ \vdots \\ NM_{sand} \end{matrix} \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{16} \\ b_{21} & b_{22} & \cdots & b_{26} \\ \vdots & \vdots & \ddots & \vdots \\ b_{71} & b_{72} & \cdots & b_{76} \end{pmatrix} \quad (1)$$

The costs of activity drivers (per unit) can be represented as a column vector (Eq. 2). For example, the unit costs of activity drivers $b_1 - b_6$ are as follows:

c_1: Depreciation cost of equipment per unit of processed CR (sieving and screening machines) (AT1)

c_2: Energy cost of equipment per unit for CR processing (sieving and screening machines) (AT1)

c_3: Depreciation Cost of carbonization equipment usage per unit of carbonized material (AT2)

c_4: Energy cost of carbonization equipment usage (AT2)

c_5: Depreciation cost of CO_2 collect equipment usage per unit at the cement plant (AT3)

c_6: Energy cost of CO_2 collect equipment usage (AT3)

$$C = \begin{pmatrix} c_1 \\ \vdots \\ c_6 \end{pmatrix} \quad (2)$$

Thus, vector AC consists of items that represent the costs of activities (per unit) computed using Eq. 3 :

$$AC = B \cdot C = \quad (3)$$

$$\begin{pmatrix} b_{11} & \cdots & b_{16} \\ \vdots & \ddots & \vdots \\ b_{71} & \cdots & b_{76} \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ \vdots \\ c_6 \end{pmatrix} = \begin{pmatrix} ac_1 \\ \vdots \\ ac_7 \end{pmatrix}$$

The next step involves allocating the respective quantities to the activities for the final product. Figure 1 also visualizes the quantity of materials at each activity. For instance, $'Q_{\mathrm{CR}}'$ represents the amount of CR arriving at the recycling facility, while Q_{RCA} and Q_{WCP} represents the amount of materials arriving at the concrete plant. Thus, those parameters can be represented as follows:

q_1 : Material quantity (CR) transported from a demolished site to a recycling plant (Q_{CR})

q_2 : Material quantity at RF (RCA and WCP, not carbonized: Q_{RCA}, Q_{WCP})

q_3: Material quantity at cement plant (CO_2 and cement: Q_{CO2}, Q_{Cem})

q_4: Natural materials quantity transported to a concrete plant (gravel and sand: Q_{NM}^{gravel}, Q_{NM}^{sand})

Therefore, the matrix 'Q' represents the allocation of material quantities to each activity (Eq. 4). In this matrix, the horizontal lines represent the quantity of materials ($q_1 \ldots q_4$) at different facilities, while the vertical lines correspond to the different activities ($a_1 \ldots a_7$). For example, $'q_{11}'$ stands for the quantity of CR transported from a demolishing site (DS) to a recycling plant, while $'q_{22}'$ and $'q_{32}'$ represents the quantity of RA and WCP at the recycling facility correspondently (not carbonized).

$$\mathrm{Q}= \begin{matrix} DS \\ RF \\ CM \\ NM \end{matrix} \begin{pmatrix} q_{11} & q_{12} & \cdots & q_{17} \\ q_{21} & q_{22} & \cdots & q_{27} \\ q_{31} & q_{32} & \ddots & q_{37} \\ q_{41} & q_{42} & \cdots & q_{47} \end{pmatrix} \quad (4)$$

Final product cost (PC) can be determined by Eq. 5:

$$PC = Q \cdot \mathrm{A}C = \quad (5)$$

$$\begin{pmatrix} q_{11} & q_{12} & \cdots & q_{17} \\ q_{21} & q_{22} & \cdots & q_{27} \\ \vdots & \vdots & \ddots & \vdots \\ q_{41} & q_{42} & \cdots & q_{47} \end{pmatrix} \cdot \begin{pmatrix} ac_1 \\ ac_2 \\ \vdots \\ ac_7 \end{pmatrix} = \begin{pmatrix} pc_1 \\ pc_2 \\ \vdots \\ pc_4 \end{pmatrix}$$

For the environmental cost calculation, a matrix 'E' is utilized to represent the environmental cost per unit of the considered material. In this matrix, the horizontal rows correspond to various activities (a_1 to a_7) represented above, such as CR, RA, WCP, and CO_2 processing, while the vertical columns indicate CO_2 emissions per unit of material during specific processes (e_1). Environmental parameters are illustrated in Figure 1 as $E_1 - E_5$ and AE_1 - AE_4 and represented as column values below:

e_1: Amount of CO_2 emissions by each equipment per unit of material *(a_i)* during recycling and carbonization processing (E1, E2)

e_2: Amount of CO_2 absorption during the RCA and WCP carbonization (AE1)

e_3: An environmental effect associated with soil and water contamination, as well as land occupation in landfills (AE2).

e_4: Amount of CO_2 emissions by each equipment per unit of flue CO_2 collection (E3)

e_5: Amount of captured CO_2 per unit of flue gases at the cement plant (AE3)

e_6: Amount of CO_2 emissions per unit of cement produced (E5)

e_7: Environmental effect related to the unit of natural material (gravel) processing for the concrete manufacturing (E4)

e_8: Savings associated with the substitution of natural materials (gravel) per unit (AE4)

e_9: An environmental effect related to the unit of natural material (sand) processing for concrete manufacturing (E4)

e_{10}: Savings associated with the substitution of natural materials (sand) per unit (AE4)

The matrix 'E' is represented as Eq. 6. For instance, the matrix element $'e_{11}'$ indicates the amount of CO_2 emissions produced per unit of CR processed into RCA and WCP. In this context, CO_2 emissions are represented as negative values, while CO_2 sequestration is denoted as positive values.

$$E = \begin{matrix} CR \\ RCA \\ \vdots \\ NM_{sand} \end{matrix} \begin{pmatrix} e_{11} & e_{12} & \cdots & e_{1\,10} \\ e_{21} & e_{22} & \cdots & e_{2\,10} \\ \vdots & \vdots & \ddots & \vdots \\ e_{71} & e_{72} & \cdots & e_{7\,10} \end{pmatrix} \quad (6)$$

The environmental costs of activity drivers (per unit) can be represented as a column vector 'EC' (Eq. 7). The unit costs of activity drivers $\mathrm{e}_1 - \mathrm{e}_{10}$ are as follows:

$ec_1, ec_2, ec_4 - ec_7, ec_9$: Environmental cost related to the unit of material recycling and carbonization

ec_3: Environmental costs per unit of disposed material due to soil and water contamination and landfill occupation.

ec_8: Cost per unit of substituted natural materials (gravel)

ec_{10}: Cost per unit of substituted natural materials (sand)

$$EC = \begin{pmatrix} ec_1 \\ \vdots \\ ec_{10} \end{pmatrix} \quad (7)$$

Thus, vector 'EEC' consists of items that represent the environmental costs of activities (per unit) computed using Eq. 8:

$$\mathrm{EEC} = E \cdot \mathrm{E}C = \quad (8)$$

$$\begin{pmatrix} e_{11} & \cdots & e_{1\,10} \\ \vdots & \ddots & \vdots \\ e_{71} & \cdots & e_{7\,10} \end{pmatrix} \cdot \begin{pmatrix} ec_1 \\ \vdots \\ ec_{10} \end{pmatrix} = \begin{pmatrix} eec_1 \\ \vdots \\ eec_7 \end{pmatrix}$$

Final product environmental cost (PEC) can be determined by Eq. 9:

$$PEC = Q \cdot \mathrm{EEC} = \quad (9)$$

$$\begin{pmatrix} q_{11} & q_{12} & \cdots & q_{17} \\ q_{21} & q_{22} & \cdots & q_{27} \\ q_{31} & q_{32} & \ddots & q_{37} \\ q_{41} & q_{42} & \cdots & q_{47} \end{pmatrix} \cdot \begin{pmatrix} ec_1 \\ ec_2 \\ \vdots \\ ec_7 \end{pmatrix} = \begin{pmatrix} pec_1 \\ pec_2 \\ \vdots \\ pec_4 \end{pmatrix}$$

3 Case study

The Hong Kong concrete waste market was chosen as a case study due to the city's proactive approach in tackling local challenges. The government has implemented various initiatives aimed at enhancing CDW management and achieving carbon neutrality. Currently, the government's policy direction emphasizes further reduction of CDW and aligns with the global movement towards carbon neutrality. This strategy is reinforced by key government documents, including "Hong Kong's Climate Action Plan 2030+" [25], "Waste Blueprint for Hong Kong 2035" [26], and "The Chief Executive's 2022 Policy Address" [27]. The "Climate Action Plan 2030+" specifically outlines strengthened measures to combat climate change and establishes a target for reducing carbon emissions by 2030.

3.1 Data collection

To illustrate the practicality and feasibility of the proposed matrices, an assessment was conducted for the Hong Kong market, focusing on the Tseung Kwan O Area 137 Construction Waste Sorting Facility (TKO137SF) as a basis for integrating demolished concrete carbonization technology. Decisions and defined variables were quantified using reliable data obtained from a range of credible sources, including sustainability reports from government agencies, Hong Kong standards, a Life Cycle Assessment tool database,

the Chinese life cycle database V0.8, and production specifications for recycled concrete. This information was incorporated into the ABC model matrixes to produce final results and facilitate a comparison of the two scenarios under consideration. The average quantity of incoming demolished concrete was determined based on statistics from the Environmental Protection Department and the Census and Statistics Department. The estimated annual amount of concrete rubble (CR) is 80,002 tonnes [28], assuming that 15% of this recycled material is transformed into WCP [29]. Consequently, the capacities of all considered equipment were calculated based on the average daily and yearly waste turnover. Separate calculations indicate that carbonizing this quantity of CR byproduct requires a total of 8,144 tonnes of CO_2 annually. Additionally, the amounts of substitute cement, gravel, and sand were determined based on standard concrete mix proportions.

3.2 Data processing and discussion

The obtained data is utilized to populate the matrices (B, C, Q, E, and EC), which are necessary for subsequent calculations, ultimately leading to the formation of the 'PC' and 'PEC' matrices. The final matrices associated with the environmental cost of the traditional method are represented in Eq. 10 (PEC_{trad}). In this context, the final product cost matrix (PC_{trad}) is not included, as it only accounts for additional equipment and resources while keeping the existing recycling processes unchanged. Negative numbers represent the loss of CO_2, while positive numbers indicate preserved CO_2. The vector element $'pc_1'$ reflects the recycling of CR through traditional methods, excluding any consideration of additional equipment, which is consequently equal to '0'. The vector element 'pc_2' is quantified at 2.4 million HKD per year, signifying the environmental impact related to soil and water contamination, as well as land occupation in landfills, due to WCP being directed to landfills instead of being recycled. Additionally, the vector element $'pc_3'$ represents the annual CO_2 emissions resulting from cement production that could potentially be mitigated through the use of WCP to capture CO_2 from the cement plant, amounting to 1.5 million HKD. Similarly, 'pc4' illustrates the annual CO_2 emissions associated with the expenditure on natural resources, totalling 125,505 HKD.

$$PEC_{trad} = \begin{pmatrix} 0.00 \\ -2{,}400{,}060 \\ -1{,}495{,}289 \\ -124{,}505 \end{pmatrix} \quad (10)$$

The final matrices for the product cost in the proposed recycling process are represented in Eq. 11, which details expenses related to additional technology integration and energy consumption. The vector element $'pc_1'$, approximately 3,964HKD, indicates the annual amortization cost and energy consumption needed to upgrade the existing equipment for recycling the annual quantity of CR and producing WCP in addition to RCA. This cost is relatively low, as only an additional mobile vibrating screen for sieving is required alongside the current capabilities. The elements $'pc_2'$ and $'pc_3'$ represent the additional equipment amortization costs and energy consumption for the carbonization of RCA and WCP, as well as the flue CO_2 collection, respectively. The energy consumption of the CO_2 collection equipment is quite high (19 kWh/tonne), resulting in an annual cost of nearly 1.5 million HKD.

In terms of environmental impact (Eq.12), CO_2 emissions occur during the processing of CR into RCA and WCP with $'pec_1'$ equaling -736HKD. In the subsequent carbonization stage (with $'pec_2'$ costing 6.3 million HKD per year), the CO_2 absorbed by RCA and WCP significantly outweighs the emissions from the energy used in the process. Additionally, saving landfill space plays a crucial role in enhancing environmental benefits. The vector element $'pec_3'$, approximately 5.1 million HKD per year, represents savings in cement manufacturing due to flue CO_2 collection and the substitution of cement with WCP, which has a considerable environmental impact. The highest figure, 28.3 million HKD per year (represented by 'pec_4'), reflects the conservation of natural materials (gravel and sand) in terms of preserving these resources and avoiding energy expenditure for their processing.

$$PC_{altern} = \begin{pmatrix} 3{,}964 \\ 139{,}540 \\ 1{,}479{,}910 \\ 0.00 \end{pmatrix} \quad (11)$$

$$PEC_{altern} = \begin{pmatrix} -763 \\ 6{,}294{,}267 \\ 5{,}090{,}160 \\ 28{,}275{,}866 \end{pmatrix} \quad (12)$$

Figure 2 illustrates that the annual costs for additional equipment investment and energy consumption in the proposed approach closely match the environmental costs of traditional methods. This alignment indicates that the alternative approach offers significant environmental benefits across all facilities involved. Moreover, by reducing carbon emissions, these facilities can engage in the carbon credit market, allowing them to buy and sell carbon credits. This provides a financial incentive for adopting greener technologies while contributing to sustainability goals. In summary, the proposed approach not only balances economic and environmental costs but also creates opportunities for revenue through carbon credits.

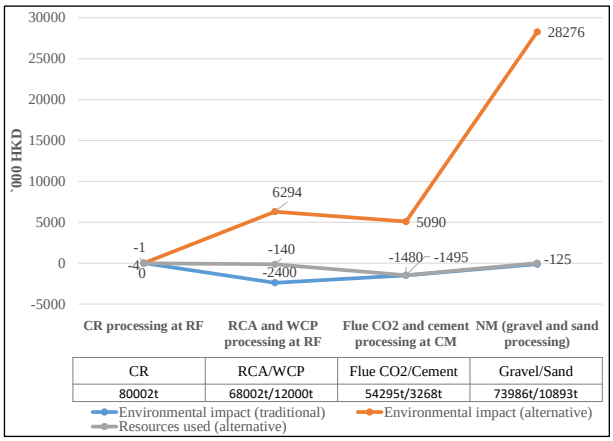


Figure 2. Cost comparison between traditional and proposed approaches.

4 Conclusions

This research explored the practicality and effectiveness of applying active carbonation techniques to WCP and RCA, using flue gases from cement manufacturing as a substitute source of CO_2 for the carbonization process. This innovative approach utilizes flue gases from cement plants for carbonation, significantly reducing the environmental costs associated with cement production and the processing and consumption of natural materials.

The comparison of environmental impacts between traditional and alternative demolished concrete recycling methods (PEC_{trad} and PEC_{altern}) reveals significant differences across all operations in the analyzed facilities. RCA and WCP processing at recycling facilities exhibits a 362% increase from traditional to alternative methods, transforming from -2,400 to 6,294kHKD. The collection of flue CO_2 and cement processing at the cement plant shows an even larger percentage change of 440%, rising from -1,495 to 5,090kHKD. The most significant difference is in natural material processing (gravel and sand), showing an increase of approximately 227kHKD in absolute terms, attributed to the conservation of natural materials and reduced energy expenditure for processing. These significant percentage increases and the shift from negative to positive values in most processes indicate a fundamental difference in environmental impact assessment between the two methodologies. Considering the environmental benefits of the proposed recycling approach that incorporates carbonization technology, the government should implement initiatives to promote this technology further, as it will enhance its ESG evaluation and align with overall policies aimed at reducing carbon emissions.

The universalization of the matrix can be expanded to adapt it to different configurations within the CDW supply chain. However, it's crucial to take into account specific conditions in the recycling market, such as issues related to illegal dumping, which may necessitate modifications to the model. The data were taken as the average quantity of demolished concrete, with other parameters adjusted to simulate and analyze the differences between approaches. The same matrices can be reused. Future research will explore broader implementation challenges, including logistical constraints and site-specific conditions, while examining how the Hong Kong-based model can be adapted to diverse global markets through considerations of varying regulatory frameworks, market maturity, consumer behavior patterns, and regional infrastructure development stages.

Acknowledgment

This work was supported by Guangdong Special Support Talent Program—Innovation and Entrepreneurship Leading Team (2019BT02S593), RGC Research Impact Fund (R7036-22), RGC Theme-based Research Scheme (T32-707-22-N), RGC GRF project (17202124), and PPR (2024.A8.154.24D).

References

[1] Z. Zengfeng, L. Courard, S. Groslambert, T. Jehin, A. Léonard, and X. Jianzhuang, "Use of recycled concrete aggregates from precast block for the production of new building blocks: An industrial scale study," *Resources, Conservation and Recycling,* vol. 157, p. 104786, 2020.

[2] J. Xiao, Y. Xiao, Y. Liu, and T. Ding, "Carbon emission analyses of concretes made with recycled materials considering CO2 uptake through carbonation absorption," *Structural Concrete,* vol. 22, pp. E58-E73, 2021.

[3] D. Oh, T. Noguchi, R. Kitagaki, and H. Choi, "Proposal of demolished concrete recycling system based on performance evaluation of inorganic building materials manufactured from waste concrete powder," *Renewable and Sustainable Energy Reviews,* vol. 135, 2021, Art no. 110147.

[4] S. K. Kaliyavaradhan, T.-C. Ling, and K. H. Mo, "Valorization of waste powders from cement-concrete life cycle: A pathway to circular future," *Journal of Cleaner Production,* vol. 268, 2020, Art no. 122358.

[5] C. S. Poon and D. Chan, "Feasible use of recycled concrete aggregates and crushed clay brick as unbound road sub-base," *Construction and building materials,* vol. 20, no. 8, pp. 578-585, 2006.

[6] F. Kaddah, H. Ranaivomanana, O. Amiri, and E. Rozière, "Accelerated carbonation of recycled

concrete aggregates: Investigation on the microstructure and transport properties at cement paste and mortar scales," *Journal of CO2 Utilization,* vol. 57, p. 101885, 2022.

[7] E. Kravchenko, C. Qin, Z. Lin, and C. W. W. Ng, "Effect of polyvinyl alcohol on the CO2 uptake of carbonated steel slag," *Construction and Building Materials,* vol. 375, p. 130761, 2023.

[8] K. Ouyang, J. Liu, S. Liu, B. Song, H. Guo, G. Li, and C. Shi, "Influence of pre-treatment methods for recycled concrete aggregate on the performance of recycled concrete: A review," *Resources, Conservation and Recycling,* vol. 188, p. 106717, 2023.

[9] S. Jamil, J. Shi, and M. Idrees, "Effect of various parameters on carbonation treatment of recycled concrete aggregate using the design of experiment method," *Construction and Building Materials,* vol. 382, p. 131339, 2023.

[10] J. Zhang, C. Shi, Y. Li, X. Pan, C.-S. Poon, and Z. Xie, "Influence of carbonated recycled concrete aggregate on properties of cement mortar," *Construction and Building Materials,* vol. 98, pp. 1-7, 2015.

[11] S. K. Kaliyavaradhan, L. Li, and T.-C. Ling, "Response surface methodology for the optimization of CO2 uptake using waste concrete powder," *Construction and Building Materials,* vol. 340, p. 127758, 2022.

[12] L. Liu, J. Ha, T. Hashida, and S. Teramura, "Development of a CO2 solidification method for recycling autoclaved lightweight concrete waste," *Journal of materials science letters,* vol. 20, no. 19, pp. 1791-1794, 2001.

[13] Y. Wu, H. Mehdizadeh, K. H. Mo, and T.-C. Ling, "High-temperature CO2 for accelerating the carbonation of recycled concrete fines," *J Build Eng,* vol. 52, 2022, Art no. 104526.

[14] J. Sun, M. F. Bertos, and S. J. Simons, "Kinetic study of accelerated carbonation of municipal solid waste incinerator air pollution control residues for sequestration of flue gas CO 2," *Energy & Environmental Science,* vol. 1, no. 3, pp. 370-377, 2008.

[15] L. Liu, Y. Liu, X. Tian, and X. Chen, "Superior CO2 uptake and enhanced compressive strength for carbonation curing of cement-based materials via flue gas," *Construction and Building Materials,* vol. 346, p. 128364, 2022.

[16] D. Xuan, B. Zhan, and C. S. Poon, "Assessment of mechanical properties of concrete incorporating carbonated recycled concrete aggregates," *Cement and Concrete Composites,* vol. 65, pp. 67-74, 2016.

[17] N. Kemache, L.-C. Pasquier, E. Cecchi, I. Mouedhen, J.-F. Blais, and G. Mercier, "Aqueous mineral carbonation for CO2 sequestration: From laboratory to pilot scale," *Fuel processing technology,* vol. 166, pp. 209-216, 2017.

[18] L. Izoret, T. Pernin, J.-M. Potier, and J.-M. Torrenti, "Impact of industrial application of fast carbonation of recycled concrete aggregates," *Applied Sciences,* vol. 13, no. 2, p. 849, 2023.

[19] Y. Jiang, B. Li, S. Liu, J. He, and A. G. Hernandez, "Role of recycled concrete powder as sand replacement in the properties of cement mortar," *Journal of Cleaner Production,* vol. 371, p. 133424, 2022.

[20] (2004). *Guidelines for Selective Demolition & On Site Sorting*.

[21] S. Besklubova, E. Kravchenko, B. Q. Tan, and R. Y. Zhong, "A feasibility analysis of waste concrete powder recycling market establishment: Hong Kong case," *Environmental Impact Assessment Review,* vol. 103, p. 107225, 2023.

[22] S. Besklubova, B. Q. Tan, R. Y. Zhong, and N. Spicek, "Logistic cost analysis for 3D printing construction projects using a multi-stage network-based approach," *Automat Constr,* vol. 151, p. 104863, 2023.

[23] P. Quesado and R. Silva, "Activity-based costing (ABC) and its implication for open innovation," *Journal of Open Innovation: Technology, Market, and Complexity,* vol. 7, no. 1, p. 41, 2021.

[24] M. Jourdaine, P. Loubet, G. Sonnemann, and S. Trébucq, "The ABC-LCA method for the integration of activity-based costing and life cycle assessment," *Business Strategy and the Environment,* vol. 30, no. 4, pp. 1735-1750, 2021.

[25] EB. "Hong Kong's Climate Action Plan 2030+." www.hkgbc.org.hk/eng/engagement/file/ClimateActionPlanEng.pdf (accessed 1 February 2024.

[26] EB. "Waste Blueprint for Hong Kong 2035." https://www.eeb.gov.hk/sites/default/files/pdf/waste_blueprint_2035_eng.pdf (accessed.

[27] HKSARGovernment. "The Chief Executive's 2022 Policy Address." https://www.policyaddress.gov.hk/2022/public/pdf/policy/policy-full_en.pdf (accessed 31 January 2023.

[28] EPD. "Construction Waste Disposal Charging Scheme." https://www.epd.gov.hk/epd/english/environmentinhk/waste/prob_solutions/background.html#:~:text=l%20The%20Charging%20Regulation%20sets,to%20the%20polluter%20pays%20principle (accessed 1 October 2023.

[29] E. Kravchenko and S. Besklubova, "Value stream assessment of the sustainable concrete recycling process with sequestration of CO2 from flue gases," *Sustainable Production and Consumption,* 2024.

Harnessing Digital Twins for Construction Site Management

Ahmad Al Ayoubi[1] and Ashtarout Ammar[1], PhD

[1]Myers-Lawson School of Construction, Virginia Polytechnic Institute and State University, USA
alayoubi@vt.edu, aammar@vt.edu

Abstract

Digital Twins (DTs) is transforming construction by providing real-time, data-informed solutions that improve project workflows and can aid in achieving decarbonization objectives. In construction, DT was mainly implemented in the operation & maintenance phase. However, by combining DT's functionalities of real-time monitoring, predictive analytics, sophisticated simulations, and collaborative decision-making, construction site management can be revolutionized by enabling site managers and stakeholders to refine construction processes, enhance resource use, lower emissions, boost quality, and reduce expenses. These features tackle essential issues like operational inefficiencies, resource depletion, and communication failures, providing a more integrated and sustainable method for managing construction site activities.

This study uses a Systematics literature review (SLR) to thoroughly examine the function of DTs for onsite construction management, emphasizing their application in managing safety, logistics and material, emissions and energy, quality assurance and detection, real-time decision support, dust, noise, and vibration monitoring, digital collaboration and issue resolution, and real-time worker productivity tracking. Conversely, DTs can facilitate accurate monitoring of carbon emissions and encourage eco-friendly construction methods by integrating technologies such as Internet of Things sensors, cloud computing, and artificial intelligence. Furthermore, DT-driven workflows, including prefabrication and modular building, promote waste reduction and energy conservation, harmonizing onsite activities with net-zero carbon objectives. This research highlights the crucial functions of DTs in optimizing onsite construction workflows, showing their ability to transform the industry and aid in moving towards a more sustainable and decarbonized built environment.

Keywords – Digital-Twins, Decarbonization, Construction Site Management, Construction Planning

1 Introduction

The construction industry is at a pivotal juncture, facing persistent challenges such as inefficiencies, cost overruns, schedule delays, and the need to meet stricter sustainability standards [1]. Despite advances in project management tools and building techniques, construction companies continue to seek new technologies to enhance productivity, achieve sustainability goals, and improve customer satisfaction in an increasingly competitive and complex market [2]. Emerging technologies, particularly DTs, have garnered significant attention as a potential solution to these challenges due to their ability to provide real-time insights and predictive capabilities [3].

DTs present a breakthrough by creating dynamic virtual models of physical construction systems that integrate real-time data, predictive analytics, and advanced simulations. These virtual replicas imitate onsite conditions and processes, enabling stakeholders to monitor material conditions, track equipment performance, and visualize workflows in real time, fostering proactive decision-making and improved quality control [4]. The ability of DTs to merge the physical and virtual environments through data centralization significantly improves communication across stakeholders and reduces errors, aligning construction projects with sustainability and environmental standards [5].

One of the key contributions of DTs is their ability to support sustainability efforts by enabling real-time monitoring of emissions, energy consumption, and waste generation on construction sites. By integrating Internet of Things (IoT) sensors with cloud-based platforms, DTs provide comprehensive feedback loops that ensure construction activities remain aligned with sustainability targets [7, 8]. Moreover, the optimization capabilities of DTs allow teams to simulate and refine workflows before implementation, which is particularly valuable in complex projects where early identification of inefficiencies can save significant time and resources [8]. For instance, DT-enabled simulations of crane operations can help predict potential conflicts in crowded sites, allowing site managers to redesign workflows to prevent delays and accidents [9].

However, despite these promising capabilities, key research gaps remain. There is limited empirical evidence

demonstrating the impact of DT-enabled onsite workflows on decarbonization and operational efficiency. Additionally, the industry lacks standardized methodologies for implementing DTs in onsite construction, as well as case studies that illustrate best practices across various project types and regions. Addressing these gaps is essential for validating the role of DTs in transforming onsite construction.

To bridge this gap and improve the understanding of the potential of DTs for construction site management, this paper examines in detail the critical role of the four DTs' key functions (i.e., real-time monitoring, predictive analytics, simulation, and collaborative decision-making) in managing construction sites by optimizing schedules and enhancing safety. Depending on project site needs, site managers can integrate these functions to create various onsite applications, optimizing efficiency and decision-making [10]. Building on this concept, this study analyzes eight specific applications fundamental for efficient construction site management that result from these functions' combinations: 1) safety management, 2) logistics and material management, 3) emissions and energy management, 4) quality assurance and defect detection, 5) real-time decision support, 6) dust noise and vibration monitoring, 7) digital collaboration and issue resolution, 8) real-time worker productivity tracking.

By examining these four core functions and their eight applications (detailed in Section 3), this paper highlights how DTs can increase efficiency, reduce costs, time, and complexity, improve onsite construction workflows, address persistent industry challenges, and contribute to sustainable project outcomes, particularly in supporting decarbonization efforts.

2 Methodology

2.1 Systematic Literature Review Approach

This study adopts a systematic review methodology guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework to ensure transparency, replicability, and comprehensiveness [11]. The PRISMA methodology was employed to identify, screen, and select relevant academic literature on DTs applications for on-site construction. The review focused on identifying peer-reviewed articles (i.e., peer-reviewed conferences and journals) that provide insights into DTs functionalities such as real-time monitoring, predictive maintenance, workflow optimization, and collaborative decision-making, as adopted from [10].

The systematic search for this study was conducted using two major academic/research databases, Web of Science and Google Scholar, for their extensive coverage of broad search capabilities and coverage of emerging research trends [12, 13]. To capture a comprehensive range of studies, we carefully chose search word strings. These search strings utilized Boolean operators and keywords aligned with the study's focus on DTs for on-site construction; the strings are as follows:
"Digital Twins" And "Construction Quality" Or "Real-Time Monitoring" Or "Predictive Maintenance" Or "Digital Twins" And "Construction" And "Simulation" And "Optimization" Or "Digital Twins" And "Workflow" Or "Digital Twins" And "Stakeholder Collaboration" And "Construction" Or "Construction Quality" Or "Digital Twins Applications" And "Sustainability" Or "Digital Twins" And "Construction" And "IoT Sensors" Or "Digital Twins" And "Construction Safety" Or "Defect Detection" Or "Digital Twins" And "Construction" Or "Material Management" Or "Digital Twins" And "On-Site Construction" Or "Digital Twins Framework" And "Construction Project Management."

These search strings were continuously refined based on the relevance and quality of the initial search results, ensuring a comprehensive and accurate representation of the topic. In addition to the search strings, multiple filters and inclusion/exclusion criteria, as presented in Table 1, were applied to ensure that only relevant papers were considered for a more comprehensive review.

Table 1. Inclusion criteria of the research.

Criteria	Inclusion Criteria
Publication Type	Peer-reviewed journal and conference papers.
Language	Papers published in English.
Timeframe	Studies that were published between 2018 and 2024.

The initial search resulted in 80 peer-reviewed articles extracted from Google Scholar and Web of Science databases. After applying the filters based on the inclusion/exclusion criteria and reviewing the full text to ensure that the extracted studies align with the research objective, 50 journals were considered for in-depth analysis. Figure 1 below outlines the study selection process, showing the breakdown of how many papers were extracted from each database and how they were filtered based on the inclusion/exclusion criteria.

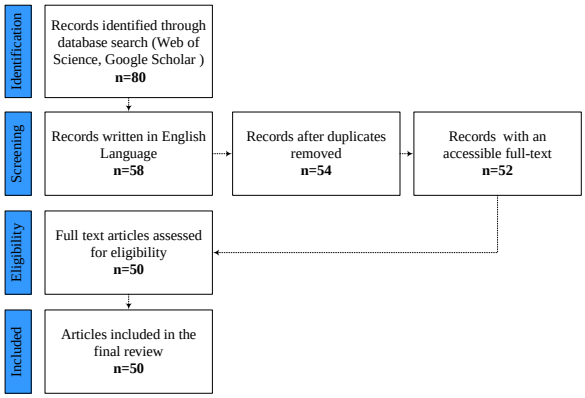


Figure 1. PRISMA Flow Diagram

2.2 Data Extraction and Synthesis

The 50 selected articles were then reviewed carefully, and key points and findings were extracted. The analysis focused on the contributions of DTs for on-site construction, and data was then categorized based on the following areas:

1. General Key Functions.
2. Specific applications.

This methodology provides a detailed account of the systematic review process used, adhering to the PRISMA framework, and highlights the search strings and criteria that were adopted.

3 Digital Twins Functionalities and Applications for Construction Site Management

The findings extracted from the reviewed articles were divided into two sections: the first section consists of the general key functions of DTs. In contrast, the second section consists of the on-site construction applications of DTs. The general key functionalities section goes over how each function works and how they can be applied. In contrast, the application section discusses in detail how when the functions are combined simultaneously, they can influence and improve the on-site construction processes and workflows, affecting the overall quality, efficiency, decarbonization, and sustainability goals.

3.1 General Key Functions of Digital Twins in Construction

3.1.1 Real-Time Monitoring and Feedback

Real-time monitoring is foundational to DTs, providing continuous visibility and insight into the on-site processes. Through IoT sensors and cloud-based platforms, DTs enable real-time tracking of material conditions, structural integrity, and equipment performance, such as cranes [9]. By monitoring performances, management can keep track of CO_2 emissions, keep track of idle time and fuel consumption for machinery, make instant decisions based on real-time data, or backup decisions that might be contradictory, such as switching to electric-powered equipment [14]. Real-time monitoring is essential in every application under the DT umbrella. By using real-time monitoring and feedback, the engineers and site teams can have insights and information from the virtual realm [15]. This empowers the site team to conduct modifications based on informative insights extracted from real-time data without wasting time, effort, money, and human resources, which brings projects closer to sustainability and environmental goals.

3.1.2 Predictive Analytics

Predictive analytics leverage DTs analytics to predict equipment failures, structural degradation, material shortages, and logistic needs. By using AI, DT can analyze historical and real-time data to identify patterns and anomalies that can indicate potential risks and failures, allowing for timely interventions [16]. For instance, vibration data from cranes monitored by DTs can detect stress or wear-and-tear, enabling teams to schedule maintenance before failures occur, extending the lifecycle of equipment, thus reducing the need for replacements and minimizing the environmental impact of manufacturing and transport [7, 17]. This information reflects on the current production flows and is essential for planning and control systems [18]. Currently, there is an underutilization of data, and this allows for missed opportunities to enhance maintenance efficiency and strategy, improving safety, and reducing costs. [19].

By optimizing equipment ‘health,’ machinery will operate more efficiently, consume less fuel, and emit fewer greenhouse gases. Through predictive analytics, DTs aids in improving on-site construction efficiency, reducing construction-related risks, and ensuring the overall quality of the project [20].

3.1.3 Simulation and Optimization

Simulations powered by DTs are invaluable for optimizing on-site construction workflows and validating designs before implementation. Performance simulation is the domain that duplicates and forecasts multiple aspects of construction by applying fundamental physical concepts and engineering methodologies to a computer-based mathematical model [21]. DTs allow teams to test alternative scenarios, identify inefficiencies, and optimize resource allocation. For instance, simulating crane movements on congested construction sites can help identify conflicts, optimize arrangements, enhance operational efficiency while minimizing risks, and reduce delays, as poor scheduling and planning often contribute to budget overruns in construction projects [9].

Simulation tools also drive decarbonization by enabling construction teams to assess the environmental impact of materials and methods before execution. For instance, DTs can simulate workflows and identify opportunities to minimize energy consumption, such as scheduling heavy equipment usage during off-peak energy hours to leverage cleaner electricity.

3.1.4 Collaborative Decision-Making

Collaboration is central to successful construction projects, and DTs facilitate this by providing a unified platform for data sharing and analysis. Cloud-enabled DT systems can accomplish near-real-time data exchange between a digital replica and the physical environment, which can help designers make management site execution decisions. [22].

In the context of decarbonization, collaborative platforms supported by DTs simulate built environmental changes and assist in decision-making for environmental

optimization through platforms, thus reducing the need to visit the site and delaying actions [23].

3.2 Applications for Digital Twins in On-Site Construction

In addition to the general key functions, DTs enables a range of applications in site construction by integrating multiple functionalities simultaneously. Rather than requiring multiple DTs, a single DT equipped with the necessary IoT sensors *(depending on the case and need)* can support eight key applications.

While not exhaustive, these applications represent essential site operations, demonstrating how four fundamental DT functions can drive efficiency, optimization, and sustainability for on-site construction management. By leveraging DTs technology, construction projects can enhance operational workflows while contributing to decarbonization efforts. This section examines eight critical applications where DTs play a transformative role in modern site operations in construction.

3.2.1 Safety Management

Safety management is a top priority in construction, and DTs enhance this by integrating IoT-enabled wearables and real-time monitoring systems. DTs can track worker locations, proximity to hazards, and adherence to safety protocols, preventing accidents and improving compliance [24]. As demonstrated by [25], site management can use drones to monitor potential hazards, allowing them to prevent incidents from occurring. Furthermore, by utilizing real-time analysis, DTs enables site teams in their preparedness to respond to possible accidents [26] properly. This allows the site to not only be prepared in case of emergencies but possibly prevent them altogether.

3.2.2 Logistics and Material Management

DTs streamline material tracking to ensure that resources are utilized efficiently and minimize waste. By monitoring material inventories and deliveries in real-time, DTs enable precise ordering schedules and prevent overstocking or spoilage [27]. Various technologies are adopted into DTs to be able to track materials and stay on top of logistics, such as Radio-Frequency Identification (RFID), which detects items, materials, or targets by utilizing radio frequencies for data processing [28] or Blockchain, which allows for traceability throughout chain activities that are fragmented or isolated [29]. While tracking is fundamental in material management and logistics, other important aspects result from it, such as recycling and waste reduction.

By tracking materials and always knowing the leftovers present on-site from deliveries, excavations, or demolitions, recycling and waste reduction become critical, easier, and needed for better sustainability and environmental metrics [30]. By tracking these activities, DTs can help identify the sources of high embodied carbon emissions (ECE) emissions that are present in materials and products in construction and buildings [31] in real-time [28]. It also includes the activities used to transform these materials for buildings [32]. Through this DTs application, a site can keep track of its material handling and quantity management, resulting in a better ability to manage recycling, reusing, and waste management.

3.2.3 Emissions and Energy Management

As of 2021, carbon emissions are currently on a global rise, with buildings making up to 37% of carbon emissions and 34% of energy consumption, with the construction industry (concrete, steel, and aluminum) occupying 4% and 9%, respectively [33]. These numbers are estimated to rise as the number of floor areas is estimated to rise by 75% between 2020 and 2050, of which 80% will be in developing economies, which is equivalent to the floor area of Paris being constructed every week till 2050 [34].

With the use of DT, reducing energy consumption and carbon emissions has become easier. DTs can calculate shadow areas, allowing project managers to use renewable energy resources (RES) and hybrid systems better and monitor and simulate energy expenditure, thus contributing to reducing the site's dependency on fossil fuels and furthering the decarbonization agenda [35].

Furthermore, DTs can provide transformative solutions by utilizing real-time data, predictive analytics, and sensors such as smart meters to reduce emissions and energy usage and identify areas that require improvement [43, 44]. This helps site management know if something is left operating by mistake, ensure no consumption occurs when not needed, and keep a better eye on energy expenditure.

3.2.4 Quality Assurance and Defect Detection

For quality assurance and defect detection, DTs can delve into details that are not purely construction-based and are often carried out on-site. For instance, DTs can be applied to the welding of large excavation motor arms [38]. Also, DTs can be used to form a model that can adapt to a variety of situations to monitor the quality of spot welding (i.e., structural steel welding) to achieve dynamic control over welding quality and ensure product stability [39].

Early defect detection reduces the material and energy demands associated with rework. By preventing this waste, DTs can directly contribute to lower emissions and resource conservation. By utilizing predictive analytics and real-time feedback, DTs provide the ability to predict the condition of machines and prevent unplanned halts in work. In some cases, such as diesel-operated motors, it can lead to a reduction in air pollutants produced by ill-maintained diesel engines.

3.2.5 Real-Time Decision Support

Decision-making has been revolutionized for on-site applications by DTs. By integrating IoT devices and sensors, DTs allow for a dynamic decision-making environment. By using up-to-date real-time deviations, DTs enhance planning and scheduling and optimize modular construction processes while ensuring seamless integration and minimizing disruption in cases like precast concrete assembly [40]. As mentioned previously in Section 3.2.2, DTs improve material management by providing visibility and traceability of resources [41]. This feature allows site managers to anticipate and resolve material shortages and enable predictive decision-making to forecast risks and bottlenecks to ensure uninterrupted progress [42]. Also, as discussed in section 3.2.4, when DTs provides a prompt to perform maintenance preemptively, it can provide advice on when to perform it and ensure that the site management is well-informed to understand the implications better and take a well-informed approach.

Furthermore, DTs also support the reduction of the carbon footprint by optimizing resource use, reducing waste through real-time decision-making, and performing energy consumption planning to adjust energy use [43].

3.2.6 Dust, Noise, and Vibration Monitoring

Dust, noise, and environmental monitoring are becoming increasingly important. With DTs, site managers can monitor these factors continuously to improve the management of environmental risks. This data helps mitigate health hazards associated with dust exposure, implement noise-reduction strategies, and offer comprehensive monitoring solutions for construction operations to reduce vibrations [44].

In a case study conducted by the authors of [1], DT models were able to control tunnel dust levels, ensure safe working conditions, and adhere to environmental guidelines. These applications underscore DTs' potential in reducing and possibly in the future eliminating environmental impacts and ensuring regulatory compliance with green metrics [37].

3.2.7 Digital Collaboration and Issue Resolution

In construction projects, conflicts often occur. This might be due to miscommunication, design discrepancies, scheduling clashes, or multiple other reasons. Multiple approaches can be used to reduce such conflicts. The authors of [45] showed how DTs could create a shared platform for all stakeholders to access real-time data, reduce misunderstandings, and accelerate the decision-making process, while [46] integrated data mining with building information modelling (BIM) in a DTs platform to enhance project management, and allow for swift changes in the schedule by detecting conflicts and inconsistencies. Similarly, the authors of [47] noted that DTs can foster collaboration between practitioners, stakeholders, and authorities by providing a unified interface for problem-solving and discussions. When such platforms are applied, conflict resolution significantly improves through the decision-making capabilities of DTs by creating an accountable record of decisions and changes [48]. By reducing conflicts, rework also decreases significantly, reduces construction time, and improves construction/project management. This implies that DTs not only improve project outcomes but also significantly contribute to the industry's social and environmental goals [43].

3.2.8 Real-Time Worker Productivity Tracking

DTs are reshaping worker productivity tracking. The authors of [40] highlighted how the use of synchronized DTs for managing workflows in construction ensures worker productivity aligns with project milestones.
DTs can identify worker output inefficiencies by using a cyber-physical system, which allows them to provide actionable insights to improve workforce performance [49].

While some approaches focus on human productivity monitoring, other approaches focus on increasing human productivity via human-robot collaboration to optimize both human and machine productivity [50]. Productivity doesn't always decrease due to uncontrollable events. They might be caused due to fatigue or health issues. To avoid such issues, the authors of [51] discussed how wearable devices can monitor laborers' health and avoid problems that might arise due to such issues.

By offering real-time insights, DTs empower construction managers and superintendents to monitor and improve worker productivity dynamically without sacrificing health, cost, and quality.

4 Results & Conclusion

This study highlights the potential of DT technology in on-site construction, demonstrating how it integrates real-time monitoring and feedback, predictive maintenance, simulation and optimization, and collaborative decision-making to enhance site management, efficiency, and decarbonization efforts. By applying these four key functions using the eight applications discussed in section 3, DT can address safety concerns and maintenance issues, monitor and track workers & worker productivity, enhance decision-making & collaboration, increase quality of work and reduce the effort needed for recycling and material management.

The findings illustrate that DTs play a crucial role in ensuring that site management has a proper hold on what is going on on-site. With DTs receiving real-time feedback, tracking inventory, prompting action when an incident might occur, and tracking workers, site management will be able to manage better and relay

progress to the higher-ups, increasing collaboration efficiency and critical decision-making.

Beyond efficiency gains, DTs significantly contribute to sustainability and decarbonization. By minimizing rework, reducing waste, preventing equipment failure, reducing fossil fuel dependency, and optimizing energy consumption, DTs help construction projects lower carbon emissions and overall environmental impact. Their ability to monitor pollutants, track emissions, and support recycling & material reuse efforts further strengthens their role in site decarbonization initiatives. Figure 2 below illustrates how the four functions are applied to support each of the eight identified specific applications, ultimately contributing to decarbonization efforts in on-site construction.

In essence, DTs represent a major advancement in modern site operations, offering a scalable, data-driven solution to enhance overall productivity. As the industry continues to embrace digitalization and sustainability, DTs will be critical in driving innovation, improving resilience, and shaping a more efficient, collaborative, and eco-friendly built environment. Future research should focus on standardized DTs frameworks to allow for one DT to perform seamlessly between projects and improve interoperability further to enhance automation and predictive capabilities in site construction.

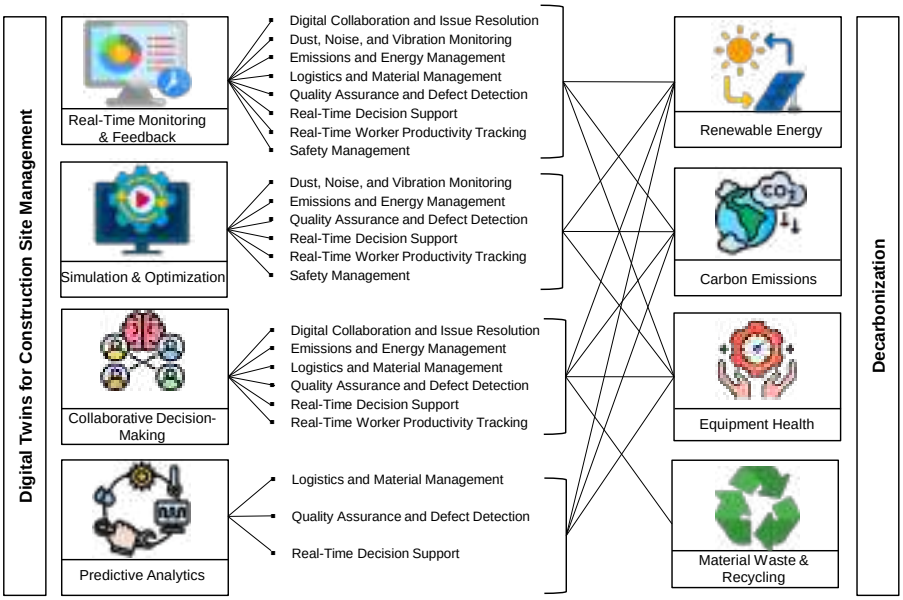


Figure 2. Digital Twins functionalities and applications for construction site management in supporting decarbonization

5 Bibliography:

[1] L. Rigó, J. Fabianová, M. Lokšík, and N. Mikušová, "Utilising Digital Twins to Bolster the Sustainability of Logistics Processes in Industry 4.0," *Sustainability*, vol. 16, no. 6, p. 2575, Mar. 2024, doi: 10.3390/su16062575.

[2] M. Javaid, A. Haleem, R. P. Singh, R. Suman, and E. S. Gonzalez, "Understanding the adoption of Industry 4.0 technologies in improving environmental sustainability," *Sustainable Operations and Computers*, vol. 3, pp. 203–217, 2022, doi: 10.1016/j.susoc.2022.01.008.

[3] C. J. Turner, J. Oyekan, L. Stergioulas, and D. Griffin, "Utilizing Industry 4.0 on the Construction Site: Challenges and Opportunities," *IEEE Trans. Ind. Inf.*, vol. 17, no. 2, pp. 746–756, Feb. 2021, doi: 10.1109/TII.2020.3002197.

[4] T. W. Kang and Y. Mo, "A comprehensive digital twin framework for building environment monitoring with emphasis on real-time data connectivity and predictability," *Developments in the Built Environment*, vol. 17, p. 12, Mar. 2024, doi: 10.1016/j.dibe.2023.100309.

[5] R. Klinc and Ž. Turk, "Construction 4.0 – Digital Transformation of One of the Oldest Industries," *Economic and Business Review*, vol. 21, no. 3, Dec. 2019, doi: 10.15458/ebr.92.

[6] A. Harode, W. Thabet, and P. Dongre, "A tool-based system architecture for a digital twin: a case study in a healthcare facility," *ITcon*, vol. 28, pp. 107–137, Feb. 2023, doi: 10.36680/j.itcon.2023.006.

[7] S. A. Talmaki and V. R. Kamat, "Sensor Acquisition and Allocation for Real-Time Monitoring of Articulated Construction Equipment in Digital Twins," *Sensors*, vol. 22, no. 19, p. 7635, Oct. 2022, doi: 10.3390/s22197635.

[8] M. Afzal *et al.*, "Delving into the Digital Twin Developments and Applications in the Construction Industry: A PRISMA Approach," *Sustainability*, vol.

15, no. 23, Art. no. 23, Jan. 2023, doi: 10.3390/su152316436.
[9] P. Li, J. Xie, J. Ding, Z. Zhao, W. Wu, and G. Q. Huang, "Real-time Locating System-enabled Digital Twin for Crane Operation Safety Monitoring on Construction Sites," in *2024 IEEE 20th International Conference on Automation Science and Engineering (CASE)*, Bari, Italy: IEEE, Aug. 2024, pp. 2689–2696. doi: 10.1109/CASE59546.2024.10711454.
[10] H. Jebelli, S. Asadi, I. Mutis, R. Liu, and J. Cheng, Eds., *Digital Twins in Construction and the Built Environment*. Reston, VA: American Society of Civil Engineers, 2024. doi: 10.1061/9780784485606.
[11] M. Angioi and C. E. Hiller, "Systematic Literature Reviews," in *Research Methods in the Dance Sciences*, T. Welsh, J. P. Ambegaonkar, and L. Mainwaring, Eds., University Press of Florida, 2023, p. 0. doi: 10.5744/florida/9780813069548.003.0018.
[12] V. K. Singh, P. Singh, M. Karmakar, J. Leta, and P. Mayr, "The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis," *Scientometrics*, vol. 126, no. 6, pp. 5113–5142, Jun. 2021, doi: 10.1007/s11192-021-03948-5.
[13] A.-W. Harzing and S. Alakangas, "Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison," *Scientometrics*, vol. 106, no. 2, pp. 787–804, Feb. 2016, doi: 10.1007/s11192-015-1798-9.
[14] E. Negri, L. Fumagalli, and M. Macchi, "A Review of the Roles of Digital Twin in CPS-based Production Systems," *Procedia Manufacturing*, vol. 11, pp. 939–948, 2017, doi: 10.1016/j.promfg.2017.07.198.
[15] Y. Fu, G. Zhu, M. Zhu, and F. Xuan, "Digital Twin for Integration of Design-Manufacturing-Maintenance: An Overview," *Chin. J. Mech. Eng.*, vol. 35, no. 1, p. 80, Dec. 2022, doi: 10.1186/s10033-022-00760-x.
[16] H. H. Hosamo, P. R. Svennevig, K. Svidt, D. Han, and H. K. Nielsen, "A Digital Twin predictive maintenance framework of air handling units based on automatic fault detection and diagnostics," *Energy and Buildings*, vol. 261, p. 111988, Apr. 2022, doi: 10.1016/j.enbuild.2022.111988.
[17] L. M. B. Ventura, F. de O. Pinto, A. Gioda, and M. de A. D'Agosto, "Inspection and maintenance programs for in-service vehicles: An important air pollution control tool," *Sustainable Cities and Society*, vol. 53, p. 101956, 2020, doi: https://doi.org/10.1016/j.scs.2019.101956.
[18] H. G. Ballard, "THE LAST PLANNER SYSTEM OF PRODUCTION CONTROL," 2000.
[19] F. Tao, Q. Qi, A. Liu, and A. Kusiak, "Data-driven smart manufacturing," *Journal of Manufacturing Systems*, vol. 48, pp. 157–169, Jul. 2018, doi: 10.1016/j.jmsy.2018.01.006.
[20] W. Wang *et al.*, "BIM Information Integration Based VR Modeling in Digital Twins in Industry 5.0," *Journal of Industrial Information Integration*, vol. 28, p. 100351, Jul. 2022, doi: 10.1016/j.jii.2022.100351.
[21] P. De Wilde, "Building performance simulation in the brave new world of artificial intelligence and digital twins: A systematic review," *Energy and Buildings*, vol. 292, p. 113171, Aug. 2023, doi: 10.1016/j.enbuild.2023.113171.
[22] G. White, A. Zink, L. Codecá, and S. Clarke, "A digital twin smart city for citizen feedback," *Cities*, vol. 110, p. 103064, Mar. 2021, doi: 10.1016/j.cities.2020.103064.
[23] A. Walters, "The local governance of digital technology – implications for the cityscale digital twin." Accessed: Dec. 24, 2024. [Online]. Available: https://www.cdbb.cam.ac.uk/research/digital-built-environment/local-governance-digital-technology-implications-cityscale
[24] W. Cheng, Y. Pan, Z. Ma, Y. Cai, Y. Li, and T. Peng, "Tunnel Construction Site Monitoring and Digital Twin System," in *2024 26th International Conference on Advanced Communications Technology (ICACT)*, Pyeong Chang, Korea, Republic of: IEEE, Feb. 2024, pp. 389–394. doi: 10.23919/ICACT60172.2024.10471763.
[25] S. Alizadehsalehi, I. Yitmen, T. Celik, and D. Arditi, "The effectiveness of an integrated BIM/UAV model in managing safety on construction sites," *International Journal of Occupational Safety and Ergonomics*, vol. 26, no. 4, pp. 829–844, Oct. 2020, doi: 10.1080/10803548.2018.1504487.
[26] M. Shariatfar, A. Deria, and Y.-C. Lee, "Digital Twin in Construction Safety and Its Implications for Automated Monitoring and Management," in *Construction Research Congress 2022*, Arlington, Virginia: American Society of Civil Engineers, Mar. 2022, pp. 591–600. doi: 10.1061/9780784483961.062.
[27] S. Becker, S. Einizinab, S. Radanovic, K. Khoshelham, K. Mirzaei, and Y. Fang, "REALITY CAPTURE METHODS FOR REMOTE INSPECTION OF BUILDING WORK," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. XLVIII-1/W2-2023, pp. 275–281, Dec. 2023, doi: 10.5194/isprs-archives-XLVIII-1-W2-2023-275-2023.
[28] S. K. Yevu, E. K. Owusu, A. P. C. Chan, S. M. E. Sepasgozar, and V. R. Kamat, "Digital twin-enabled prefabrication supply chain for smart construction and carbon emissions evaluation in building projects," *Journal of Building Engineering*, vol. 78, p. 107598, Nov. 2023, doi: 10.1016/j.jobe.2023.107598.
[29] X. Li *et al.*, "Blockchain-Enabled IoT-BIM Platform for Supply Chain Management in Modular Construction," *J. Constr. Eng. Manage.*, vol. 148, no. 2, p. 04021195, Feb. 2022, doi: 10.1061/(ASCE)CO.1943-7862.0002229.
[30] S. Züst, R. Züst, V. Züst, S. West, O. Stoll, and C. Minonne, "A graph based Monte Carlo simulation supporting a digital twin for the curatorial management of excavation and demolition material flows," *Journal of Cleaner Production*, vol. 310, p. 127453, Aug. 2021, doi: 10.1016/j.jclepro.2021.127453.

[31] M. Robati, P. Oldfield, A. A. Nezhad, D. G. Carmichael, and A. Kuru, “Carbon value engineering: A framework for integrating embodied carbon and cost reduction strategies in building design,” *Building and Environment*, vol. 192, p. 107620, Apr. 2021, doi: 10.1016/j.buildenv.2021.107620.

[32] M. Sandanayake, G. Zhang, and S. Setunge, “Estimation of environmental emissions and impacts of building construction – A decision making tool for contractors,” *Journal of Building Engineering*, vol. 21, pp. 173–185, Jan. 2019, doi: 10.1016/j.jobe.2018.10.023.

[33] I. Hamilton, Harry Kennard, O. Rapf, and J. Kockat, “2022 Global Status Report for Buildings and Construction: Towards a Zero-emission, Efficient and Resilient Buildings and Construction Sector,” United Nations Enivronmental Programme, 2022.

[34] S. Bouckaert, A. F. Pales, and C. McGlade, “Net Zero by 2050 A Roadmap for the Global Energy Sector,” International Energy Agency, 2021.

[35] V. Bocullo *et al.*, “A Digital Twin Approach to City Block Renovation Using RES Technologies,” *Sustainability*, vol. 15, no. 12, p. 9307, Jun. 2023, doi: 10.3390/su15129307.

[36] A. Francisco, N. Mohammadi, and J. E. Taylor, “Smart City Digital Twin–Enabled Energy Management: Toward Real-Time Urban Building Energy Benchmarking,” *Journal of Management in Engineering*, vol. 36, no. 2, p. 04019045, Mar. 2020, doi: 10.1061/(ASCE)ME.1943-5479.0000741.

[37] C. R. Corrado, S. M. DeLong, E. G. Holt, E. Y. Hua, and A. Tolk, “Combining Green Metrics and Digital Twins for Sustainability Planning and Governance of Smart Buildings and Cities,” *Sustainability*, vol. 14, no. 20, p. 12988, Oct. 2022, doi: 10.3390/su142012988.

[38] Q. Zhang, R. Xiao, Z. Liu, J. Duan, and J. Qin, “Process Simulation and Optimization of Arc Welding Robot Workstation Based on Digital Twin,” *Machines*, vol. 11, no. 1, p. 53, Jan. 2023, doi: 10.3390/machines11010053.

[39] J. Dong, J. Hu, and Z. Luo, “Quality Monitoring of Resistance Spot Welding Based on a Digital Twin,” *Metals*, vol. 13, no. 4, p. 697, Apr. 2023, doi: 10.3390/met13040697.

[40] Y. Jiang *et al.*, “Digital twin-enabled real-time synchronization for planning, scheduling, and execution in precast on-site assembly,” *Automation in Construction*, vol. 141, p. 104397, Sep. 2022, doi: 10.1016/j.autcon.2022.104397.

[41] Y. Yang, M. Li, C. Yu, and R. Y. Zhong, “Digital twin-enabled visibility and traceability for building materials in on-site fit-out construction,” *Automation in Construction*, vol. 166, p. 105640, Oct. 2024, doi: 10.1016/j.autcon.2024.105640.

[42] Y. Wei, Z. Lei, and S. Altaf, “An Off-Site Construction Digital Twin Assessment Framework Using Wood Panelized Construction as a Case Study,” *Buildings*, vol. 12, no. 5, p. 566, Apr. 2022, doi: 10.3390/buildings12050566.

[43] H. Omrany, A. Mehdipour, and D. Oteng, “Digital Twin Technology and Social Sustainability: Implications for the Construction Industry,” *Sustainability*, vol. 16, no. 19, p. 8663, Oct. 2024, doi: 10.3390/su16198663.

[44] N. Babazadeh, J. Teizer, H.-J. Bargstädt, and J. Melzner, “Predictive simulation of construction site noise emissions from heavy equipment,” *Smart and Sustainable Built Environment*, vol. ahead-of-print, no. ahead-of-print, Jan. 2024, doi: 10.1108/SASBE-08-2023-0226.

[45] C. Boje, A. Guerriero, S. Kubicki, and Y. Rezgui, “Towards a semantic Construction Digital Twin: Directions for future research,” *Automation in Construction*, vol. 114, p. 103179, Jun. 2020, doi: 10.1016/j.autcon.2020.103179.

[46] Y. Pan and L. Zhang, “A BIM-data mining integrated digital twin framework for advanced project management,” *Automation in Construction*, vol. 124, p. 103564, Apr. 2021, doi: 10.1016/j.autcon.2021.103564.

[47] A. Ammar *et al.*, “Digital Twins in the Construction Industry: A Perspective of Practitioners and Building Authority,” *Front. Built Environ.*, vol. 8, Jun. 2022, doi: 10.3389/fbuil.2022.834671.

[48] D. Lee, S. H. Lee, N. Masoud, M. S. Krishnan, and V. C. Li, “Integrated digital twin and blockchain framework to support accountable information sharing in construction projects,” *Automation in Construction*, vol. 127, p. 103688, Jul. 2021, doi: 10.1016/j.autcon.2021.103688.

[49] A. A. Akanmu, C. J. Anumba, and O. O. Ogunseiju, “Towards next generation cyber-physical systems and digital twins for construction,” *ITcon*, vol. 26, pp. 505–525, Jul. 2021, doi: 10.36680/j.itcon.2021.027.

[50] X. Wang, C.-J. Liang, C. C. Menassa, and V. R. Kamat, “Interactive and Immersive Process-Level Digital Twin for Collaborative Human–Robot Construction Work,” *J. Comput. Civ. Eng.*, vol. 35, no. 6, p. 04021023, Nov. 2021, doi: 10.1061/(ASCE)CP.1943-5487.0000988.

[51] D.-G. J. Opoku, S. Perera, R. Osei-Kyei, and M. Rashidi, “Digital twin application in the construction industry: A literature review,” *Journal of Building Engineering*, vol. 40, p. 102726, Aug. 2021, doi: 10.1016/j.jobe.2021.102726.

A Comprehensive Framework for Evaluating Blockchain-Based Systems: Embodied Carbon Management in Construction Industry

K.L. Kuan[1], Helen H.L. Kwok[1], Jack C.P. Cheng[1], K.L. Hui[1] and Alexis K.H. Lau[1]

[1]The Hong Kong University of Science and Technology, Hong Kong, China
klkuan@connect.ust.hk, hlkwokab@connect.ust.hk, cejcheng@ust.hk, klhui@ust.hk, alau@ust.hk

Abstract

Embodied carbon in the construction industry is a significant contributor to global carbon emissions. With growing emphasis on carbon neutrality and carbon pricing, the need for effective carbon accounting has become increasingly important. However, inefficient carbon information sharing, driven by the complex and fragmented nature of construction supply chains and the complicated quantification process, hinders carbon accounting. A blockchain-based system is believed to offer a potential solution, but limited research exists on its application. Additionally, existing blockchain-based studies often focus on computable metrics during evaluation, overlooking system design objectives and industry stakeholders' needs. To address these gaps, this study proposes a public blockchain-based carbon management system for embodied carbon information sharing and quantification in the construction industry. It also introduces a comprehensive evaluation framework to verify and validate the system. The research's contributions include: (i) exploring the application of blockchain for sharing and quantifying embodied carbon information in construction supply chains, and (ii) introducing a comprehensive evaluation framework to assess the functional suitability, performance efficiency, usability, system effectiveness, and industry impacts of a blockchain-based system.

Keywords –

Embodied carbon, Information Sharing, Blockchain, Evaluation, Verification, Validation

1 Introduction

According to the World Green Building Council, the construction industry accounts for 39% of global energy-related carbon emissions [1]. A building's life cycle comprises four phases: production, construction, operation, and demolition [2]. Among these, embodied carbon—which encompasses all emissions associated with materials and construction processes, from manufacturing to disposal—accounts for 11% of global carbon emissions [1].

During the 26th United Nations Climate Change Conference (COP26) in 2021, a call was made to achieve net-zero emissions for all new and existing assets by 2050 [3]. Additionally, the Intergovernmental Panel on Climate Change (IPCC), in its Sixth Assessment Report, emphasized the significance of carbon pricing as a key tool for cost-effective climate change mitigation [4]. These developments highlight the rising importance of accounting for embodied carbon in the industry.

Many researchers agree that poor carbon information sharing is a significant issue, including inaccurate and insufficiently detailed carbon data [5], [6] and a lack of data transparency and traceability [6], often due to traditional database systems. Blockchain technology, with features such as immutability, traceability, and transparency, has been explored for applications in the construction industry, including information management [7], [8] and carbon management [9], [10]. Consequently, blockchain is believed to offer promising solutions for improving carbon information sharing. While blockchain-based approaches for embodied carbon management have improved data integrity and privacy [11], challenges remain in ensuring the traceability of carbon information sharing and the quantification process

Moreover, existing blockchain-based solutions for the construction industry primarily focus on computable performance metrics (e.g., network latency, system transactions per minute), neglecting usability, adaptability and effectiveness. There is a lack of comprehensive evaluation methods to verify and validate blockchain-based systems, particularly in assessing whether they meet their intended objectives, align with stakeholder needs, and have long-term adoption potential.

This research investigates how blockchain enhances carbon information sharing and quantification, and proposes a comprehensive evaluation framework for the blockchain-based system. Therefore, this research aims

to address the following two research questions:

- How can a blockchain-based carbon management system enhance traceability in information sharing and the quantification process of embodied carbon in the construction supply chain?
- How can a comprehensive evaluation framework be developed to verify and validate a blockchain-based carbon management system?

2 Related Works

2.1 Blockchain applications in construction industry

Unlike traditional databases, where data is controlled and accessed by a central authority, which can lead to potential risks such as data tampering and reduced trust among participants [12] , blockchain is a decentralized ledger technology enabling secure, transparent transactions on a peer-to-peer network. Transactions are validated by consensus mechanisms and linked in blocks with unique hashes [13]. Therefore, the key characteristics of blockchain include transparency, immutability, traceability, and smart contracts executed by self-executing code [14]. Due to its capabilities and feasibility, researchers have begun exploring its potential applications in the construction industry since 2017. Two key areas of focus for blockchain in the construction industry are information management [7], [8] and supply chain management [15], [16]. Several studies have explored the application of blockchain to carbon management. For instance, blockchain has been integrated into certification and supply chains [9] [17], life cycle assessment [18], and embodied carbon estimation [10] and tracking [11]. However, while blockchain has been applied to carbon management in these contexts, existing studies tend to focus on simplified supply chain models. These models overlook the complexities involved in the carbon quantification process within the construction supply chain. This paper expands on previous research by incorporating smart contracts into the quantification process while maintaining data traceability.

2.2 Evaluations of blockchain-based in construction industry

There are several studies related to blockchain-based systems that have been developed in various areas, such as supply chain management. Examples include blockchain-based systems for transforming personal protective equipment supply chain operations [19], facilitating inventory sharing between suppliers and retailers [20], and enhancing product traceability in the healthcare supply chain [21]. These systems are typically evaluated through smart contract validation, which focuses on functionality, error resolution, and transaction simulation, alongside additional assessments such as cost analysis and security evaluations. Security assessments often include factors such as confidentiality, data integrity, availability, non-repudiation, and protection against man-in-the-middle attacks.

In the construction industry, several blockchain-based systems have been proposed with distinct evaluation methods. For example, a study aimed at improving supply chain traceability and information sharing in precast construction focused its evaluation on information sharing management, real-time control of scheduling, information traceability (e.g., number of transactions per minute), and economic performance at a macro level [7]. Another study on cross-border logistics supervision and data sharing in modular construction evaluated product accountability, data traceability (e.g., time to record and trace information), and system security, including data immutability, authentication, and authorization [22]. Additionally, research linking permissioned blockchain to IoT and BIM platforms for off-site production management in modular construction assessed metrics such as storage cost, IoT data transaction integrity, authorization, and authentication [23]. Similarly, a blockchain-based framework for carbon management toward construction product certification was evaluated based on security, performance, blockchain network latency, and cost [17].

However, their evaluations often focus on computable metrics without addressing other important dimensions, such as the usability and effectiveness of the systems. This reliance on traditional metrics fails to capture the holistic value of blockchain systems and how they address industry-specific challenges inherent in their design. This study expands on previous research by adopting a more comprehensive evaluation approach for evaluating the performance and effectiveness of the proposed blockchain-based system.

3 Methodology

This study employs the Design Science Research (DSR) methodology [24], a structured approach that supports systematic problem-solving activities in research. The methodology begins with the identification of research problems and objectives, accomplished through a comprehensive literature review of existing studies on blockchain-based applications and their evaluation methodology, as identified in Section 2.

Following this, the system framework of the blockchain-based carbon management system and the comprehensive evaluation framework for blockchain-based systems were designed and developed. Subsequently, a system prototype will be created based

on the designed framework and demonstrated through a real-life case study in the construction industry to validate its feasibility. The prototype will then be rigorously evaluated using the proposed comprehensive evaluation framework to verify and validate how well it meets its design objectives and addresses the identified challenges. Finally, both the system framework and the evaluation framework underwent practical assessment through structured communication with relevant stakeholders.

3.1 Design and Development

3.1.1 System Framework

Referring to the previous conference paper [11], the study focuses on the integration of blockchain technology with carbon management across the entire construction supply chain to enhance the quality and traceability of embodied carbon data. A system framework and operational workflow were proposed. In the framework, the system consists of traditional off-chain carbon management functions, encompassing user registration and data input, and a decentralized public database (on-chain) with self-executing smart contracts is utilized to store significant project information and carbon emissions data to make carbon footprints transparent, reliable, and trackable. The stakeholders involved are developers, architects, consultants, contractors, and product suppliers. Carbon data input parameters are divided into three groups based on the building life cycle: Materials in New Construction (A1–A3), Upstream Transportation (A4), and Construction Site Activities (A5). Besides, the study introduced three technical highlights: (i) Role-Based Access Control for Functionality, allowing every stakeholder to have a clearly defined role and functionality within the system, (ii) Hierarchical Hashing Strategy for Validation, enabling multiple validations by various parties to ensure data quality before data is stored on-chain, and (iii) Selective Data Disclosure Before On-chain, allowing input parties the right to select the data to be disclosed or protected before uploading it to the blockchain and sharing it with the public.

The proposed system framework improves carbon information sharing by enhancing data integrity, transparency, and privacy through recording carbon data on-chain. However, the embodied carbon quantification process, which is integral to information sharing, is a complex task that cannot be accomplished solely by a single recording smart contract. To address this challenge, the smart contract functionality has been advanced to incorporate both recording and quantification capabilities.

The Smart Contract - Carbon Quantification & Recording functionality quantifies total carbon emissions while ensuring that key carbon information is accurately recorded and shared on the blockchain. It also allows stakeholders to protect sensitive information by recording only its corresponding hash value on-chain. The smart contract owners can be suppliers (responsible for permanent materials and upstream transportation emissions) or contractors (responsible for temporary materials and site activity emissions). To ensure accuracy and integrity, all data must be validated by designated parties, while input parties retain the ability to select which parameters are disclosed on-chain. Inputs to the system include, but are not limited to, product specifications, emission factors, quantities, and transportation distances. Outputs generated by the system encompass key carbon information (e.g., emission source information, emission factor, transportation distance), hash values of carbon and validation process data, and total project emissions. Both input and output data are combined to form a transaction that is recorded on the blockchain, ensuring immutability, traceability, and transparency for the project team and the public. Figure 1 presents the updated system framework.

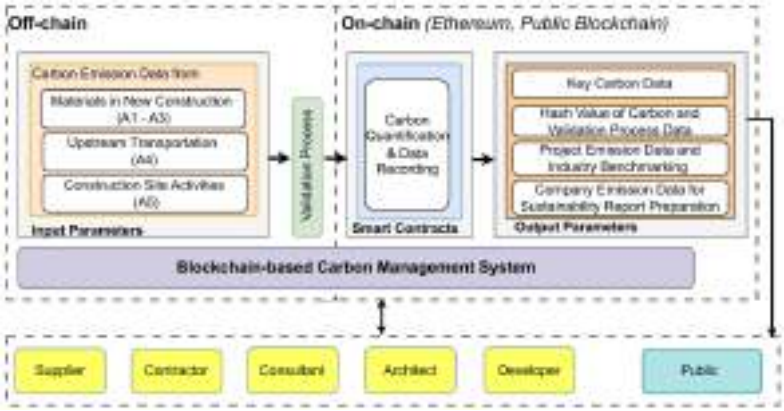


Figure 1: System framework

The updated operational workflow (Figure 2) begins with the architect creating a project account, inputting project details, and adding stakeholders. Suppliers and contractors then input carbon data, which senior staff review and validate before submitting it for further

validation. Once validated, senior staff confirm the data for blockchain upload and select key data for public disclosure before invoking the smart contract. Finally, the developer quantifies and records the total project emissions, holding the final authority in the process.

In summary, the proposed blockchain-based carbon management system enhances embodied carbon information sharing by improving data integrity, traceability, and transparency while preserving privacy. It also leverages smart contracts to automate and streamline the quantification process, ensuring accurate and reliable carbon data management.

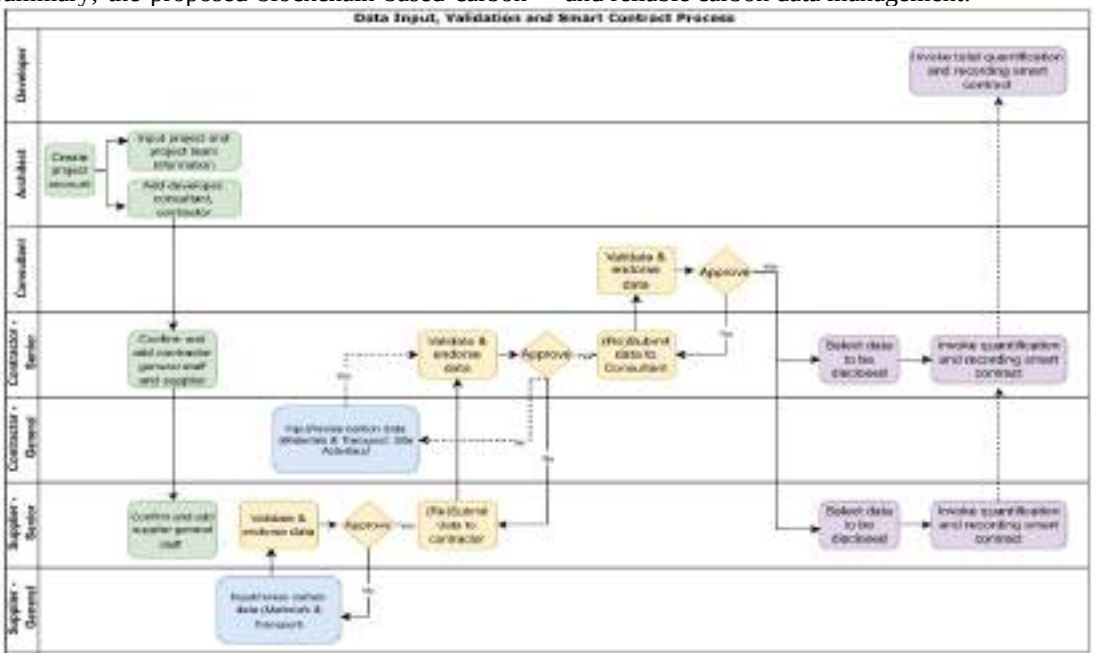


Figure 2: Operational workflow (Green - User/Project Initialization; Blue – Data Input; Orange – Validation Process; Purple - Smart Contracts)

3.1.2 Evaluation Framework

In DSR methodology, evaluation focuses on measuring performance, collecting feedback, and determining how closely the artifact aligns with its stated objectives. This typically includes verifying how well the artifact meets its specified design objectives (verification) and validating whether it effectively solves the identified problem and meets stakeholder needs (validation).

ISO 29119 [25] is a software testing standard that outlines key concepts of software testing, including test processes and techniques. It defines eight testing types: functional suitability, performance efficiency, compatibility, usability, reliability, security, maintainability, and portability. For the verification and validation of the proposed blockchain-based system, three testing types have been identified as priorities: functional suitability, performance efficiency, and usability. These testing types evaluate how well the system addresses its core functionalities, ensures efficient operation, and is practical for adoption by stakeholders. On the other hand, the proposed systems are usually a standalone solution specifically designed for blockchain-based deployment, making compatibility and portability testing less relevant at this stage. Furthermore, reliability testing and security may not prioritized in this phase of development due to the inherent resilience and cryptographic safeguards provided by the public blockchain platform, which inherently ensures data integrity and fault tolerance. Similarly, maintainability is not a focus, as the system remains in the prototype development phase.

For functional suitability testing, the functionality of the system should first be clearly identified. Blockchain-based systems are typically dual-systems, comprising a traditional management system (off-chain) and a decentralized database (on-chain). Each component serves distinct purposes with specific objectives and enhancements, making it essential to evaluate functionality across these two integrated parts. The functionality of the system can be broadly categorized into two areas: (i) Core System Functions and (ii) Functional Enhancement Features. Core System Functions include fundamental operations such as project and project team registration in the off-chain system, as well as data recording and the execution of smart contracts in the on-chain system. These functions are essential for the system's basic operation and compliance with its design objectives. Functional Enhancement Features encompass advanced strategies and mechanisms designed to address critical challenges, such as concerns related to data integrity, traceability, and privacy.

For Performance Efficiency, the testing attributes

typically include scalability, cost, response time, and transaction latency. These attributes are critical for evaluating the system's ability to handle increasing workloads efficiently, maintain operational cost-effectiveness, and deliver acceptable performance under varying conditions. However, the relevance and emphasis of these attributes may vary depending on the specific characteristics and nature of the blockchain-based system being tested. For example, systems with concerns about response time and transaction latency would prioritize these attributes due to the need for immediate updates and decision-making.

Moreover, as noted earlier, the proposed blockchain-based systems often incorporate specific enhancement features aimed at addressing various challenges and achieving distinct objectives. These features, such as enhanced transparency, traceability, and data integrity, are designed to meet the unique demands of their application contexts. Consequently, system effectiveness should be one of the key metrics in the evaluation framework, as it is important for assessing whether the system successfully fulfills its intended objectives.

Table 1: Proposed evaluation framework for the blockchain-based system

Stage		Test attributes	Test Techniques	Test method	Participants
Verification: Meets design objectives		• Functional Completeness • Functional Correctness • Compliance and Non-Compliance Conditions	• Scenario testing	• Prototype Testing	• System designer
Validation: Solves problem and meets needs	(i) Performance Efficiency	• Scalability, cost, response time, and transaction latency (varies in relevance depending on the characteristics and nature of the system being tested)	• Worst-case Scenario Analysis	• Case-based simulation	• System designer
	(ii) Usability	• Functional Appropriateness • User Satisfaction • User Experience • Learnability • User Error Protection	• Use Case Testing	• Surveys (quantitative & qualitative)	• Testing parties
	(iii) System Effectiveness	• Subject to specific enhancement features of the system being tested	• Use Case Testing Comparison with Existing Tools	• Simulation • Sensitivity analysis • Surveys (quantitative & qualitative)	• System designer • Testing parties
	(iv) Industry impacts	• Adoption Challenges • Short-term and Long-term Benefits/Changes (which vary depending on the objectives of the system being tested)	• Impact Assessment • User acceptance testing	• Surveys (quantitative & qualitative) • Semi-structured interviews	• Testing parties • Stakeholders in industry

In addition, industry impacts should be incorporated into the evaluation framework, encompassing both short-term and long-term effects, as well as the adoption of the blockchain-based system within the industry. Since the proposed blockchain-based systems often represent significant innovations, immediate evaluation results from functional, performance, and usability testing may not fully capture their overall value. Assessing industry impacts provides valuable insights into how the system contributes to broader industry goals over time. This dimension is particularly important for understanding the system's potential to influence industry standards and practices, especially regarding long-term development.

Based on the above, a comprehensive evaluation framework for the blockchain-based system has been developed and is divided into two primary stages, as outlined in Table 1: verification and validation. The verification stage focuses on ensuring that the system meets its design specifications and objectives by evaluating attributes such as functional completeness, functional correctness, and compliance through scenario-based prototype testing. The system designer, who possesses the most in-depth knowledge of the system's architecture, functionality, and expected outcomes, serves as the participant in this process.

The validation stage assesses the system's ability to

address the identified problem and fulfill stakeholder needs. It is divided into four dimensions: (i) performance efficiency, which evaluates scalability, cost, response time, and transaction latency through worst-case scenario analysis and case-based simulation, varies in relevance depending on the characteristics and nature of the blockchain-based system; (ii) usability, which tests functional appropriateness, user satisfaction, user experience, learnability, and error protection through case testing and surveys with participants ; (iii) system effectiveness, which is evaluated based on specific enhancement features using use case testing and comparisons with existing tools, both in quantitative approaches (e.g., simulation models and sensitivity analysis) and qualitative approaches (e.g., surveys); and (iv) industry impacts, which assess adoption challenges as well as short-term and long-term benefits, are evaluated through impact assessments, user acceptance testing, surveys, and semi-structured interviews. Testing parties and industry stakeholders should be included in the interviews to ensure more representative results and enhance the potential for practical implementation.

This comprehensive framework provides a robust evaluation of blockchain-based systems by assessing both technical performance and practical usability to ensure the systems meet their design objectives, deliver short-term changes and long-term value to stakeholders.

4 Implementation and results

4.1 System Framework

A prototype was developed using Ethereum for blockchain deployment and Etherscan for monitoring transactions. Figure 3 shows the prototype's user interface, and Figure 4 displays an example of data uploaded in the blockchain transaction.



Figure 3: User interface of the prototype

- *Project Background:* This module captures key project information, including project details and team member information.
- *Create and View Upfront Carbon:* This module supports the input and visualization of data related to the substructure and superstructure phases, managing materials, specifications, transportation, quantities, emissions, and site activities to ensure accurate upfront carbon recording.
- *Blockchain Integration:* This module manages the blockchain-related aspects of the system, including transaction execution, the recording of hash values, and ensuring data traceability, linking off-chain data while preserving carbon information integrity.

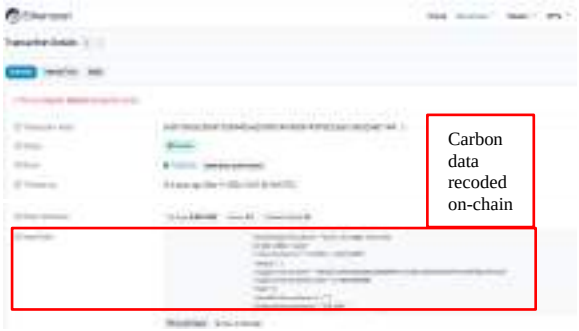


Figure 4: An example of Etherscan transaction

4.2 Evaluation Framework

Referring to Section 3.1, the proposed blockchain-based carbon management system consists of both off-chain and on-chain components. Its objective is to support embodied carbon information sharing and the quantification process in the construction industry, with enhancements in data integrity, transparency, traceability, and privacy achieved through three highlighted approaches and strategies. Based on the characteristics and nature of the proposed system, corresponding verification items based on its functionalities have been identified and summarized in Table 2, while the framework for validation is summarized as follows.

- *Performance Efficiency*: Scalability, cost estimation and transaction latency are considered. Scalability is a critical issue for the application and adoption of blockchain-based systems, ensuring the system can handle increasing data volumes and transaction loads effectively. Cost estimation addresses stakeholder concerns and ensures the system's long-term feasibility, while latency is a key factor in real-time performance.
- *Usability*: The testing parties are divided into three groups: stakeholders involved in the input and validation process (e.g., suppliers, contractors, and consultants), stakeholders responsible for the total project carbon quantification process (e.g., developers), and other stakeholders (e.g., architects and the public). These groups are assessed using data collection methods such as the System Usability Scale, task completion times, and error rates in surveys.
- *System Effectiveness*: This focuses on evaluating

the system's ability to meet its objectives, particularly in enhancing data integrity, transparency, traceability, and privacy, both through quantitative approaches (e.g., simulation models and sensitivity analysis) and qualitative approaches (e.g., surveys).

- *Industry Impacts*: This evaluates adoption challenges, as well as the short- and long-term impacts of the proposed system on carbon management practices in the construction industry. The impact assessment may include performance indicators related to short-term impacts (e.g., behavior change scales, level of stakeholder collaboration and data accessibility) and long-term industry impact (e.g., sustainability impact, adoption motivations and barriers at the project, company, and industry levels). The interviews will include testing parties and key stakeholders directly affected by the system (e.g., contractors and developers), as well as industry policy bodies (e.g., councils). This approach ensures representative results, enhances practical implementation potential, and provides a roadmap for long-term adoption.

Table 2: Verification Framework Summary for the Proposed Blockchain-Based Carbon Management System

	Verification items		Off-chain	On-chain
Core System Functions	i) Information Sharing and Reporting	Data Collection (input parameters)	✓	
		Data Reporting (output parameters)		✓
	ii) User Authentication and Authorization	Role-based Access Control Approach	✓	
		Authentication and Authorization	✓	✓
	iii) Smart Contract	Quantification and Recording		✓
Functional Enhancement Features	i) Data Integrity	Hierarchical Hashing Validation Strategy	✓	
		Immutable Records of Validation Process Data		✓
	ii) Transparency	Open Access to Data		✓
	iii) Traceability	Transaction Logs		✓
		Hashed Detailed Data and Validation Records		✓
	iv) Privacy	Selective Data Disclosure before Chain		✓

5 Conclusion

Embodied carbon in construction significantly contributes to global emissions, but poor information sharing hinders its accounting. Blockchain offers a solution, but is underexplored. Also, most studies focus on computable metrics during evaluation, neglecting critical aspects of how blockchain systems align with their design objectives. This study proposes a public blockchain-based system to enhance carbon information sharing in construction supply chains and develops a comprehensive evaluation framework. This study uses the DSR approach, including the identification of problems and research objectives, the design and development of the system, and its comprehensive evaluation framework .

The contributions of this study: First, it explores how a public blockchain-based carbon management system supports the embodied carbon information sharing and quantification process in the construction industry. Second, it introduces a novel evaluation framework that extends beyond traditional metrics, offering a holistic approach to assess functional suitability, performance efficiency, usability, system effectiveness, and industry impacts. A key research limitation is the reliance on stakeholder participation, which is essential for successful system implementation and evaluation but requires additional time and resources, restricting adoption without incentive mechanisms. Future research will focus on developing a framework to utilize the proposed system to support green finance and incentivize adoption within the construction industry.

Acknowledgments

The work was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region (Project No. T31-603/21-N).

References

[1] "Embodied Carbon," World Green Building Council. Accessed: May 10, 2023. [Online]. Available: https://worldgbc.org/advancing-net-zero/embodied-carbon/

[2] M. N. Nwodo and C. J. Anumba, "A review of life cycle assessment of buildings using a systematic approach," *Building and Environment*, vol. 162, p. 106290, Sep. 2019.

[3] T. C. Group, "COP26 & the Future of the Construction Industry | Concord Blog," The Concord Group. Accessed: May 23, 2024. [Online]. Available: https://concord-cc.com/blog/cop26-and-future-of-construction/

[4] Intergovernmental Panel On Climate Change (Ipcc), *Climate Change 2022 – Impacts, Adaptation and Vulnerability: Working Group II Contribution to the Sixth Assessment Report of the*

Intergovernmental Panel on Climate Change, 1st ed. Cambridge University Press, 2023.

[5] C. De Wolf, F. Pomponi, and A. Moncaster, "Measuring embodied carbon dioxide equivalent of buildings: A review and critique of current industry practice," *Energy and Buildings*, vol. 140, pp. 68–80, Apr. 2017.

[6] V. Venkatraj and M. K. Dixit, "Challenges in implementing data-driven approaches for building life cycle energy assessment: A review," *Renewable and Sustainable Energy Reviews*, vol. 160, p. 112327, May 2022.

[7] Z. Wang, T. Wang, H. Hu, J. Gong, X. Ren, and Q. Xiao, "Blockchain-based framework for improving supply chain traceability and information sharing in precast construction," *Automation in Construction*, vol. 111, p. 103063, Mar. 2020.

[8] B. Zhong, H. Wu, L. Ding, H. Luo, Y. Luo, and X. Pan, "Hyperledger fabric-based consortium blockchain for construction quality information management," *Front. Eng. Manag.*, vol. 7, no. 4, pp. 512–527, Dec. 2020.

[9] K.-H. Liu, S.-F. Chang, W.-H. Huang, and I.-C. Lu, "The Framework of the Integration of Carbon Footprint and Blockchain: Using Blockchain as a Carbon Emission Management Tool," in *Technologies and Eco-innovation towards Sustainability I*, A. H. Hu, M. Matsumoto, T. C. Kuo, and S. Smith, Eds., Singapore: Springer Singapore, 2019, pp. 15–22.

[10] M. N. N. Rodrigo, S. Perera, S. Senaratne, and X. Jin, "Systematic development of a data model for the blockchain-based embodied carbon (BEC) Estimator for construction," *ECAM*, vol. 29, no. 9, pp. 3311–3330, Nov. 2022.

[11] K. L. Kuan, Y. Xu, C. Zheng, J. S. M. Chan, J. C. P. Cheng, and A. K.-H. Lau, "A Blockchain-Based Approach For Embodied Carbon Management Along The Construction Supply Chain," presented at the EC3 Conference 2024, in Computing in Construction, vol. 5. European Council on Computing in Construction, 2024, pp. 0–0.

[12] S. Islam and K. U. Apu, "Decentralized Vs. Centralized Database Solutions In Blockchain: Advantages, Challenges, And Use Cases," *Global mainstream journal of innovation, engineering & emerging technology*, vol. 3, pp. 58–68, Aug. 2024.

[13] D. J. Scott, T. Broyd, and L. Ma, "Exploratory literature review of blockchain in the construction industry," *Automation in Construction*, vol. 132, p. 103914, Dec. 2021.

[14] "Blockchain for Social Impact - Moving Beyond the Hype," Business and Human Rights Gateway. Accessed: Mar. 14, 2025. [Online]. Available: https://humanrights.wbcsd.org/project/blockchain-for-social-impact-moving-beyond-the-hype/

[15] X. (Alice) Qian and E. Papadonikolaki, "Shifting trust in construction supply chains through blockchain technology," *ECAM*, vol. 28, no. 2, pp. 584–602, May 2020.

[16] J. H. Yoon and P. Pishdad-Bozorgi, "State-of-the-Art Review of Blockchain-Enabled Construction Supply Chain," *J. Constr. Eng. Manage.*, vol. 148, no. 2, p. 03121008, Feb. 2022.

[17] Y. Xu *et al.*, "A blockchain-based framework for carbon management towards construction material and product certification," *Advanced Engineering Informatics*, vol. 61, p. 102242, Aug. 2024.

[18] A. Zhang, R. Y. Zhong, M. Farooque, K. Kang, and V. G. Venkatesh, "Blockchain-based life cycle assessment: An implementation framework and system architecture," *Resources, Conservation and Recycling*, vol. 152, p. 104512, Jan. 2020.

[19] I. A. Omar, M. Debe, R. Jayaraman, K. Salah, M. Omar, and J. Arshad, "Blockchain-based Supply Chain Traceability for COVID-19 personal protective equipment," *Computers & Industrial Engineering*, vol. 167, p. 107995, May 2022.

[20] I. A. Omar, R. Jayaraman, M. S. Debe, H. R. Hasan, K. Salah, and M. Omar, "Supply Chain Inventory Sharing Using Ethereum Blockchain and Smart Contracts," *IEEE Access*, vol. 10, pp. 2345–2356, 2022.

[21] A. Musamih *et al.*, "A Blockchain-Based Approach for Drug Traceability in Healthcare Supply Chain," *IEEE Access*, vol. 9, pp. 9728–9743, 2021.

[22] L. Wu, X. Li, R. Zhao, W. Lu, J. Xu, and F. Xue, "A blockchain-based model with an incentive mechanism for cross-border logistics supervision and data sharing in modular construction," *Journal of Cleaner Production*, vol. 375, p. 133460, Nov. 2022.

[23] L. Wu, W. Lu, F. Xue, X. Li, R. Zhao, and M. Tang, "Linking permissioned blockchain to Internet of Things (IoT)-BIM platform for off-site production management in modular construction," *Computers in Industry*, vol. 135, p. 103573, Feb. 2022.

[24] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A Design Science Research Methodology for Information Systems Research," *Journal of Management Information Systems*, vol. 24, no. 3, pp. 45–77, Dec. 2007

[25] C. Patrício, R. Pinto, and G. Marques, "A Study on Software Testing Standard Using ISO/IEC/IEEE 29119-2: 2013," in *Recent Advances in Intelligent Systems and Smart Applications*, M. Al-Emran, K. Shaalan, and A. E. Hassanien, Eds., Cham: Springer International Publishing, 2021, pp. 43–62.

A Comprehensive Approach for Automated Anomaly Detection and Enhancement of EPC Datasets for Decarbonisation

Sahar Mirzaie and Zhihao Zhang

Institute for Sustainable Built Environment, Heriot-Watt University, UK
s.mirzaie@hw.ac.uk, zz5001@hw.ac.uk

Abstract

Energy Performance Certificates (EPCs) are pivotal for evaluating building energy efficiency, informing decarbonisation policies, and supporting retrofit. However, errors in EPC datasets compromise their reliability. This study develops an automated anomaly detection framework integrating Machine Learning (ML) and auxiliary datasets to enhance EPC data accuracy at scale. By reviewing validation methods and identifying gaps, this paper proposes a systematic framework for improving data quality assurance, reinforcing EPCs as a key resource for policy and decarbonisation efforts.

Keywords –

Energy Performance Certificates (EPCs), Machine Learning, Data Validation, Automating Quality Assurance, Decarbonisation

1 Introduction

Energy Performance Certificates (EPCs) have emerged as a cornerstone of energy efficiency policy across Europe [1], providing standardised metrics to evaluate the energy performance of buildings. Introduced in 2007, EPCs facilitate the monitoring of policy progress, guide retrofitting initiatives and inform decisions regarding financial incentives. In the United Kingdom (UK), an EPC is mandatory when a building is constructed, sold, or rented, are required for various funding schemes (e.g. the Green Deal) and remains valid for ten years. With 27 million domestic EPCs in the UK, these datasets offer key insights into national building energy performance [2].

In the UK, the Minimum Energy Efficiency Standards (MEES), introduced in 2018, prohibit private landlords from renting properties with an EPC rating of F or G, with the effectiveness of MEES heavily reliant on EPC accuracy [2,3]. Stock models use adapted EPC calculations to estimate input uncertainties and energy efficiency policy impact at scale [4].

Despite their extensive use, EPCs have faced criticism due to frequent errors and inconsistencies, such as discrepancies in assessments of the same property [3,5,6]. These issues undermine their credibility as policy tools and limit their application in areas such as retrofitting and carbon reduction planning. Addressing these challenges requires the implementation of rigorous validation and editing methodologies to ensure data accuracy and reliability.

Several studies have examined the accuracy of average assessment using the UK EPC dataset, as summarised in Table 1. They acknowledge that the results are not necessarily applicable to a single EPC. Yuan and Choudhary [3] compared the original and renewed EPCs of office buildings with small cooling and heating systems published between 2010 and 2020, looking at changed asset rating using clustering analysis. The non-domestic EPCs offer little information about the building's physical conditions, therefore they assumed a change in the recommendations reflects a change in the building conditions. They conclude that buildings with unchanged conditions can obtain a decreased or increased EPC rating in subsequent assessments.

With residential buildings responsible for a significant share of carbon emissions and the majority of the 2050 building stock already in place, ensuring the accuracy and reliability of EPCs as a tool for evaluating progress for decarbonisation of the domestic building stock is vital. Therefore, this work focuses on domestic EPCs.

To produce an EPC for a residential building, certified energy assessors collect data and, in many cases, make assumptions about the building's energy-related features, such as construction materials, heating systems, and insulation. This information is processed using the Government's Standard Assessment Procedure (SAP) for new builds or the reduced Standard Assessment Procedure (rdSAP) for existing

homes. The SAP algorithm models the building in a steady-state condition to calculate theoretical heat loss, using estimated thermal properties, standardised heating patterns and occupancy, and fixed fuel price assumptions. It then assigns an Energy Efficiency Rating (EER) on a 100-point scale, based on energy costs per square metre, along with an associated EPC band ranging from A (most efficient) to G (least efficient) [2].

SAP and rdSAP are set to be replaced by the Home Energy Model (HEM) in the second half of 2026. Unlike SAP's steady-state modelling approach, HEM will utilise dynamic modelling and updated methodologies to provide a more accurate representation of real-world energy usage, carbon emissions and heating demand [7]. Calculation methodologies change over time, and assessors use different software and have different approaches to assessment. All these cause variations in the results [5,8]. This study however is not concerned with the causes of these variations but rather identifying potential EPC errors for a dwelling. Comparison of the gap between the modelling results and actual energy use of the building is beyond the scope of this work. That said, similar to other studies, it is assumed that the dwelling's rating should not change in repeated assessments unless there has been a change to the fabric or systems, reflected in their recommendations [3,8].

This study evaluates existing EPC validation methods, focusing on their strengths and limitations, particularly for enabling automated EPC data applications. It explores the integration of auxiliary datasets and machine learning (ML) techniques to enhance the automation of validation processes and data quality. While much prior research has concentrated on repeated EPC assessments [3,5,8,9], this paper adopts a broader approach by examining variables and inputs involved in generating a single EPC as well. Through this wide-ranging analysis, this study proposes a comprehensive framework to improve automation of EPC data quality assurance, contributing to the enhancement of EPC reliability, thereby supporting their effective application in energy efficiency policies and broader research contexts.

2 Literature Review

To understand how EPCs have been validated a comprehensive literature review of Scopus database was conducted, using 'Energy Performance Certificate' or 'EPC' in the title, with a particular focus on domestic properties and validation techniques. Additional studies were incorporated using the 'snowball' approach. The search results indicated that relevant publications on this topic emerged from 2016 onwards.

Table 1. Overview of EPC Validation Studies in the UK: Key Characteristics, Approaches, and Analytical Methods

Ref.	Year	Country	Domestic	Sample Size	Sample Period	EPC values	Validation Approach	Validation levels	Statistical or Machine Learning Approach
[3]	2023	England and Wales	N	5400 simple office buildings	2010-2020	Asset Rating, Standard Emission Rate (SER), Building Emission Rate (BER), Recommendations	Changed asset rating using clustering analysis.	2	Local weighted polynomial regression to study the distribution. Clustering analysis for association pattern recognition: GMM.
[5]	2017	England and Wales	D	145 audits of 29 dwellings	2016	EER, age, construction, and property type, physical properties such as floor area, fabric and systems efficiency	At least 4 independent energy assessments	3	-
[8]	2019	England and Wales	D	1.6 million EPCs from 13 million dwellings	2009-2016	EER	Repeated EPC assessments and change in EER. Exploratory data analysis to decide how to apply this model to available data.	2	A semi-parametric statistical model of how measurement error contributes to variation between repeated measurements
[9]	2019	England and Wales	D	8324 dwellings with at least 3 EPCs for RF 9772 dwellings with at least 3 EPCs at the postcode for RF	2008-2016	EER, inspection & lodgement dates, physical properties such as built form, floor area, flat floor level, fabric description, glazing type	EPCs for the same properties and inconsistent data across properties with same postcode.	1,2	ML (Random Forest) to identify anomalies

EPCs have been examined from three angles in

academic literature: their accuracy and reliability, their function as policy instrument, and lastly the value of the input data used for alternative applications [3]. To reliably use EPCs for the latter objectives, it is essential to understand the accuracy of the data as EPC records commonly contain errors.

2.1 Validating EPC Data

Studies using EPCs for policy and alternative applications often incorporate validation methods, though these vary in depth and scope. For policy-oriented applications, such as enforcing energy efficiency standards or guiding retrofit programmes, researchers commonly validate EPC data by comparing it with independent energy audits or alternative building datasets. For instance, Jenkins et al. [5] highlighted inconsistencies in EPC ratings through cross-validation with third-party assessments, emphasising the need for greater accuracy to support policies like the MEES.

Pasichnyi et al. [10] adapted the six-level validation framework by Simon [11] to EPC data validation, ranging from basic structural checks (Level 0) to cross-referencing with external datasets (Level 5):

- Level 0, Data Structure Validation: Evaluates the format and structure of the EPC dataset, checking variable types and overall organisation but not the data itself.
- Level 1, Intra-Dataset Validation: Conducts variable consistency checks between EPCs by applying physical and numerical validation procedures.
- Level 2, Longitudinal Dataset Validation: Tracks changes over time by comparing different versions of the same dataset, therefore it is referred to as inter-dataset checks. Both EPC corrections (revision checks) and updates (time-series checks) are level 2 validation.
- Level 3, Mirror Validation: Compares the EPC dataset against alternative data sources, such as data from independent energy audits, alternative energy models, or smart meter data, to identify discrepancies. Mirror checks are inter-source intra-dataset checks.
- Level 4, Within Data Collector Validation: Cross-references the EPC dataset with additional datasets from the same data collector, such as government records or organisational databases.
- Level 5, Between Data Collectors Validation: Integrates external datasets from other organisations or regions, such as geographical or taxation data.

The source of data used for validation separates these validation levels from each other. Besides level 0 that looks at data structure, 1 and 2 validation levels are based on the same source of data (e.g. EPC dataset in isolation). In level 3 to level 5, auxiliary data sources are used. Simon [11] states that the higher levels are more challenging due to dataset accessibility, but do not necessarily mean increased accuracy or reliability. Most existing EPC accuracy studies, however, concentrate on Levels 0–3, which involve consistency checks within datasets or across repeated assessments of the same property [3,5,8,9].

Most EPC auditing studies assess accuracy by comparing inputs from repeated EPCs for the same property (level 2). In other words, they are consistency checks. This is a useful approach to identify anomalies when at least 2 EPCs are available for comparison. However, it does not determine which EPC is correct or improve the data. Additionally, many properties have only one EPC record. For example, this limited the study of Yuan and Choudhary [3] to 4% of the target building type.

Hardy and Glew [9] analysed EPCs for the same properties (level 2) and reported significant errors in EPC data, estimating that 27% of records contained at least one issue, often due to variations in assessors' interpretations of building features. The errors include duplicate records differing only by dates, unlikely changes in building features such as wall or roof types over time, and anomalies in energy efficiency products (e.g. reduced insulation or glazing performance). Many errors result from differences in opinion among EPC assessors regarding building features such as age band, floors, walls, and systems or difficulties in assessing the location of flats or maisonettes within a building or the surrounding conditions [3,9]. In interviews with 20 assessors, Gledhill et al. [6] discovered that in some cases, these variations could be intentional: "*EPCs for sale and rental market are manipulated sometimes to make a property more attractive by the estate agent and the EPC isn't as accurate as it should be*".

Errors in EPC building characteristics vary across different variables. Built form classification, particularly for end-terrace or semi-detached properties, is the most inconsistent [9]. Errors in floor and wall types were observed in 10-15% of EPCs, while wall insulation and glazing performance errors occurred in around 5% [9]. Another study found that roof efficiency and total floor area exhibited the least consistency across samples. In contrast, heating system and wall efficiency data were the most consistent [5,6]. Hardy and Glew [9] found that flats and maisonettes present more errors, often due to difficulties in assessing their location within a building or the surrounding conditions, and because the built form of flats and maisonettes is not used in SAP calculations, the assessor might not pay attention to the built form. In another study, Crawley et al. [8] found that the magnitude of error decreases as the building performance increases, whereby G-rated dwellings have more errors than A-rated ones.

Quantifying the magnitude of these errors and their impact on EER is a subject of debate. Jenkins et al.'s EPC 'Mystery Shopper' study [5] quantified the effect of these errors on EER through a level 3 mirror

validation against known building info. At least four assessors evaluated the same 29 volunteer households across England and Wales, leading to an average variation of 11 points per dwelling, with one property differing by over 30 points. The results are similar to a Swedish domestic building by Hårsman et al. [12] who identified 20% variation across assessors and Crawley et al. [8] who found that the size of error could even result in wrong EPC rating. Hardy and Glew [9] concluded that an individual mistake on an EPC assessment is unlikely to make a considerable difference in the EER. In other words, identifying a single error does not necessitate removing an EPC from the analysis, although this is the most common pre-processing approach as described in section 2.2.

Higher-level validation approaches (Levels 3 to 5) leverage auxiliary datasets, such as geographic information systems (GIS), property registers, and socioeconomic data, to cross-check EPC data against other datasets. While these methods provide a more comprehensive perspective on data quality, their application is limited by challenges in accessing and integrating diverse datasets.

There is a wide range of place-based data from the Census, mostly published at, at least, one of the following geographies that were primarily designed for the publication of the Census statistics. Output Areas (OAs) are the lowest level of geographical areas. Lower Layer Super Output Areas (LSOAs) group OAs, usually four or five, containing between 400 and 1,200 households and have a resident population between 1,000 and 3,000 people. LSOAs are nested within the boundaries of Middle Layer Super Output Areas (MSOAs), which are nested within the boundaries of Local Authority Districts (LADs). These geographical areas are homogenous in terms of population size. So, one can compare like-for-like when looking at changes over time and when comparing different areas [13].

Building energy efficiency varies locally [14], with low-income households bearing higher costs of low-carbon policies in the UK due to a higher percentage of their income on energy bills [15]. EPC data linked with socioeconomic information can help identify areas at higher risk of fuel poverty [16]. Socioeconomic data at LSOA cannot be used for quantification but can be used for comparison across small areas, changes over years, and the domains/types of deprivation, particularly for financial aid and policy. For example, Yuan and Choudhary [3] demonstrated how combining EPC data with domestic gas consumption and social factors can inform a more locally tailored design of residential heating policies. No study has validated these levels to date; this study addresses that gap.

Studies focused on building energy retrofitting often use EPC recommendations and auxiliary data like GIS and property registers. Gupta and Gregg [17] designed retrofit packages for UK domestic buildings based on EPC carbon emissions and used a city-level energy mapping model to identify suitable neighbourhoods and houses for targeted retrofits, promoting bulk installations to reduce costs.

In alternative applications of EPCs, such as assessing retrofit potential, analysing fuel poverty, or evaluating the impact on real estate market value, validation often involves linking EPC data with auxiliary datasets. These include socioeconomic data, property transaction records, and GIS-based spatial analyses. For example, studies have combined EPC ratings with the Index of Multiple Deprivation (IMD) to identify areas of high energy inefficiency or fuel poverty risk [3,14]. However, these studies frequently lack systematic validation of the underlying EPC data itself, relying primarily on the end-use metrics rather than scrutinising the quality of inputs. This gap underscores the importance of integrating rigorous validation frameworks into studies leveraging EPC data for broader applications.

2.2 Improving EPC Data

Following data validation, the identified errors are either removed or corrected through substitution or imputation during the data editing stage [11]. Yet, most studies do not correct the EPC data but instead remove without quantification of the acceptable level of error in EPC data. Removing EPCs or datapoints from evaluations can introduce a bias into the results. Besides, in studies that aim to utilise EPC values for alternative applications, removing is not an option as it is necessary to determine the most accurate estimates.

One of the applications of EPCs studies their impact on property sales and rental prices [18–20]. These studies associate physical, locational, and neighbourhood characteristics to the property value. Fuerst et al. examined if there is an economic case for energy efficient dwellings in rental and purchase price. They found that 37% and 17% of EPCs in Wales from 2003 to 2014 did not record the age and number of bedrooms in turn. These EPCs were excluded from their study and whether a systematic bias was presented in non-recording EPCs is not known [18].

Not all EPC variables hold the same level of importance in analyses, and the decision to include or exclude an EPC with missing data depends on which variable is missing. Similarly, when comparing repeated EPCs, not all variations represent significant inconsistencies that warrant reporting. For instance, in existing dwellings assessed using RdSAP, certain variables are assigned default values. For example, the dwelling's type and age are used to estimate window areas, while age and wall type are used to infer U-values for walls, roofs, and floors [18]. Consequently, age (or vintage class) emerges as a critical variable.

Hardy and Glew [9] showed that ML has the potential for improving EPC auditing by identifying inconsistencies across houses within the same postcode (level 1) and EPC updates for the same house (level 2). They limited the selection to postcodes and houses with at least three EPCs, with both analyses misclassifying 7% of ‘correct’ data as errors while identifying 62% and 80% of the errors, respectively.

3 Methodology

A comprehensive EPC validation and improvement framework was developed following a systematic review of validation approaches within the literature, publicly available databases, and domestic EPC records in England and Wales. The framework was implemented in Python and follows the flowchart shown in Figure 1 and described below.

Leeds, the second most populous area and an important economic hub in the UK, was chosen as a case-study to illustrate the approach. Domestic EPC records in Leeds were sourced from the UK's Open Data Communities platform published by the Ministry of Housing, Communities and Local Government [21].

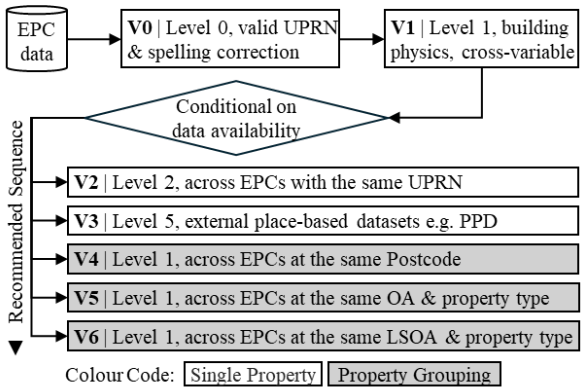


Figure 1. EPC Data Validation Engine

The six-level validation framework was applied to assess data quality of EPCs using publicly available datasets. Validation was applied based on reliability, prioritising property-specific checks over area-based analysis. This distinction is reflected in Figure 1, where property-specific validations are marked in white (V0 to V3), while validations based on property groupings are shaded in grey (V4 to V6).

Level 3 can benefit from manual user inputs or independent energy audits for mirror validation. It could include smart meter data load shape analysis and energy use disaggregation [22]. This study however is focused on automated approaches in the absence of access to building specific information. Moreover, due to unavailability of other data from the data collector, no level 4 validation is considered.

In the UK, unique property reference numbers (UPRN) are used to record anonymised public EPCs. Therefore, the level 0 validation (V0) ensures that all EPCs have a valid UPRN and no spelling inconsistencies. The dataset was then filtered to include valid EPC records within the latest 10 calendar years (2014 up to 2024), acknowledging the 10-year validity of EPCs.

Level 1, internal consistency within dataset, involves applying building physics and numerical validation procedures. V1 building physics validation is informed by leveraging insights from the literature and domain expertise. While statistical validation occurs at different tiers of geographical areas, i.e. properties within the same postcode (V4), OA (V5), and LSOA (V6). In higher geographical areas, other than properties with the same postcode, the validation is more for inconsistency checks or potential anomaly detection rather than value correction through statistical analysis and ML.

As shown in Figure 1, proceeding to the next steps beyond building physics and cross-variable checks, depends on the availability of sufficient data. For example, Level 2 (V2) compares former EPC reports of the same dwelling using UPRN, if available.

Level 5 involves cross validation with external place-based datasets, including socioeconomic data. These datasets provide contextual insights that also assist with extending the application of validated EPC data. To append auxiliary socioeconomic data, the ONS Postcode Lookup library [23] was used to retrieve OAs and LSOAs for every postcode. V3 utilises HM Land Registry Price Paid Data (PPD) [24] as an example of auxiliary property-level data. PPD reports property sales in England and Wales in every calendar year [24]. To maximise matched records for better analysis, PPD was mapped to EPCs using all possible postcode and address line combinations, ensuring inspection and price year align.

To construct datasets for anomaly detection, this study applies and enhances the approach proposed by Hardy and Glew [9], where artificial errors were introduced by modifying subsets of EPCs. UPRN and postcode were used as two identifiers to group EPCs. Only groups containing at least three EPCs were retained in V2 and V4. Within each sub-dataset, about one-third of the EPCs were randomly selected, with single building parameters replaced by the values randomly drawn from a different group. The modified EPCs were labelled as 'False', while the remaining ones were labelled as 'True'. Categorical variables were transformed based on their similarity within the same group. For example, if a postcode had three EPCs with roof types of flat, pitched, and pitched, the similarity values would be 0.33, 0.66 and 0.66, respectively. Numerical variables were transformed into percentage values relative to the mean. For example, if a UPRN has three EPCs with floor areas of 30, 40, and 50, the percentage values would be 0.75, 1, and 1.25. These features allow the Random Forest (RF) model to detect potentially erroneous records.

To further develop their approach, this study also considers two additional identifiers: (1) the combination of OAs and property types (V5) and (2) the combination of LSOAs and property types (V6). Given the larger number of properties within higher geographical areas such as OAs and LSOAs, only groups with at least ten EPCs were retained. Furthermore, they trained the model on an entire sub-dataset, which increases the risks of data leakage and model generalisation ability, causing it to perform overoptimistic conclusions due to the absence of testing datasets [25]. To mitigate this issue, the four

sub-datasets were split into an 80% training set and 20% testing set, as it is empirically the best division into the training and testing sets [26]. Due to their fast computation speed and efficiency, gradient boosting classification models including CatBoost, LightGBM and XGBoost were also applied for anomaly detection, along with the RF model [27].

4 Results

391,441 EPCs were identified in the Leeds LAD (GSS code E08000035). 9,750 of these were removed due to null UPRNs (V0). 190,221 EPCs were within the considered 10 calendar years.

V1 identifies intra-variable inconsistencies in EPCs. Improving the EPC values was achieved as per the following steps depending on the variable class: numeric, e.g. floor height, or character, e.g. hot water description. Categorical variables were assessed by unique values. For example, in Leeds, there are more than 250 unique wall, floor, and roof type descriptions. In validation level 0, spelling errors that resulted in unique values were corrected. Many of these unique values refer to the same construction composition and properties, thus can be combined into a smaller list of variations. The process of standardising these variables results in creating a rule-based dictionary that associates relevant variables to cross-check values and improve the properties where needed.

The key categorical variables are the construction age bands and property types, which were harmonised. Property type can also be compared to PPD database in level 5 validation (V3) to ensure accuracy. These key variables are essential because they are the basis for assumptions about the fabric construction and systems in the building. For example, rdSAP makes assumptions about the wall composition and thermal resistance properties based on these. One example of intra-variable improvement is entries with 'Tenure' labelled as 'Unknown', whereby conditional logic was applied based on the 'Transaction type' column to infer likely tenure types. Transactions marked as private or social rentals were reassigned accordingly.

One focal example of categorical variables is when looking across fields related to the heating source and system, where gas might be mentioned as the 'main fuel', while 'mains gas flag' might be listed as N and 'main heat description' might suggest electric storage heaters. A detailed intra-variable logic was developed to review all heating and hot water related properties. Combined with other level 1 and 2 validations, these were identified and revised where possible.

Figure 2Figure **2** summarises the accuracy evaluation of various ML models (V2, V4, V5, and V6). CatBoost achieved the highest accuracy of approximately 85% in the postcode-based dataset (V4). In the UPRNs-based dataset (V2), CatBoost also outperformed other models with an accuracy of around 81%, closely followed by LightGBM, while RF showed the lowest performance. For the sub-dataset based on the combination of LSOAs or OA and property types (V5 and V6), RF showed superior performance, achieving around 84% and 85% accuracy, respectively. Overall, CatBoost and RF were the top-performing models.

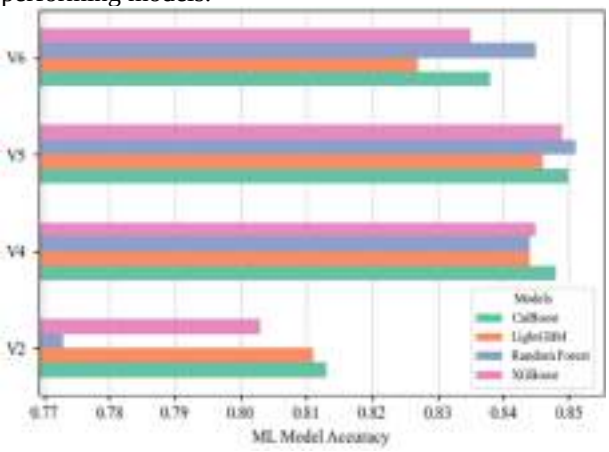


Figure 2: Accuracy comparison of ML models

Based on the accuracy results, the best-performing ML model for each validation was selected for further analysis using confusion matrices, as shown in Table 2. In V2, the model correctly classified 2622 'True' and 967 'False' EPCs, and misclassified 311 as 'False' and 517 as 'True'. While in V4, the model correctly predicted 27660 'True' and 11008 'False' records but misclassified overall 6949 EPCs.

Table 2: Confusion matrices for the best ML models

Level		False	True	Class Error
V2	False	967	517	0.34
	True	311	2622	0.10
V4	False	11008	4240	0.27
	True	2709	27660	0.08
V5	False	10740	4063	0.27
	True	2569	27103	0.08
V6	False	10532	4664	0.30
	True	2436	28150	0.07

5 Discussion

This study develops a data-driven EPC validation framework, addressing key research gaps. By integrating ML and rule-based corrections, it detects anomalies and refines data quality. Results highlight the value of auxiliary datasets, multi-level validation, and ML in anomaly detection.

5.1 Evaluation of ML Model Performance

ML models performed well across validation levels, with CatBoost and RF as top performers. In the postcode-based dataset (V4), CatBoost achieved the highest accuracy of 85%, while in the UPRNs-based dataset (V2), it led with 81% accuracy, followed by LightGBM. These results suggest location-based

grouping improves predictions, likely due to spatial correlations in EPC characteristics.

For datasets combining geographical areas and property types (V5 and V6), RF excelled (85% accuracy), indicating hierarchical property aggregation enhances anomaly detection. This aligns with prior studies [9] on validation accuracy through geographical grouping. However, False Positives remain an issue, with misclassified 'True' EPCs slightly increasing in V6 (0.30) from V4 and V5 (0.27). While True class error consistently improved (0.10 in V2 to 0.07 in V6), distinguishing erroneous records remains a challenge as the dataset scales.

5.2 Implications of EPC Validation and Policy Relevance

The effectiveness of EPCs as a policy tool depends on their accuracy. The UK government's EPC reform [28], set for 2026, aims to improve reliability by introducing a fabric efficiency rating, heating system type, and an emissions classification. The proposed reforms highlight the increasing reliance on EPCs for energy policy, financial incentives, and decarbonisation planning. However, EPC accuracy issues can undermine their effectiveness in these applications [3,5,9]. The findings reinforce the need for incorporating systematic validation frameworks into EPC auditing to ensure reliability.

This study also demonstrates that auxiliary datasets, such as HM Land Registry PPD, can enhance EPC validation, offering a new avenue beyond traditional consistency checks. While previous studies linked EPCs to property pricing [18–20], this research shows that they can also be leveraged to validate EPC inputs. Future policy reforms should consider integrating external datasets into EPC validation frameworks to improve reliability at scale.

5.3 Challenges and Future Considerations

While the proposed methodology has proven effective, several limitations warrant further research. First, geographical validation (postcode, OA, LSOA) detects anomalies but does not pinpoint specific errors in EPCs. While property-level validation (V3) improves accuracy, future studies should explore hybrid approaches that integrate spatial validation with assessor verification. Mirror validation (Level 3), using smart meter data or thermal imaging, could further enhance real-world accuracy.

This framework does not account for assessor-related inconsistencies, despite evidence that assessor experience, software versions, and subjective interpretation influence EPCs [5,6]. Future research should examine assessor-driven variations and develop standardised correction mechanisms to further refine EPC datasets.

6 Conclusions

This study demonstrates that systematic validation frameworks, integrating both rule-based corrections and ML techniques, can significantly enhance EPC reliability. By applying a multi-level validation approach and leveraging auxiliary datasets, this research identifies key limitations in conventional EPC validation methods and offers a scalable, data-driven pathway for improvement. ML models achieved up to 85% accuracy, highlighting the potential of automated anomaly detection. However, False Positives and rating imbalances remain challenges, requiring further refinement. Future research should expand access to external datasets, integrate mirror validation techniques, and develop scalable, interoperable validation systems to further improve EPC accuracy and utility.

References

[1] EPBD, Directive 2010/31/EU of the European Parliament and of the Council of 19 May 2010 on the energy performance of buildings (recast), 2010. http://data.europa.eu/eli/dir/2010/31/oj/eng (accessed December 20, 2024).

[2] BEIS, Non-domestic private rented property: minimum energy efficiency standard - landlord guidance, (2023). https://www.gov.uk/guidance/non-domestic-private-rented-property-minimum-energy-efficiency-standard-landlord-guidance (accessed December 20, 2024).

[3] M. Yuan, R. Choudhary, Energy Performance Certificate renewal — An analysis of reliability of simple non-domestic buildings' EPC ratings and pragmatic improving strategies in the UK, Energy Policy 178 (2023) 113581. https://doi.org/10.1016/j.enpol.2023.113581.

[4] M. Hughes, J. Palmer, V. Cheng, D. Shipworth, Global sensitivity analysis of England's housing energy model, Journal of Building Performance Simulation 8 (2015) 283–294. https://doi.org/10.1080/19401493.2014.925505.

[5] D. Jenkins, S. Simpson, A. Peacock, Investigating the consistency and quality of EPC ratings and assessments, Energy 138 (2017) 480–489. https://doi.org/10.1016/j.energy.2017.07.105.

[6] T. Gledhill, W. Swan, R. Fitton, A practitioner study into the variability of UK domestic energy assessments, International Journal of Building Pathology and Adaptation ahead-of-print (2023). https://doi.org/10.1108/IJBPA-10-2022-0167.

[7] Scottish Government, Energy Performance Certificate (EPC) reform: consultation, 2023. https://www.gov.scot/publications/energy-performance-certificate-epc-reform-consultation/ (accessed December 24, 2024).

[8] J. Crawley, P. Biddulph, P.J. Northrop, J. Wingfield, T. Oreszczyn, C. Elwell, Quantifying the Measurement Error on England and Wales EPC Ratings, Energies 12 (2019) 3523. https://doi.org/10.3390/en12183523.

[9] A. Hardy, D. Glew, An analysis of errors in the Energy Performance certificate database, Energy Policy 129 (2019) 1168–1178. https://doi.org/10.1016/j.enpol.2019.03.022.

[10] O. Pasichnyi, J. Wallin, F. Levihn, H. Shahrokni, O. Kordas, Energy performance certificates — New opportunities for data-enabled urban energy policy instruments?, Energy Policy 127 (2019) 486–499. https://doi.org/10.1016/j.enpol.2018.11.051.

[11] A. Simon, Definition of validation levels and other related concepts, European Commission. Eurostat., 2013. https://circabc.europa.eu/sd/a/43c3c89d-8ab4-4f39-af3f-07185035ff80/ASA_AAP_154_Annex%201_Definition_of_validation_levels_etc.doc (accessed December 24, 2024).

[12] B. Hårsman, Z. Daghbashyan, P. Chaudhary, On the quality and impact of residential energy performance certificates, Energy and Buildings 133 (2016) 711–723. https://doi.org/10.1016/j.enbuild.2016.10.033.

[13] ONS, 2011 Census geographies - Office for National Statistics, (2024). https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeographies/2011censusgeographies (accessed January 15, 2025).

[14] J. Ahlrichs, S. Wenninger, C. Wiethe, B. Häckel, Impact of socio-economic factors on local energetic retrofitting needs - A data analytics approach, Energy Policy 160 (2022) 112646. https://doi.org/10.1016/j.enpol.2021.112646.

[15] A. Owen, J. Barrett, Reducing inequality resulting from UK low-carbon policy, Climate Policy 20 (2020) 1193–1208. https://doi.org/10.1080/14693062.2020.1773754.

[16] R. Camboni, A. Corsini, R. Miniaci, P. Valbonesi, Mapping fuel poverty risk at the municipal level. A small-scale analysis of Italian Energy Performance Certificate, census and survey data, Energy Policy 155 (2021) 112324. https://doi.org/10.1016/j.enpol.2021.112324.

[17] R. Gupta, M. Gregg, Targeting and modelling urban energy retrofits using a city-scale energy mapping approach, Journal of Cleaner Production 174 (2018) 401–412. https://doi.org/10.1016/j.jclepro.2017.10.262.

[18] F. Fuerst, P. McAllister, A. Nanda, P. Wyatt, Energy performance ratings and house prices in Wales: An empirical study, Energy Policy 92 (2016) 20–33. https://doi.org/10.1016/j.enpol.2016.01.024.

[19] F. Fuerst, M.F.C. Haddad, H. Adan, Is there an economic case for energy-efficient dwellings in the UK private rental market?, Journal of Cleaner Production 245 (2020) 118642. https://doi.org/10.1016/j.jclepro.2019.118642.

[20] F. Fuerst, P. McAllister, A. Nanda, P. Wyatt, Does energy efficiency matter to home-buyers? An investigation of EPC ratings and transaction prices in England, Energy Economics 48 (2015) 145–156. https://doi.org/10.1016/j.eneco.2014.12.012.

[21] Department of Levelling Up, Housing & Communities, Energy Performance of Buildings Data England and Wales, (2024). https://epc.opendatacommunities.org/domestic/search (accessed December 21, 2024).

[22] M. Taheri, L. Jacob, C. Parry, S. Mirzaie, A. Hermanowicz, I. Vrachimi: Alan, The efforts towards development of an energy efficiency upgrade platform, in: IBPSA, 2021: pp. 3300–3307. https://doi.org/10.26868/25222708.2021.30910.

[23] Postcode to OA (2021) to LSOA to MSOA to LAD (November 2024) Best Fit Lookup in the UK, (n.d.). https://open-geography-portalx-ons.hub.arcgis.com/datasets/ons::postcode-to-oa-2021-to-lsoa-to-msoa-to-lad-november-2024-best-fit-lookup-in-the-uk/about (accessed March 14, 2025).

[24] HM Land Registry, Price Paid Data, GOV.UK (2024). https://www.gov.uk/government/statistical-data-sets/price-paid-data-downloads (accessed December 24, 2024).

[25] M.A. Bouke, A. Abdullah, An empirical study of pattern leakage impact during data preprocessing on machine learning-based intrusion detection models reliability, Expert Systems with Applications 230 (2023) 120715. https://doi.org/10.1016/j.eswa.2023.120715.

[26] A. Gholamy, V. Kreinovich, O. Kosheleva, Why 70/30 or 80/20 Relation Between Training and Testing Sets: A Pedagogical Explanation, International Journal of Intelligent Technologies & Applied Statistics 11 (2018) 105–111. https://doi.org/10.6148/IJITAS.201806_11(2).0003.

[27] C. Bentéjac, A. Csörgő, G. Martínez-Muñoz, A comparative analysis of gradient boosting algorithms, Artif Intell Rev 54 (2021) 1937–1967. https://doi.org/10.1007/s10462-020-09896-5.

[28] DESNZ and MHCLG, Open consultation - Reforms to the Energy Performance of Buildings regime, GOV.UK (2024). https://www.gov.uk/government/consultations/reforms-to-the-energy-performance-of-buildings-regime (accessed December 24, 2024).

Transforming Construction Waste into Resources for 3D Printed Concrete

Ekaterina Kravchenko[1]*, Muhammad Huzaifa Raza[2], Svetlana Besklubova[2], Georgy Lazorenko[3]

[1] *Southern Federal University, Rostov-on-Don, Russia*
[2] *Department of Data and Systems Engineering, The University of Hong Kong, Hong Kong*
[3] *Novosibirsk State University, Pirogov Street, 2, Novosibirsk, Russia*
Emails: ekakra@sfedu.ru; huzaifa@connect.hku.hk; lanabess@hku.hk; glazorenko@yandex.ru
*Corresponding author

Abstract

The construction sector faces significant challenges in balancing environmental sustainability, economic feasibility, and compliance with required strength standards. This study investigates the potential of recycled materials, specifically waste concrete fine aggregates (WCA), for use in geopolymer mortar for 3D concrete printing as a sustainable alternative to conventional construction methods. The research examines the fresh and hardened properties of 3D-printed geopolymer mortar. Experimental results demonstrate that WCA-based mixtures achieve superior compressive strength and dimensional stability compared to natural fine aggregates (NFA). Life cycle analysis (LCA) further highlights the environmental benefits of WCA-based 3D mixture, showing a 3.4% reduction in global warming potential (GWP) and a decrease in terrestrial acidification compared to NFA. The study also underscores the importance of orientation in 3D printing, with significant improvements in compressive strength observed in all tested directions: perpendicular (64.2%), lateral (77.2%), and longitudinal (57.7%). This research emphasizes the promise of 3D concrete printing with recycled materials as a transformative approach to sustainable construction, reducing environmental impacts while maintaining structural performance.

Keywords – Waste concrete; Geopolymers; Life cycle analysis; Environmental impact.

1 Introduction

The construction sector is undergoing a transformation driven by the integration of technologies, with 3D concrete printing emerging as a revolutionary approach to construction processes [1]. This method allows the direct conversion of digital building models into physical structures, eliminating the need for labor-intensive and resource-demanding formwork [2]. The advantages of 3D concrete printing include faster project completion, reduced material waste, and the ability to create intricate geometries with precision and minimal labor costs [2].

In addition to improving construction efficiency, 3D concrete printing presents an opportunity to incorporate recycled materials, such as fly ash (FA) [3], ground granulated blast furnace slag (GGBS), and waste concrete-based recycled fine aggregates [4,5]. These materials not only mitigate environmental issues associated with industrial and construction waste disposal but also reduce dependency on traditional resources like Ordinary Portland Cement (OPC). The use of OPC is a significant contributor to carbon emissions and resource depletion, necessitating sustainable alternatives [6]. According to an estimate, the manufacture of each ton of OPC requires the mining of 1.5 tons of limestone and releases around 0.5 tons of CO_2 into the atmosphere [6]. OPC production accounts for 5-7% of overall greenhouse gas emissions, making it the fourth most major source after petroleum, coal, and natural gas. [7] Moreover, the use of OPC in 3D concrete printing technique has a variety of constraints due to its setting characteristics. These environmental and technical challenges highlight the need for alternative low-CO_2 binders. In this context, geopolymers have emerged as a promising solution, as they can be formulated using industrial by-products [3], thereby reducing waste disposal in open environments and mitigating carbon emissions.

In recent years, the recycling of construction and demolition waste, such as recycled concrete, crushed brick, and ceramic aggregates, has gained significant research interest. Utilizing waste concrete as a substitute for natural resources has led to the development of recycled aggregate concrete, a sustainable product addressing both resource scarcity and waste disposal

challenges [8]. Ding et al. [9] investigated recycled fine aggregates in 3D-printed concrete, reporting slightly lower compressive and flexural strengths compared to natural sand, with significant anisotropy in mechanical properties. Panda and Tan [3] conducted a study on fly ash-based geopolymer mortar for 3D concrete printing and found that the mixture with a sand-to-binder ratio of 1.5 exhibited the best extrudability with a yield stress of 0.6 kPa. The optimal open time for printing was 15–20 minutes, preventing deformation of the lower layers. The mixture was successfully used to print a 60 cm model in 20 minutes with good interlayer adhesion. In 3D printing processes, the integration of FA, GGBS and WCA presents unique challenges and opportunities. Unlike standard concrete, 3D printing requires precise control over rheological properties, buildability, and layer adhesion, which necessitates a more tailored approach to mix design. The ongoing research in 3D printing is focused on optimizing these supplementary materials to ensure printability without compromising mechanical performance or sustainability benefits. While FA, GGBS, and recycled aggregate are already used in traditional concrete, their adaptation to 3D printing remains an evolving area of study aimed at maximizing their potential in automated construction processes.

This study investigates the application of geopolymer binders and WCA in 3D concrete printing, focusing on their fresh and hardened properties, environmental impacts, and compliance with strength standards. By evaluating the performance and sustainability of these materials, this research aims to demonstrate the viability of integrating recycled resources into 3D concrete printing as a pathway toward more sustainable and resilient construction practices.

2 Materials and Research Methodology

2.1 Materials and Methods

Class F fly ash, adhering to the ASTM C618 standard, and GGBS, conforming to the C989 standard, were utilized as precursors. Industrial-grade sodium hydroxide pellets were used as the alkaline activator for the binder. Silica sand with an average particle size of 0.5 μm served as the natural fine aggregate, while recycled fine aggregate derived from waste concrete (WC), with a similar particle size, was employed as a replacement for the natural fine aggregate.

The mix design for the two samples, 3D-NFA (natural fine aggregate) and 3D-WCA (waste concrete aggregate) is as follows: For the 3D-NFA sample, the binder consists of 60 g of FA, 25 g of GGBS, and 15 g of sodium hydroxide. The aggregate used is 100 g of natural fine aggregate, with no recycled fine aggregate included. The water-to-binder ratio is 0.20, and no admixture is added. For the 3D-WCA sample, the binder composition remains the same, with 60 g of FA, 25 g of GGBS, and 15 g of sodium hydroxide. However, the aggregate used consists of 100 g of recycled fine aggregate, with no natural fine aggregate included. The water-to-binder ratio is increased to 0.25, and 0.1 g of sucrose-based admixture is added. The mix composition for the two specimens was determined through preliminary laboratory experiments, optimizing the water content and sodium hydroxide dosage based on the findings of these tests.

The printing experiments involved producing a slab measuring 200 mm × 100 mm × 100 mm and a hollow cylinder with a diameter of 100 mm and a height of 200 mm, using both natural and recycled fine aggregates. The layer height and width were set to 10 mm and 30 mm, respectively. The material extrusion rate during the printing process was maintained at 0.55 m^3/h, and the printhead speed was consistently set to 30 mm/s. The printed sample using geopolymer binders with waste concrete and fly ash is presented in Figure 1.

To evaluate the anisotropic behavior of 3D-printed mortars, specimens were tested under perpendicular, longitudinal, and lateral loads relative to the printing direction. The printing orientation was defined as the longitudinal direction. Compression tests were conducted at a strain rate of 1 mm/min, with each test repeated three times to obtain average compressive strength values.

2.2 Life cycle analysis

To evaluate the environmental impact of materials and construction methods, a comprehensive life cycle analysis (LCA) was conducted in accordance with ISO 14040 and 14044 standards. The analysis was performed using OpenLCA 2.0.3 software, with the Ecoinvent 3.9.1 database serving as a reliable and compliant data source. The cradle-to-gate approach was adopted for the LCA, covering the production processes of key raw materials. The analysis applied the ReCiPe 2016 midpoint (H) method, focusing on critical environmental impact categories such as global warming potential (GWP), stratospheric ozone depletion, terrestrial acidification, freshwater eutrophication, fossil resource scarcity, water consumption, and fine particulate matter formation, as recommended by the Product Environmental Footprint Category Rules (PEFCR).

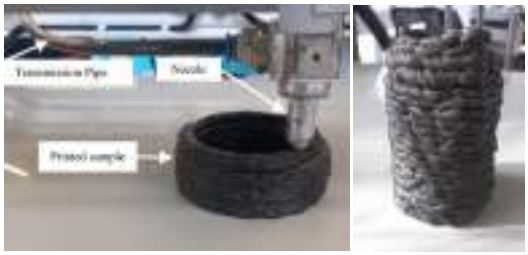
Case study data for LCA are summarized in Table 2.

To assess the environmental impact of recycled concrete aggregates in construction, the performance of concrete containing recycled aggregates was compared to that of concrete using NFA.

Table 2. Case study data for LCA

Parameters	Value and unit
Printing speed	250 mm/s
Printing time	1.381 h
Printer power	19.06
Mixing time	1.547 h
Mixer power	9 kW
Pumping/flowing time	8.738 h
Pump power	22 kW

Figure 1. 3D printed sample using geopolymer binders with waste concrete and fly ash.



3 Results and discussion

The fresh property results showed that the slump flow and spread diameter for mixtures with NFA were 14.3 mm and 138.2 mm, respectively. In contrast, replacing natural aggregates with WC aggregates increased these values to 14.5 mm and 143.2 mm, respectively, indicating reasonable flowability for all mixtures.

The maximum number of layers printed was 28 for the mixture with NFA and 24 for the mixture with WC aggregates. The Vicat value for the mixture with natural aggregates ranged between a lower limit of 5±1.5 mm and an upper limit of 21±6 mm, whereas for the mixture with artificial aggregates, the upper limit was slightly higher at 24±6 mm. These results are consistent with those previously presented in the work by Papachristoforou et al. [11].

Figure 2 presents the compressive strength results for two different mixtures (3D-WCA and 3D-NFA) tested along three principal directions: perpendicular, lateral, and longitudinal. The results reveal a significant influence of the material's extrusion process and orientation on its compressive strength. The specimens 3D-WCA and 3D-NFA exhibited higher strength values of 5.9 MPa and 3.7 MPa, respectively, along the longitudinal direction. This increase in strength can be attributed to the extrusion process, where the material is compacted and densified as it is pushed through the nozzle to print the layers. This densification leads to enhanced bonding and material integrity along the longitudinal direction, resulting in higher compressive strength. When tested in the perpendicular direction, competitive compressive strength values of 5.9 MPa and 3.5 MPa were observed for 3D-WCA and 3D-NFA, respectively. These results suggest that the material's structure in this orientation maintains a significant level of bonding and compaction. The introduction of waste concrete aggregate place as a replacement for NFA further contributed to an overall improvement in strength in this direction. In contrast, the lateral direction exhibited the lowest compressive strength among the three tested orientations. This can be explained by the lower pressure exerted during the extrusion process in this direction.

In general, when using natural aggregates, particularly sand, the printed material tends to expand and settle more freely, resulting in weaker bonding and a less compacted structure, which subsequently leads to reduced strength. A comparison of the results shows a clear trend: replacing natural aggregates with WC has significantly improved the compressive strength in all tested directions. Specifically, the compressive strength in the perpendicular direction increased by 64.2%, while the lateral and longitudinal directions showed enhancements of 77.2% and 57.7%, respectively. These improvements highlight the potential of WCA to enhance the mechanical performance of 3D-printed concrete, particularly by improving bonding and compaction in critical load-bearing directions.

The pumpability results showed a significant difference between the two mixtures. The average pumpability for 3D-NFA was 1580 mm^3/sec, whereas for 3D-WCA, it was lower at 1400 mm^3/sec. This indicates that incorporating waste concrete aggregates may reduce the ease of material flow through the pump. The lower pumpability of 3D-WCA can be attributed to the higher water absorption and irregular shape of WCA particles, which increase friction and resistance in the pumping system. Moreover, the rough texture and porosity of WCA may lead to greater water demand, affecting the overall workability of the mixture. While WCA enhances compressive strength, these findings suggest that adjustments in mix design, such as optimizing aggregate gradation, increasing superplasticizer dosage, or modifying water-to-binder ratios, may be necessary to improve pumpability and ensure smooth extrusion in 3D printing applications.

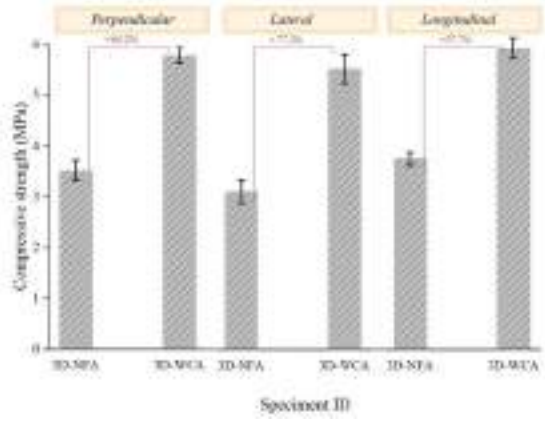


Figure 2. 7-day compressive strength of 3D printed specimens using NFA and WCA.

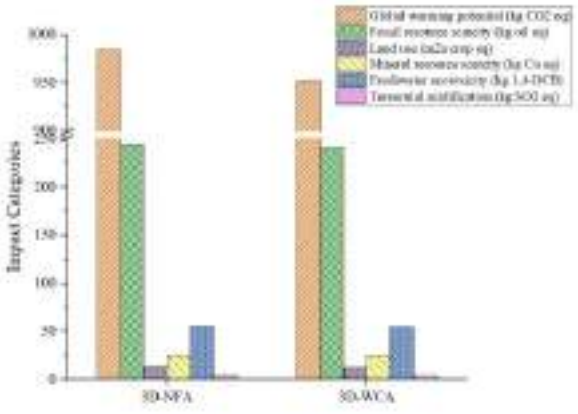


Figure 3. Comparative analysis of life cycle impact categories for 3DP-NFA and 3DP-WCA.

In LCA, GWP quantifies the amount of infrared radiation absorbed by greenhouse gases over a specified time, referenced to 1 kg of CO_2 equivalent. It measures heat trapped by greenhouse gases, which contributes to rising temperatures, extreme weather, sea-level rise, and shifting climate patterns. Analysis reveals that incorporating recycled WCA reduces GWP by 3.4% in 3D printing (Fig. 3), consistent with findings by Habibi et al. [10]. Sodium hydroxide significantly contribute to GWP due to their energy-intensive production processes.

Terrestrial acidification, another critical impact category, reflects emissions that acidify soils through the deposition of compounds like NOx, NH_3, and SO_2, evaluated relative to SO_2. These emissions lower soil pH, release heavy metals, and reduce buffering capacity, negatively impacting plants and animals. The analysis shows that using NFA results in the highest acidification potential (3.63 kg SO_2 eq), which decreases to 3.45 kg SO_2 eq when replaced with recycled aggregates. The NFA extraction process, involving blasting and quarrying, directly and indirectly, releases acidifying compounds through fossil fuel combustion, contributing more to acidification than recycled aggregates.

Despite the promising results obtained in this study, certain limitations should be acknowledged. One of the primary challenges in evaluating the fresh and hardened properties of 3D-printed concrete is the influence of the extrusion process on the material's structural integrity. Variability in extrusion pressure and nozzle geometry may introduce inconsistencies in layer bonding, potentially affecting compressive strength, particularly in the lateral direction. Additionally, while the substitution of NFA with WCA demonstrated improvements in compressive strength, further investigation is required to assess its long-term durability.

4 Conclusion

This study demonstrates the potential of recycled waste concrete for sustainable 3D concrete printing. Experimental results confirm that replacing NFA with waste concrete fine aggregate improves the hardened properties of the mixtures. The compressive strength showed significant improvements in the perpendicular (64.2%), lateral (77.2%), and longitudinal (57.7%) directions when the recycled aggregates were applied.

The LCA highlighted the environmental benefits of recycled fine aggregates, including a 3.4% reduction in global warming potential compared to natural aggregates, as well as lower terrestrial acidification potential. The reduced environmental impacts are attributed to the minimized reliance on resource-intensive natural aggregate extraction and the effective reuse of waste materials. Additionally, 3D concrete printing with waste concrete fine aggregates eliminates the need for formwork, further reducing material consumption and environmental burdens. These findings underscore the feasibility of integrating recycled materials into 3D concrete printing to achieve sustainable construction practices, offering a pathway to reduce carbon emissions, resource depletion, and environmental degradation. Future research should explore optimizing material compositions and scaling up 3D-printed structures to further validate their environmental and structural performance.

The authors acknowledged the support by the Ministry of Science and Higher Education of the Russian Federation (grant No. FSUS-2024-0027).

References

[1] B. Sun, Q. Zeng, D. Wang, & W. Zhao. Sustainable 3D printed mortar with CO2 pretreated recycled fine aggregates. *Cement and Concrete Composites*, 134, 104800, 2022.

[2] S. Besklubova, N. Spicek, R.Y. Zhong, E. Kravchenko, & J. Miroslaw. A framework for the acceptance of 3D printing technology in construction. *International Journal of Applied Science and Engineering*, 21(4), 1-25, 2024.

[3] B. Panda, M. Tan. Experimental study on mix proportion and fresh properties of fly ash based geopolymer for 3D concrete printing. *Ceramics International*, 44(9), 10258-10265. 2018.

[4] Y. Han, Z. Yang, T. Ding, & J. Xiao. Environmental and economic assessment on 3D printed buildings with recycled concrete. *Journal of Cleaner Production*, 278, 123884. 2021.

[5] G. Lazorenko, E. Kravchenko, A. Kasprzhitskii, & E.H. Fini. An evaluation of the environmental impact and energy efficiency of producing geopolymer mortar with plastic aggregates. *Resources, Conservation & Recycling Advances*, 22, 200216. 2024.

[6] Huntzinger, D. N., & Eatmon, T. A life-cycle assessment of Portland cement manufacturing: comparing the traditional process with alternative technologies. *Journal of cleaner production*, 17(7), 668-675. 2009.

[7] Heath, A., Paine, K., & McManus, M. Minimising the global warming potential of clay based geopolymers. *Journal of Cleaner Production*, 78, 75-83. 2014.

[8] Kravchenko, E., Lazorenko, G., Jiang, X., Leng, Z. Alkali-activated materials made of construction demolition waste as precursors: A review. *Sustainable Materials and Technologies*, e00829. 2024.

[9] Ding, T., Xiao, J., Zou, S., Wang, Y. Hardened properties of layered 3D printed concrete with recycled sand. *Cement and Concrete Composites*, 113, 103724. 2020.

[10] Habibi, A., A.M. Ramezanianpour, and M. Mahdikhani, RSM-based optimized mix design of recycled aggregate concrete containing supplementary cementitious materials based on waste generation and global warming potential. *Resources, Conservation and Recycling*, 2021. 167: p. 105420.

[11] Papachristoforou, M., Mitsopoulos, V., Stefanidou, M. Evaluation of workability parameters in 3D printing concrete. Procedia Structural Integrity, 10, 155-162. 2018.

Economic Viability of 3D Concrete Printing: A Comparative Study with Traditional Construction Method

Muhammad Huzaifa Raza[1]*, Svetlana Besklubova[1], Ekaterina Kravchenko[2], Ray Y. Zhong[1]

[1]Department of Data and Systems Engineering, The University of Hong Kong, Pokfulam, Hong Kong

[2]Academy of Biology and Biotechnology, Southern Federal University, Rostov-On-Don, Russia

huzaifa@connect.hku.hk, lanabess@hku.hk, ekakra@sfedu.ru, zhongzry@hku.hk

Abstract –

The construction sector's considerable environmental impact has led to a growing demand for sustainable construction technologies. This study aims to compare the cost and environmental impact of traditional construction and 3D printing construction processes, starting from the supply of raw materials to the final construction and construction waste disposal. Based on the factors identified (raw material, printing equipment, transportation, warehousing, construction and waste disposal), traditional construction and 3D printing value streams are mapped to assess the potential factors or activities contributing to the cost and environmental burden. The results showed that 3D printing offers nearly 80% lower cost at the construction phase than traditional construction. However, 3D printing hardly reduces the total cost by only 0.96% and even produces significantly higher greenhouse gas (GHG) emissions than traditional construction due to the significant transportation cost of raw materials and 3D printing equipment. The higher transportation costs in 3D printing are mainly due to the lack of a well-established market for 3D printing equipment and raw material supply. Futuristic scenario analysis revealed that 3D printing could reduce the total construction cost (starting from raw material and equipment supply to construction waste disposal) by 61% when the infrastructure for 3D printing is established in the same way as traditional construction. The findings of this study explored the sustainability potential of 3D printing and offered insights for its application in construction projects.

Keywords – 3D printing; traditional construction; cost analysis; value stream mapping; GHG emissions

1 Introduction

As the global population continues to grow, the demand for new construction has surged, leading to significant expansion in the construction industry. This sector consumes 40% of natural resources, including minerals and fossil fuels, and is responsible for nearly 40% of greenhouse gas emissions among industrial sectors. With projections estimating a global population increase of 9.8 billion by 2050 and up to 11.2 billion by 2100 [1], the strain on mineral and fossil resources is expected to escalate dramatically. In light of this environmental impact, the adoption of smart construction technologies is essential to mitigate damage. One promising solution is 3D printing, also known as additive manufacturing, which creates components by layering materials. This technology has emerged as a sustainable alternative in construction, enabling the production of buildings and components through successive layers of concrete with minimal human intervention. 3D printing (3DP) offers a significant advantage by reducing the cost and environmental impact associated with traditional plywood and metal formwork [2], as it can produce self-supporting concrete layers. Unlike conventional methods, 3D printing minimizes economic and environmental burdens while allowing for the creation of complex geometries with less reliance on intricate formwork [3].

In the construction industry, 3D printing has offered significant benefits by reducing labor costs, material waste, and construction time. In addition, 3D printing streamlines the supply chain by enabling just-in-time delivery of equipment and raw materials directly to the construction site, eliminating the need for large inventories of raw materials and formwork. In 2020, the global 3D printing market experienced impressive growth, surging by 21% compared to 2019 and reaching an estimated value of $12.6 billion [4]. Owing to these benefits, several studies have conducted an extensive economic analysis for 3D printing with reference to

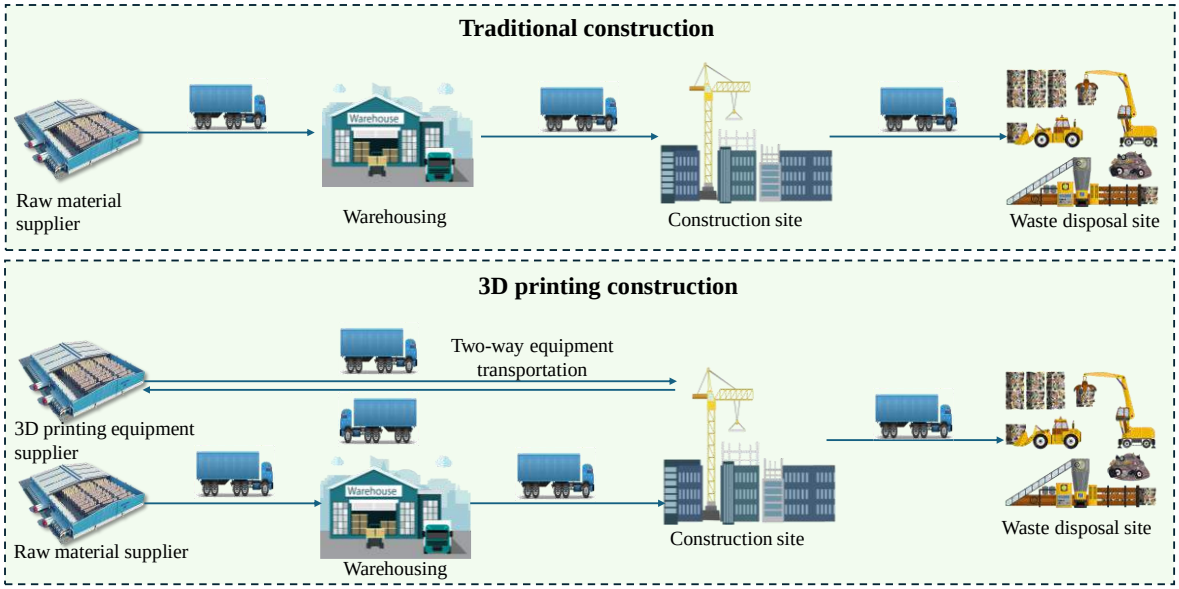


Figure 1. Supply chain configuration for traditional and onsite 3D printing construction.

traditional construction. In their study, Markin et al. [5] analyzed the cost components of 3D printing and found that labor costs contribute just 21%, while material costs account for 70% of the total construction expenses for 3D printing with foam concrete.

Weng et al. [6] conducted a comparative analysis of 3D printing and precast construction. Their findings revealed that 3D printing offered 34% less cost than precast construction. Batikha et al. [7] compared the cost of a two-story building using different construction methods. Their results showed that 3D printing yields 21% less construction cost than precast modular construction. Tobi et al. [8] conducted a cost analysis of a 3D printed house in the UK and revealed that the 3D printing method offers nearly 30% less construction cost than traditional methods. The findings of Allouzi et al. [9] revealed that the 3D printing method has the potential to reduce the material consumption cost by 65%. Abdalla et al. [10] conducted a comparison of formwork and raw material costs between construction 3D printing and traditional construction methods, revealing that 3D printing incurs a capital cost that is 78% lower, primarily due to the elimination of formwork and reduction in concrete consumption. Han et al. [11] considered a case study of a cylindrical silo for the comparative analysis of 3D printing and traditional construction methods. Their findings revealed that the construction cost (labor, electricity, formwork and fuel for operating equipment) of 3D printing is nearly 80% lower than conventional construction.

Apart from cost effectiveness, 3D printing also aids in reducing the environmental impact generated by the consumption of materials and energy due to their efficient utilization. Alhumayani et al. [12] examined the environmental effects of 3D printing compared to traditional casting methods in large-scale construction. Their findings indicated that the overall environmental impact of 3D concrete printing (3DCP) is 50% lower than that of conventional construction. In a study, Weng et al. [6] revealed that traditional construction produces nearly 86% higher carbon emissions than 3D printing construction. Batikha et al. [7] performed a comparative analysis of five construction methods—3D concrete printing (3DCP), prefabrication, cast in-situ, reinforced concrete, and both hot-rolled and cold-formed steel—for building a two-story structure. Their results showed that 3DCP generates 32% lower carbon emissions, whereas prefabricated modular construction has the highest carbon emissions among the methods studied. Han et al. [11] conducted a comprehensive sustainability analysis and revealed that 3D printing offers 2.9% less greenhouse gas emissions than conventional construction during the construction phase. Based on the extensive literature review, it can be believed that 3D printing can be a better alternative to conventional construction due to its cost-effectiveness and reduced environmental impact during the construction phase. However, Besklubova et al. [13] revealed that 3D printing construction is only cost-effective at the construction phase. In their study, it was observed that the logistics cost of 3D printing is significantly higher than that of traditional construction, which ultimately increases the overall construction cost of 3D printing. Based on the extensive literature review, it has been observed that most of the studies have compared the economic and environmental impact of 3D printing with traditional methods at the construction phase considering major factors such as labor hours, energy/fuel consumption, material consumption, and

equipment usage. Therefore, there is a need for studies on the comprehensive economic and environmental impact analysis considering the entire supply chain starting from the supply of raw materials to the waste disposal phase.

Even though 3D printing has sustainability potential due to a higher level of automation, it has been reported that the higher logistics cost can hinder its real-time application [13]. Therefore, there is a need to investigate the potential reasons for its higher logistics cost and to enhance its economic benefits. This study aims to conduct a comprehensive economic analysis to determine the potential factors contributing to the higher cost and throughout the value stream of 3D printing construction and highlight its possible economic benefits by considering a hypothetical scenario representing the future state of the 3D printing market in the construction industry. Moreover, this study also focused on the analysis of the environmental impact of both 3D printing and traditional construction and features potential environmental perks in the future state.

2 Methodology

This section covers the selection of major factors or activities for traditional construction and 3D printing that incur costs and cause environmental burdens. In the next step, the methodology for mapping these activities on the value stream is presented. After mapping the value streams, real-time case studies have been considered to assess the activities that experience significant costs and contribute to the environmental burden.

2.1 Parameters identification

First of all, the activity-based costing (ABC) approach has been used to highlight the activities that incur the resources and costs at the different phases of the construction process. The ABC method offers more precise cost information, facilitating effective monitoring of supply chain and construction strategies. In both construction methods, activities are highlighted, from the supply of the raw materials to the construction waste disposal to estimate the cost. All activities are further disintegrated into several other small factors. For transportation, the amount of material to be transported, distances, labor cost, handling cost, fuel cost and environmental cost are considered. For the construction phase, factors including labor cost, construction time, equipment depreciation, energy cost, material cost, and environmental cost are encountered. Apart from the cost, the factors contributing to the greenhouse gas emissions, including fuel (for transportation), energy and construction materials in different construction phases, are also highlighted.

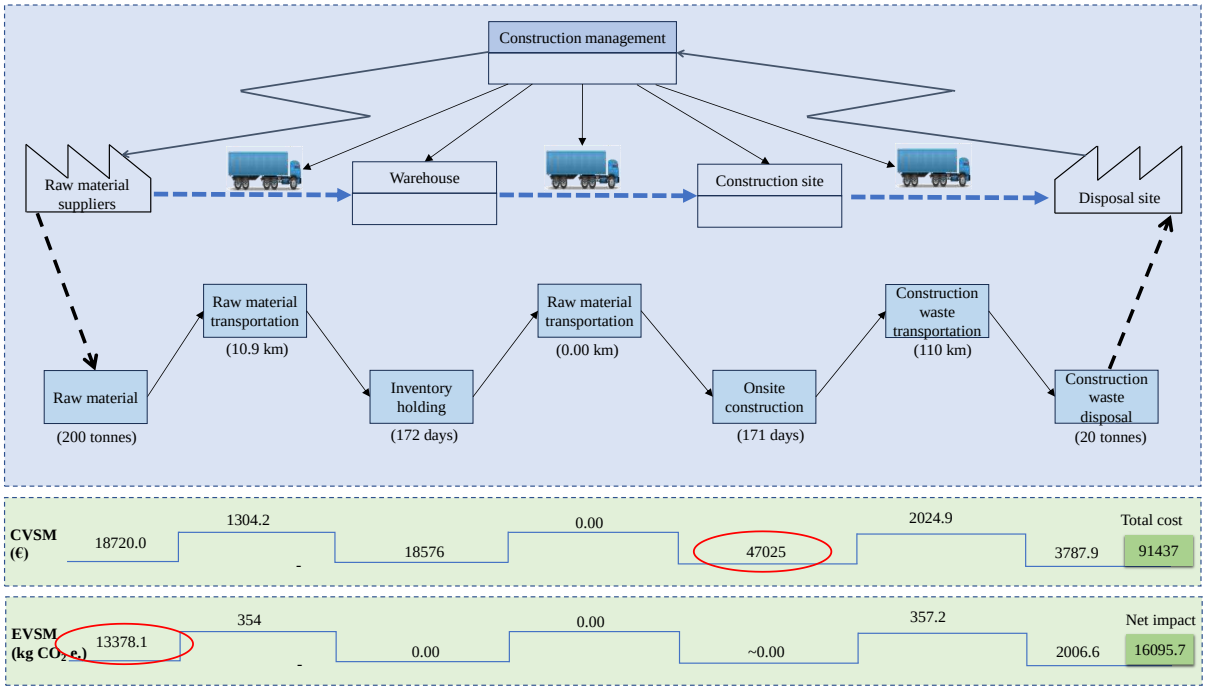
2.2 Mapping the value streams

Value stream mapping was used to highlight the cost- and emissions-intensive phases of the construction processes. Value stream mapping is a valuable tool for visualizing and quantifying activities across the construction supply chain. Figure 1 illustrates the supply chains for both traditional and 3D printing construction, which aid in the development of value streams. Within the value stream, activities related to 3D printing and traditional construction processes are clearly delineated into facility-based and non-facility-based activities. Facility-based activities include warehousing and onsite construction, while material and equipment handling and their transportation between these facilities are classified as non-facility-based [14]. For traditional construction, the value stream starts with the supply of raw materials from the suppliers to the warehouse and then to the construction site, which finally ends with the transportation of construction waste material from the construction site to the disposal site. The value stream of onsite 3D printing is nearly the same as traditional onsite construction; however, it also involves the two-way transportation of 3D printing equipment from the supplier to the construction site and back to the supplier (upon project completion). After highlighting the processes, the cost and carbon emissions are highlighted for each stage of the value stream for both traditional construction and 3D printing processes. Upon mapping the cost and environmental value streams, critical stages are highlighted in order to reduce their economic and environmental impacts.

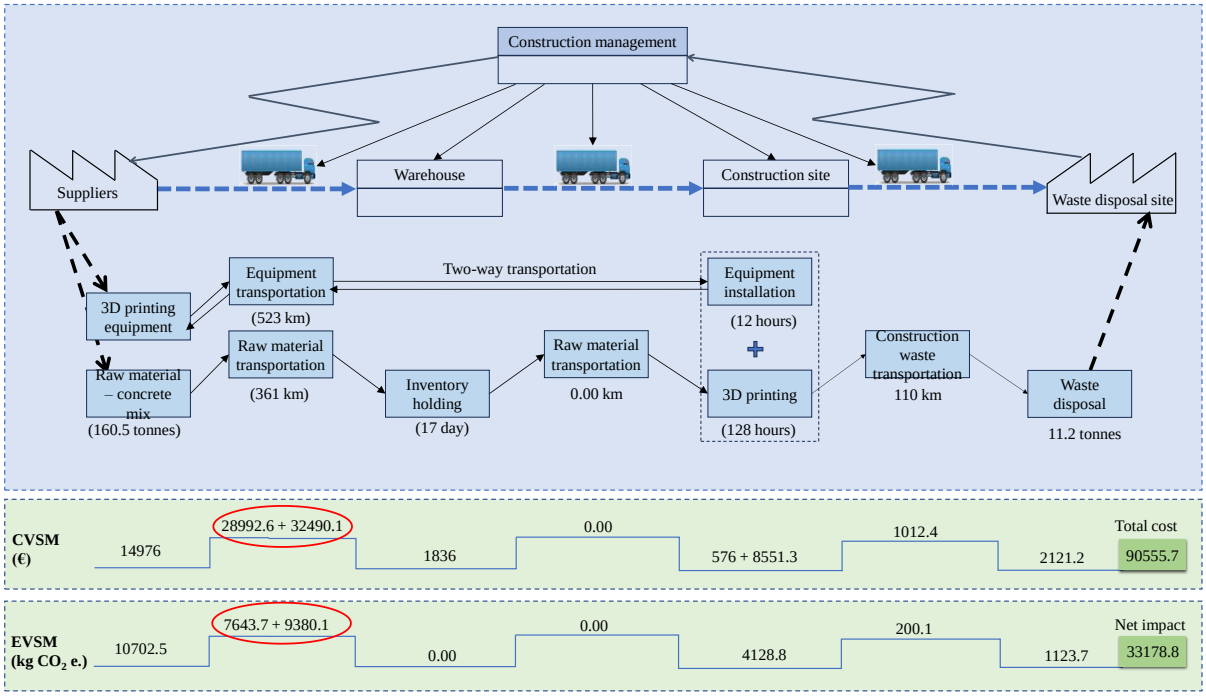
2.3 Case study and data collection

This section outlines the application of the proposed model to evaluate the feasibility of 3D printed projects in comparison to traditional construction methods. For this analysis, case studies of residential buildings in Berlin and Beckum, Germany, were selected. The onsite 3D printed project, a joint venture involving PERI, COBOD, and others, was chosen for its unconventional design and onsite construction, which aligns with industry interests. This project has received a building permit from the local authorities in Beckum, providing valuable insights. The study gathered data from open-access articles, reports, and official websites to design a technical questionnaire for operational employees, focusing on project specifics, resource logistics, and costs. Follow-up interviews with senior management addressed gaps and enriched insights. Collected data was then applied to a model to validate the feasibility and practical implementation in the case studies, demonstrating the utility of the results.

In contrast, a traditionally built house in Berlin, constructed with sand-lime blocks and having a similar floor area, was selected for comparison. Bricks or sand-



(a)



(b)

Figure 2. CVSM and EVSM for (a) traditional construction (b) 3D printing construction.

lime blocks are chosen over cast-in-place concrete in small-scale construction because they have lower structural needs (sufficient strength for low-rise buildings), are less expensive (no formwork, easier labor and logistics), and can be built more quickly (no curing delays). They are easier to manage in tight places and provide greater thermal/acoustic performance, reduced dead loads, and shrinkage-free stability. These projects meet the criteria for the case study as they are real-world examples from the same country, use the same currency, and involve a similar scope of work, offering comprehensive data on various aspects. Although the

projects were implemented in cities of different sizes (Berlin and Beckum), the cost calculations were based on average values for Germany, considering factors such as materials, equipment, and waste transportation, which often require intercity logistics. Based on the parameters identified and mathematical models developed by Besklubova et al. [13], the cost-based value streams have been determined for the comparative analysis of traditional construction and 3D printing. To assess the environmental impact of both construction techniques, a detailed methodology of Life cycle analysis presented by Raza et al. has been adopted [15].

3 Results and discussion

This section provides a comparative analysis of the traditional and 3D printing construction techniques in terms of cost and environmental impact. To illustrate the cost incurred and environmental impact produced at different stages of construction, cost-based value stream mapping (CVSM) and environmental value stream mapping (EVSM) for both traditional and 3D printing construction are considered.

3.1 Cost-based value stream mapping

Cost-based value streams of traditional and 3D printing construction are mapped to investigate the costs at the different phases. In the case of value streams, costs associated with raw materials, transportation, warehousing, labor, and construction waste disposal are considered for a low-rise building. Value streams start with the supply of raw materials, and raw material cost is incurred in the first stage. From Figure 2, it can be seen that the raw material cost of 3D printing (14976 €) is lower than that of the traditional construction process. This is primarily due to the fact that traditional construction consumes more material due to cutting, shaping, and fitting materials in intricate shapes, leading to the consumption of more material. On the other hand, 3D printing does not require any excessive material in the case of complex designs. The next step involves transporting raw materials from the supplier to the warehouse or construction site. It has been observed that traditional construction suppliers are available quite close (10.9 km) to the construction site. However, 3D printing requires specialized composition or formulation of concrete material, which is not commonly available in the market; therefore, the raw material is procured from a specific supplier from a distant place. Increased distance (361 km) between the supplier and warehouse or construction site significantly increases the transportation cost to 32490.1 €. 3D printing also involves the transportation of large sized 3D printing equipment. Similarly, in the case of 3D printing material supply, printing equipment is also delivered by particular suppliers which are not commonly available. Transportation of printing equipment involves the delivery of a large-sized gantry system, piston pump, batching plant and a silo, and equipment is returned to the supplier upon the completion of the project.

Two-way transportation from a distant place (523 km) incurs a significant transportation cost of 28992.6 €. Warehousing cost is also a distinguishing factor between traditional and 3D printing construction. The warehousing cost of traditional construction is significantly higher (18576 €) as construction materials are held for a longer time as the construction time of traditional construction is almost 171 days; therefore, material holding is considered for almost the same duration with an additional one day. On the other hand, 3D printing time is significantly shorter; therefore, raw materials are held for a short period of time (17 days). In this case study, the warehouse is considered close to the construction site; therefore, the transportation of materials from the warehouse to the construction site is negligible. On the construction phase, typical construction cost involves labor, energy and equipment cost. In traditional construction, low-rise buildings typically involve manual operations for small-scale projects, leading to substantial labor costs. As a result, the construction phase incurs significantly higher expenses, amounting to approximately €47025. Equipment and energy costs are negligible compared to labor costs. However, 3D printing involves the printing of material, which requires few operators for its installation and printing process and consumes energy. Due to the significant level of automation, the cost of the



Figure 3. Locations of construction companies in Germany are highlighted with yellow icons on the map [16].

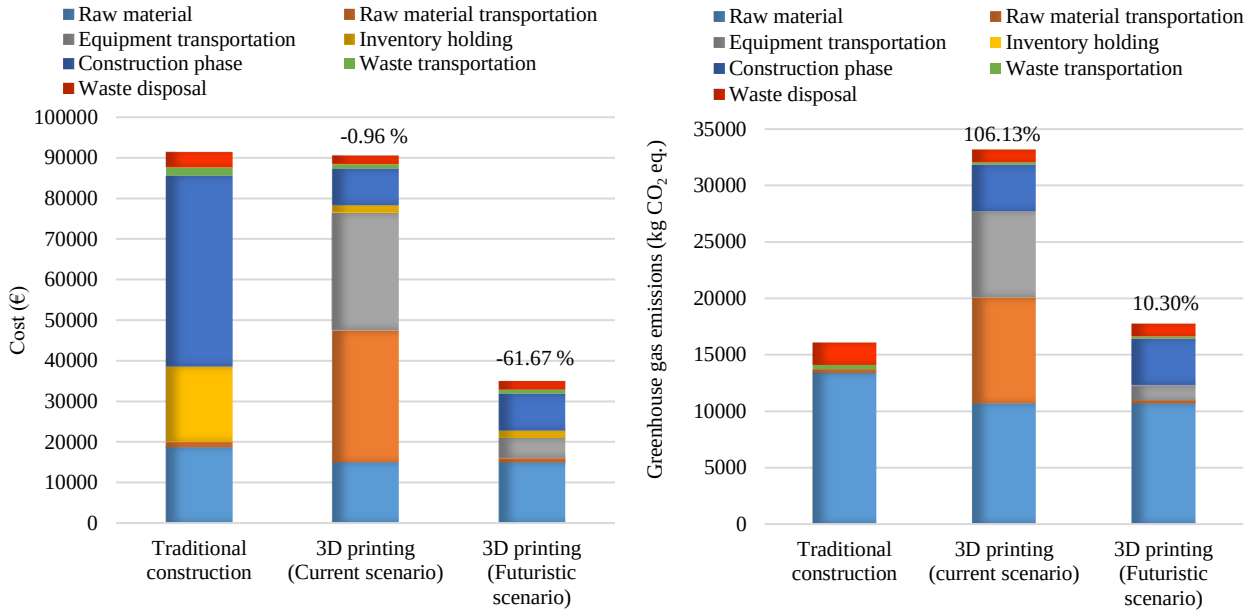


Figure 4. Comparative analysis of traditional construction with a futuristic scenario of 3D printing for (a) total construction cost (b) total greenhouse gas emissions.

construction phase is observed to be 9127.3 €, which is nearly 80% lower than that of traditional construction. In the case of waste disposal, nearly the same distance is considered for both construction methods. Analysis shows that the waste disposal cost of 3D printing (2121.2 €) is lower than that of traditional construction (3778.9 €), which is due to the lower amount of waste produced. Comparative analysis shows that even though 3D printing offers significantly lower construction costs due to a higher level of automation, its total cost (90555 €) is still quite close to the traditional construction cost (91437 €). The prime reason for the higher total cost of 3D printing is the transportation of raw materials and 3D printing equipment from distant places or facilities due to the lack of a well-established market for 3D printing.

3.2 Environmental value stream mapping

EVSM is a valuable tool used to assess the potential sources of greenhouse gas emissions throughout the value stream of the construction process [15]. The analysis shows that GHG emissions from the raw material (13378 kg CO_2 eq.) are significantly higher than any other activity in the value stream. However, the GHG emissions from the raw material for 3D printing (10702.5 kg CO_2 eq.) are lower than those from traditional construction due to the relatively lesser amount of material used. The transportation of raw materials and 3D printing equipment contributed to the GHG emissions to a significant extent of 17023 kg CO_2 eq. which are mainly due to the large distances between the suppliers and construction site. On the other hand, traditional construction hardly commits 353.98 kg CO_2 eq., which shows that transportation is the decisive factor for the selection of environmentally friendly alternatives. In the next phase of construction, 3D printing also contributes 47025 kg CO_2 eq. to the GHG emissions, which are associated with the energy consumed during the concrete mixing, pumping and printing process. As (in this study), a low-rise building is considered where most of the operations are performed manually in traditional construction; therefore, negligible GHG emissions are observed during the construction phase.

However, in the case of high-rise buildings or large construction projects, a significant amount of energy is also consumed in the concrete mixing and pumping process, which generates a competitive amount of GHG emissions. In the case of construction waste transportation and disposal, 3D printing outperformed by producing 200.05 kg CO_2 eq. and 1123.70 kg CO_2 eq. respectively, due to the generation of less construction waste than traditional construction. Net GHG emissions analysis depicts that traditional construction produces almost 51% less GHG emissions than 3D printing construction.

3.3 Futuristic scenario for 3D printing construction

From the previous discussion on value stream assessment, it has been observed that traditional construction outperformed 3D printing in terms of cost and environmental impact for low-rise buildings. The

prime reason for the higher cost and environmental impact was observed to be the larger distances between the suppliers and construction sites for 3D printing. The market and infrastructure for traditional construction are well established, and raw material suppliers (for traditional construction) are frequently available. Moreover, most of the traditional construction companies are located within a 100-kilometer radius, as shown in Figure 3. Conversely, 3D printing technology is not yet fully developed, and suppliers of its equipment and raw materials are not consistently accessible. Thus, it would be premature to conduct a comparative analysis of both techniques at this stage. Therefore, this study presents a futuristic scenario for 3D printing by considering the equipment supplier at 90 km and the raw material supplier at the same as traditional construction (10.9 km). In the current scenario, it has been observed that 3D printing hardly reduces the total cost by 1%, as shown in Figure 4a (the negative sign depicts the percentage reduction). A futuristic scenario for 3D printing shows that 3D printing could reduce the overall cost by 61.67% as compared to traditional construction. In the case of environmental impact, it has been observed that the current scenario produces almost 106% higher GHG emissions, which could be reduced to 10% higher than that of traditional construction (Figure 4b). The higher GHG emissions are associated with energy consumption due to higher levels of automation. However, in the case of large construction (where traditional construction processes also use machines for mixing and pouring), 3D printing has the potential to offer even lower GHG emissions than traditional construction.

4 Conclusion

This study outlines the key factors and activities in traditional construction and 3D printing that incur costs and impact the environment. It then presents the methodology for mapping these activities on the value stream. Finally, real-time case studies are analyzed to evaluate the activities that significantly contribute to costs and environmental burdens. The following conclusions are drawn from the findings of this study.

- In traditional construction, the primary factors driving total costs are raw materials, warehousing, and labor. In contrast, for 3D printing, transportation of raw materials and printing equipment significantly influences overall costs, primarily due to the absence of a well-established infrastructure compared to traditional construction.
- In the case of environmental burden, raw material is the major source of GHG emissions for traditional construction due to the significant amount of material used. On the other hand, transportation of raw materials and equipment remains a significant factor for 3D printing due to the large distances from the suppliers.
- In the construction phase, 3D printing offers nearly 80% less cost than traditional construction, which is mainly due to the significant automation involved. In contrast, labor remained the major factor in the higher cost of traditional construction.
- To assess the potential of 3D printing, a futuristic scenario has been presented, revealing that 3D printing can reduce overall construction costs (including transportation, warehousing, construction, and waste disposal) by nearly 61.67% if raw materials and equipment are available close to the construction site, similar to traditional construction.
- The environmental impact of 3D printing is 106% higher than that of traditional construction, which can be reduced to only 10% higher (for low-rise buildings or small-scale construction) if the raw materials and equipment are readily available. However, the environmental burden of traditional construction could be even higher than 3D printing in large-scale construction projects due to the energy consumption by batching plants and piston pumps.

This study revealed that 3D printing offers an even higher cost and environmental burden at this stage due to the lack of well-established infrastructure same as traditional construction. However, 3D printing could be the potential alternative to traditional construction due to its potential to limit the overall construction cost and environmental burden.

Acknowledgements

This work was supported by Guangdong Special Support Talent Program—Innovation and Entrepreneurship Leading Team (2019BT02S593), RGC Research Impact Fund (R7036-22), RGC Theme-based Research Scheme (T32-707-22-N), and RGC GRF project (17202124).

References

[1] UnitedNations. "Department of Economic and Social Affairs, Population Division, World population." https://www.un.org/en/desa/world-population-projected-reach-98-billion-2050-and-112-billion-2100 (accessed 2025).

[2] H. Hassan, E. Rodriguez-Ubinas, A. Al Tamimi, E. Trepci, A. Mansouri, and K. Almehairbi, "Towards

innovative and sustainable buildings: A comprehensive review of 3D printing in construction," *Automation in Construction*, vol. 163, p. 105417, 2024/07/01/ 2024, doi: https://doi.org/10.1016/j.autcon.2024.105417.

[3] M. H. Raza and R. Y. Zhong, "A sustainable roadmap for additive manufacturing using geopolymers in construction industry," *Resources, Conservation and Recycling*, vol. 186, p. 106592, 2022/11/01/ 2022, doi: https://doi.org/10.1016/j.resconrec.2022.106592.

[4] H. Everett. "3D Hubs AM Trends Report reveals 3D printing grew 21% despite Covid-19. 3D Printing Industry." https://3dprintingindustry.com/news/3d-hubs-am-trends-report-reveals-3d-printing-grew-21-despite-covid-19-189087/ (accessed 04 March 2025.

[5] V. Markin, M. Krause, J. Otto, C. Schröfl, and V. Mechtcherine, "3D-printing with foam concrete: From material design and testing to application and sustainability," *Journal of Building Engineering*, vol. 43, p. 102870, 2021/11/01/ 2021, doi: https://doi.org/10.1016/j.jobe.2021.102870.

[6] Y. Weng *et al.*, "Comparative economic, environmental and productivity assessment of a concrete bathroom unit fabricated through 3D printing and a precast approach," *Journal of Cleaner Production*, vol. 261, p. 121245, 2020/07/10/ 2020, doi: https://doi.org/10.1016/j.jclepro.2020.121245.

[7] M. Batikha, R. Jotangia, M. Y. Baaj, and I. Mousleh, "3D concrete printing for sustainable and economical construction: A comparative study," *Automation in Construction*, vol. 134, p. 104087, 2022/02/01/ 2022, doi: https://doi.org/10.1016/j.autcon.2021.104087.

[8] A. L. M. Tobi, S. A. Omar, Z. Yehia, S. Al-Ojaili, A. Hashim, and O. Orhan, "Cost viability of 3D printed house in UK," *IOP Conference Series: Materials Science and Engineering*, vol. 319, no. 1, p. 012061, 2018/03/01 2018, doi: 10.1088/1757-899X/319/1/012061.

[9] R. Allouzi, W. Al-Azhari, and R. Allouzi, "Conventional Construction and 3D Printing: A Comparison Study on Material Cost in Jordan," *Journal of Engineering*, vol. 2020, no. 1, p. 1424682, 2020/01/01 2020, doi: https://doi.org/10.1155/2020/1424682.

[10] H. Abdalla, K. P. Fattah, M. Abdallah, and A. K. Tamimi, "Environmental Footprint and Economics of a Full-Scale 3D-Printed House," *Sustainability*, vol. 13, no. 21, doi: 10.3390/su132111978.

[11] Y. Han, Z. Yang, T. Ding, and J. Xiao, "Environmental and economic assessment on 3D printed buildings with recycled concrete," *Journal of Cleaner Production*, vol. 278, p. 123884, 2021/01/01/ 2021, doi: https://doi.org/10.1016/j.jclepro.2020.123884.

[12] H. Alhumayani, M. Gomaa, V. Soebarto, and W. Jabi, "Environmental assessment of large-scale 3D printing in construction: A comparative study between cob and concrete," *Journal of Cleaner Production*, vol. 270, p. 122463, 2020/10/10/ 2020, doi: https://doi.org/10.1016/j.jclepro.2020.122463.

[13] S. Besklubova, M. H. Raza, R. Y. Zhong, and M. J. Skibniewski, "3D Printing vs. Traditional Construction: Cost Comparisons from Design to Waste Disposal Stages," in *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, 2024, vol. 41: IAARC Publications, pp. 1247-1254.

[14] M. H. Raza, S. Besklubova, and R. Y. Zhong, "Economic analysis of offsite and onsite 3D construction printing techniques for low-rise buildings: A comparative value stream assessment," *Additive Manufacturing*, vol. 89, p. 104292, 2024/06/05/ 2024, doi: https://doi.org/10.1016/j.addma.2024.104292.

[15] M. H. Raza, S. Besklubova, and R. Y. Zhong, "Comparative environmental value stream assessment of offsite and onsite 3D construction printing processes," *Procedia CIRP*, vol. 130, pp. 1012-1019, 2024/01/01/ 2024, doi: https://doi.org/10.1016/j.procir.2024.10.200.

[16] LetsBuild. "The biggest construction companies in Germany." https://www.letsbuild.com/blog/biggest-construction-companies-in-germany (accessed 06-March-2025).

A Regional Investigation of Near-Zero Energy Buildings: Assessing the Impact of PV Integration and Building Design

Elsayed Salem[1], Emad Elwakil[2] and Ashraf Salem[3]

[1,2]School of Construction Management Technology, Purdue University, United States
[1]Civil Engineering Department, Faculty of Engineering, Kafrelsheikh University, Egypt
[3]Senior Risk Manager, WSP Canada Inc., Canada
esalem@purdue.edu, eelwakil@purdue.edu, ashraf.salem@wsp.com

Abstract –

Building energy research faces challenges in predicting complex interactions, integrating renewable energy, bridging real-world applications, addressing economic feasibility, and designing buildings to adapt to climate conditions. Zero energy building (ZEB) is an innovative approach aiming to generate renewable energy equivalent to building demand, but its development is hindered by industry flexibility and complexity. The research explores the feasibility of achieving near ZEB by utilizing roof photovoltaics technology for energy production and optimizing consumption in four US climate regions. Building Information Modeling (BIM) was used in this study to simulate the building operation energy. Actual Location and real weather data were used to adopt results accurate enough for the study objective. Additionally, the local energy prices were used to costly-wise introduce an effective feasibility. Colorado's model has the highest energy and cost percentages, at (75.34%) and (88.61%), respectively. Georgia ranked second in terms of energy savings (71.21%). Due to the high local energy prices, San Francisco, Ca achieved the largest cost-saving percentage (88.61%), which is like Colorado, despite having a lower percentage of energy savings (59.83%). The study found that wall and roof construction, along with HVAC system types, are effective in optimizing building energy consumption, but its results are limited by climate regions and office building types.

Keywords –

Sustainability; Building Performance; Renewable Energy; Reconstruction; Building Operation; Retrofit; Building Design; Zero Energy Building; ZEB; Energy Analysis; Building Information Modeling; Restoration.

1 Introduction

Urban regions are moving toward sustainable energy by converting the infrastructure into energy-efficient systems. Advanced energy analysis techniques and Building Information Modeling (BIM) are being employed to forecast energy generation and demand to support self-sufficient, and energy-independent communities [1], [2]. Computational techniques provide an accurate and reliable way to address uncertainty and sensitivity analysis in energy efficiency buildings, addressing intricate linkages [3], [4], [5]. The building industry is still in the developing stage for sustainability due to the high complexity of building systems and the involvement of many stakeholders [6]. Therefore, computational techniques are used to overcome this difficulty [7], [8]. Energy and environmental concerns are addressed by integrating renewable and energy-efficient technologies [9], [10], [11]. There are many barriers to the adoption of renewable energy [12]. Reaching the point of optimal building energy performance could be impacted by several factors [13], [14], [15].

The World Green Building Council (WGBC) and the United Nations Framework Convention on Climate Change (UNFCCC) are collaborating to create a common approach for assessing and grading building sustainability. Most scholars use generative computation for aesthetic form-finding without using zero-energy-driven global performance standards, and there is no standard seal or measure for Net-Zero Energy buildings [16]. Using a Digital Twin model, ZEB buildings in the UK have successfully integrated renewable technology into their current buildings. The model has shown a 23-year return period for solar panels and wind turbines but does not provide design recommendations for new buildings [17]. The study [18] revealed that countries plan to convert traditional buildings to Zero Energy, despite challenges like high initial costs for renewable energy systems compared to overall building costs, gaining more interest and support. The study [19] analyzed cost barriers to Zero Energy Building (ZEB) implementation, revealing gaps in energy-saving calculations, comparative studies, life cycle cost analysis, and high initial costs.

BIM and cloud optimization platforms are popular building energy simulation software for simulating building materials and geometry. Cloud-based optimization tools were used to analyze an academic building in Pakistan, focusing only on roof isolation and building orientation, demonstrating the effectiveness of BIM in optimizing energy efficiency [20]. The cloud tools were utilized in an evaluation strategy for the Middle East and North Africa, analyzing building energy use and identifying solutions to reduce it. The study focused on photovoltaic panels and HVAC system enhancement, not considering alternative construction materials [10]. Furthermore, computational design methods were used to improve the energy performance of concrete residential towers in cold climates [21], and pre-trained language model-based methods were used to transform the building information model into an energy model [22].

According to the literature review, the impact of climate on designing a new ZEB or retrofitting existing buildings was not well understood. Additionally, little is known about how climate influences design factors. By examining design features and using special design scenarios to rebuild a sustainable and highly energy-efficient building using BIM software, this paper will bridge this gap by examining four different climates for an office building to achieve near Zero-Energy objective.

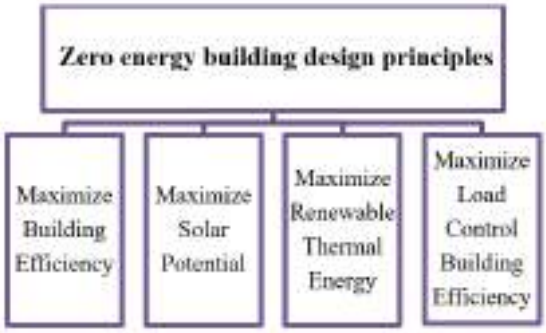


Figure 1. Zero energy building design principles adapted from [25].

Zero energy building (ZEB) is an innovative approach aiming to balance building energy demand and renewable energy production. This study aims to identify the impact of climate on achieving near ZEB by analyzing the unique design scenarios of factors such as location, design, renewable energy use, energy consumption trends, and budget when planning, erecting, restoring, and reconstructing such a high-efficiency building. The literature examining the building designs that can achieve the high-energy performance building concept is severely lacking [18], [19], [23], [24]. This study will cover this gap by modeling the operation performance of an office building using BIM, Energy analysis, and optimization tools. It also applies distinctive design scenarios for each design parameter to achieve near ZEB requirements. Finally, the optimization values for each design parameter were determined based on the impact of reducing the overall building energy consumption and cost. Figure 1 summarizes the main key steps optimize a ZEB.

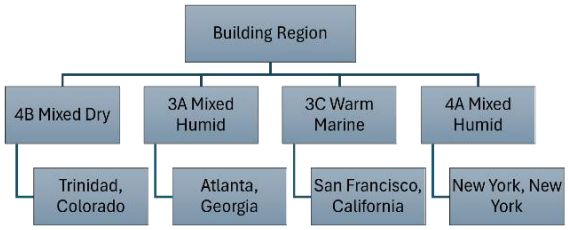


Figure 2. Different building regions

This study aims to fill the literature gap and provide a roadmap for designers and practitioners to adopt passive solar design for high-performance buildings and sustainable targets. It aims to accelerate zero-energy buildings by reducing energy consumption and introducing energy production techniques. The paper investigates the influence of climate region on building energy consumption and solar energy production in four different regions as shown in Figure 2. The climate regions classification is according to Building America and the International Energy Conservation Code (IECC) 2021 Updates as shown in Figure 3. The four locations were selected from different climate regions after excluding extreme climates that are expected to consume very high energy due to the weather.

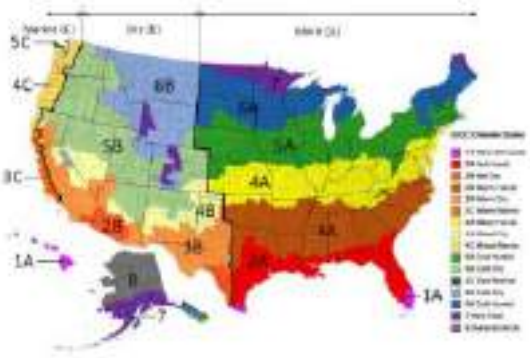


Figure 3. Climate Zone by County according to Building America and International Energy Conservation Code (IECC) 2021 Updates adapted from [26].

The study aims to identify the feasibility of adopting ZEB. It utilizes BIM simulation tools to connect the effects of different variables and deal with the complexity of the building systems. The fundamental presumptions of this research are as follows.

- The selected building represents its local

location for weather and environmental parameters.

- The results generated from the simulation tools are accurate enough for the study objective to compare energy savings for different cases.
- The secondary simulation parameters have a negligible impact when comparing different scenarios.

The study focuses on office buildings and considers the standard energy usage rates for simulating different energy modules. However, additional studies should be conducted to expand the scope due to limited case studies and climate regions.

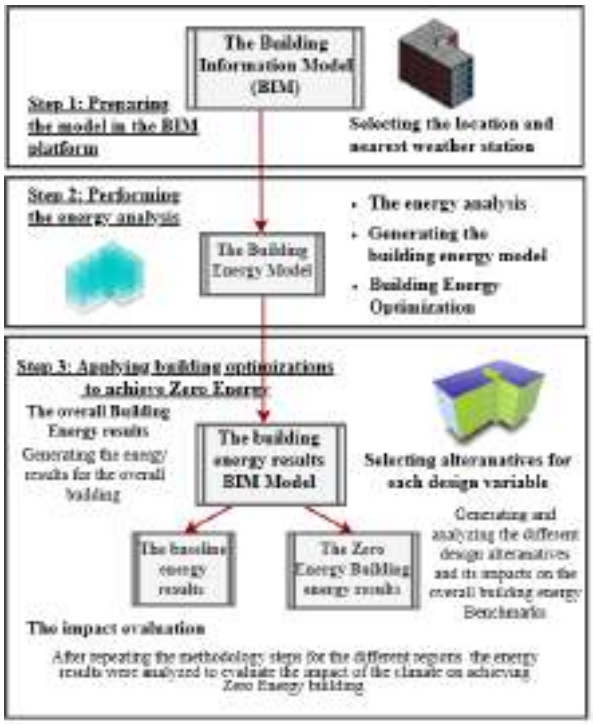


Figure 4. The Study Methodology Flow Chart.

2 Methods

This study is examining the building operation performance in 4 different locations through applying three main steps for each location, as illustrated in Figure 4, starting with step 1: Preparing the model in the BIM platform. Following that, the energy analysis is performed in step 2 to generate the building energy model. In step 3, the building energy model will be used to apply the energy optimizations to reduce the energy consumption and optimize the photovoltaic energy production trying to achieve near ZEB. Finally, the baseline energy and cost will be compared to the near ZEB energy and cost for each location.



Figure 5: BIM Model, Energy Model, and Optimization results model. (BIM model source [27])

2.1 Preparing the model in the BIM platform

Autodesk Revit was used as the BIM tool to model the building geometry and materials. Additionally, its integrated GIS (Geographic Information System) functionality enabled the incorporation of location-specific weather data, enhancing the analysis with real-world environmental conditions.



Figure 6: The selection of the building location and the nearest Weather Station.

Figure 5 illustrates BIM Model, Energy Model, and Optimization results model. For instance, Figure 6 shows the selected location and the nearest available weather station, and Figure 7 reports the monthly temperature rates for first location.



Figure 7. The monthly temperature rates in the selected location.

2.2 Performing the energy analysis

The energy analysis tools were used to perform a complete energy analysis. Figure 5 shows the generated building energy model.

Finally, the energy model for each location were exported to the optimization platform. At this step, different variables for each design parameter were applied, which gave different results for each variable. For example, different wall construction types including (R13 Wood- Uninsulated- R2CMU- R3 Metal- R13+R10 Metal- 14-inch ICF- 12.25-inch SIP- R38Wood) gave the Energy Use Intensity (EUI) results shown in Figure 8. Figure 5 illustrates the results of the BIM model generated in the optimization platform.

Wall construction Types	R13 Wood Uninsulated	R2CMU	Baseline (BIM)	R3 Metal	R13+R10 Metal	14-inch ICF	12.25-inch SIP	R38Wood
Energy Impact (kwh/m²/yr.)	+50	+30	+10	0	-5	-45	-48	-49	-50

Figure 8. The energy impact of using different Wall Construction Types in Trinidad, CO - 4B Mixed Dry region.

2.3 Applying building optimizations to achieve zero energy

Different energy scenarios are achieved by setting distinctive design parameters to a specific value. For example, wall construction materials can be set to different types. A total of 24 design parameters were used in the optimization process as given in Figure 9. A new scenario is added to the results after setting all the design parameters to a specific value. This study applied the scenario that achieves the ZEB target to compare it with the baseline scenario. The design parameters were determined based on the impact of energy consumption and cost reduction. The findings of this study were analyzed and compared at the end of this paper.

Design Parameters

- Building Orientation (Degrees)
- Daylighting & Occupancy Controls
- HVAC Types
- Infiltration (ACH)
- Lighting Efficiency
- Operating Schedule
- Panel Efficiency
- Payback limit
- Plug Load Efficiency
- Roof Construction
- Surface Coverage
- Wall Construction
- Window Glass Types - North
- Window Glass Types - East
- Window Glass Types - West
- Window Glass Types- South
- Window Shades - East
- Window Shades - North
- Window Shades - South
- Window Shades - West
- WWR - Eastern Walls
- WWR - Northern Walls
- WWR - Southern Walls
- WWR - Western Walls

Figure 9. Design Parameters

3 Results

The study results were analyzed to visualize and evaluate each building location for constructing or retrofitting existing buildings to become a near ZEB. Figure 10, and Figure 12 explain the comparison analysis for the building energy consumption and cost, respectively for the four building locations for the baseline scenario. Figure 12, and Figure 13 compare the results of building energy consumption values for the four building regions for the near ZEB scenario with the baseline results.

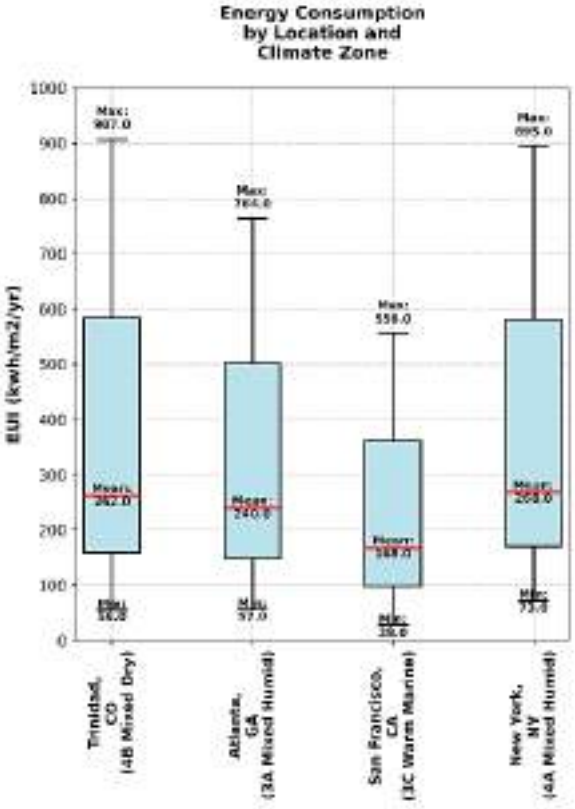


Figure 10. Comparing the Minimum, Mean, and

Maximum building energy consumption results by location and climate zone for the baseline scenario.

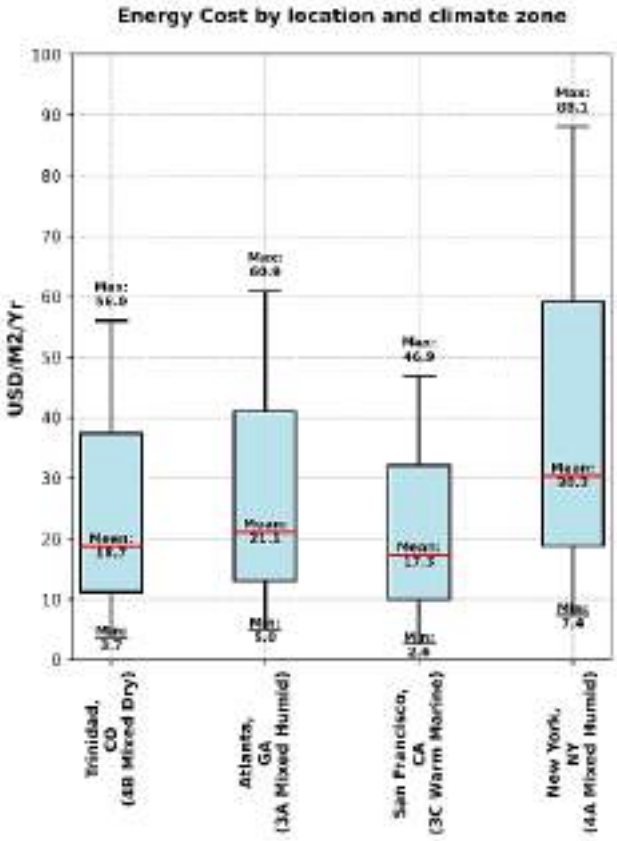


Figure 11. Energy Cost by location and climate zone for the baseline scenario.

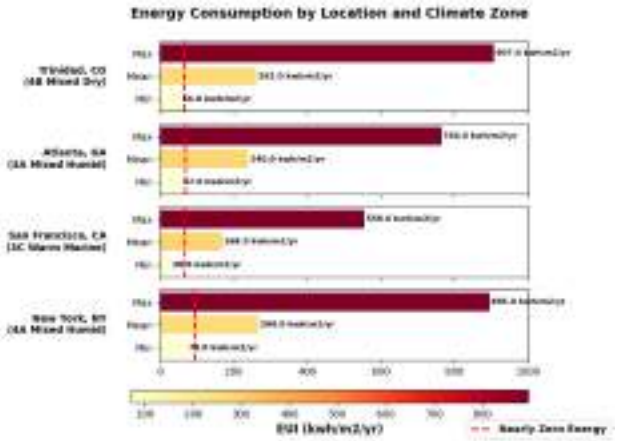


Figure 12. Comparing the building energy consumption results by location for the near ZEB scenario with the baseline results.

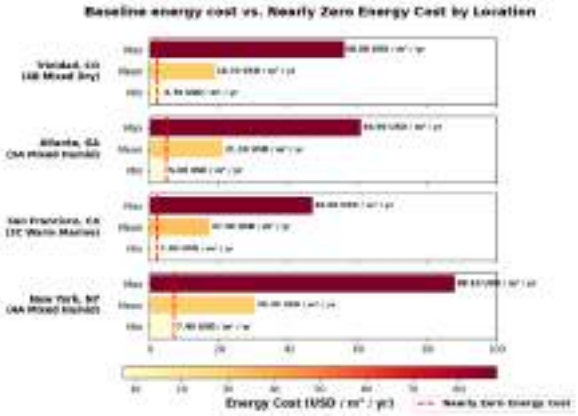


Figure 13. Comparing the building energy cost results by location for the near ZEB scenario with the baseline results.

The near ZEB results were compared with the baseline mean values to visualize the feasibility in energy and cost savings between the four locations as illustrated in Figure 14, and Figure 15.

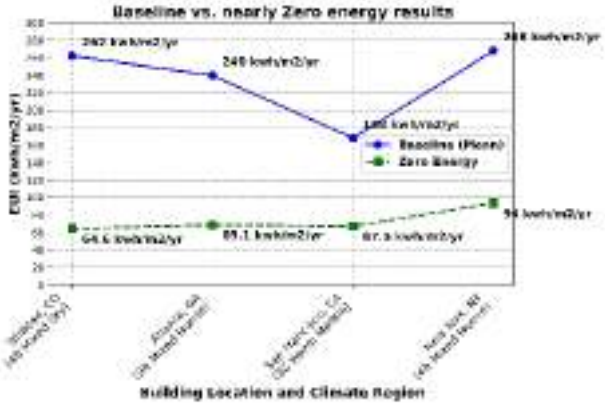


Figure 14. Comparing the near ZEB energy with the baseline Mean results.

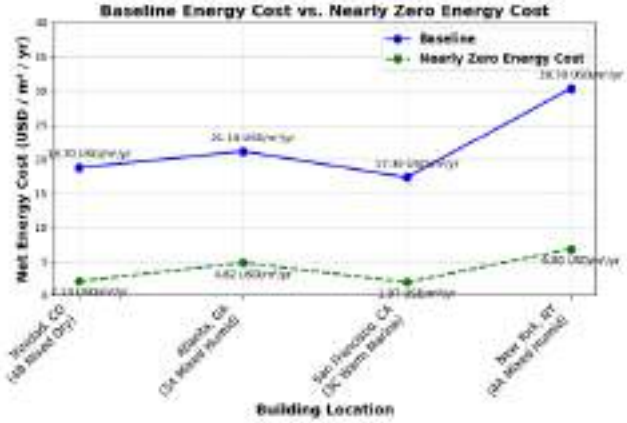


Figure 15. Comparing the near ZEB cost with the baseline Mean results.

Finally, the percentages of energy and cost savings were calculated and plotted in Figure 16

to find the best near ZEB locations. The model in Colorado recorded the best energy and best cost with (75.34%) and (88.61%). Georgia was the second-best location in energy savings (71.21%). However, the lower percentage of energy savings (59.83%) in San Francisco, CA, it recorded the highest cost saving percentage (88.61%) same as Colorado, due to the high local energy prices. New York was the 3rd in energy and cost saving percentages recording (64.93%) and (77.56%), respectively.

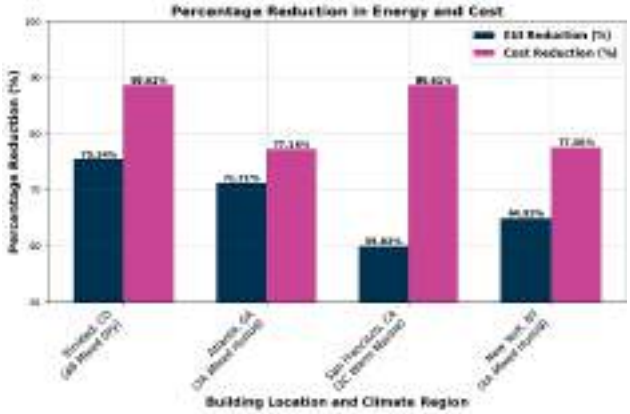


Figure 16. Comparing the percentages of reduction in energy and cost by location.

4 Discussion and Conclusion

This study assessed the influence of building location on operational performance by examining four distinct office building locations in different climate regions in United States including Colorado, Georgia, San Francisco (California), and New York. Retrofitting strategies were implemented to enhance operational efficiency and achieve the objectives of a near Zero Energy Building (ZEB). The baseline energy consumption of each location was first simulated, and subsequent performance outcomes were evaluated after applying various design alternatives. Building Information Modeling (BIM) was employed to simulate building geometry, while energy optimization tools generated energy analysis models. These models were further analyzed using an optimization platform. The evaluation adhered to the World Green Building Council's zero-energy building standards, assessing the feasibility of retrofitting, restoration, or reconstruction and the potential energy and cost savings across the four climate regions.

The study results revealed that the operation performance is highly sensitive to the building location. The model in Colorado recorded the best energy and cost savings (75.34%) and (88.61%), Respectively. Georgia was the second-best location in energy savings (71.21%). However, the lower percentage of energy savings (59.83%) in San Francisco, CA, it recorded the highest cost savings (88.61%) same as Colorado, due to the high local energy prices. New York was the 3rd in energy and cost savings (64.93%) and (77.56%), respectively.

The results are limited to the four climates included in this study and office buildings. Future work is highly recommended in other climates and for other building types. Future work will focus on validating the proposed method using real-world data to further assess its accuracy and performance. Also, simultaneously simulating a group of buildings may have additional impacts on overall building consumption. This paper contributes to the acceleration of ZEB targets and sustainable buildings by introducing an evaluation of the energy and cost savings for designers, stakeholders, and governments. This study contributes to the field by applying existing methods, such as BIM simulations and energy modeling, to specific climate regions, highlighting the importance of location-specific factors in optimizing energy efficiency. This approach offers valuable insights for real-world applications, demonstrating the potential for significant energy and cost savings based on regional conditions.

References

[1] A. Buonomano, G. Barone, and C. Forzano, "Latest advancements and challenges of technologies and methods for accelerating the sustainable energy transition," *Energy Reports*, vol. 9, pp. 3343–3355, Dec. 2023, doi: 10.1016/J.EGYR.2023.02.015.

[2] H. Elba, H. Hegazy, J. Zhang, I. M. Mahdi, I. Abdel Rashid, and H. M. Hassan, "Exploring critical barriers towards the uptake of renewable energy usage in Egypt," *Innovative Infrastructure Solutions*, vol. 9, no. 7, pp. 1–17, Jul. 2024, doi: 10.1007/S41062-024-01578-3/METRICS.

[3] W. Tian, "Towards advanced uncertainty and sensitivity analysis of building energy performance using machine learning techniques," *J Build Perform Simul*, pp. 1–8, Aug. 2024, doi: 10.1080/19401493.2024.2387071.

[4] A. M. El-Kholy, M. Y. El-Shikh, and S. K. Abd-Elhay, "Which Fuzzy Ranking Method is Best for Maximizing Fuzzy Net Present Value?," *Arab J Sci Eng*, vol. 42, no. 9, pp. 4079–4098, Sep. 2017, doi: 10.1007/S13369-017-2573-4/METRICS.

[5] E. Salem, E. Elwakil, and M. Hegab, "Emerging Trends in Construction Dispute Resolution: Integrating Technology and Collaboration for the Future," *Sustainable Civil Infrastructures*, pp.

139–150, 2024, doi: 10.1007/978-3-031-76557-5_12.

[6] E. Elwakil, "Life Cycle Cost Analysis of Energy Efficient Single Family Residence," 2019, doi: 10.14455/ISEC.res.2019.102.

[7] J. Elsamni, H. Elhegazy, I. Abdulrasheed, and M. Kohail, "A conceptual framework for sustainable construction in the Arab region," *Innovative Infrastructure Solutions*, vol. 9, no. 6, pp. 1–6, Jun. 2024, doi: 10.1007/S41062-024-01501-W/FIGURES/6.

[8] S. M. Soliman, H. M. Elhegazy, J. Zhang, I. M. Mahdi, A. K. Dessouki, and I. Abdel Rashid, "Key Factors Affecting Steel Construction Projects' Performance by Adopting Building Information Modeling (BIM) in Egypt," *Computing in Civil Engineering 2023: Visualization, Information Modeling, and Simulation - Selected Papers from the ASCE International Conference on Computing in Civil Engineering 2023*, pp. 442–450, 2024, doi: 10.1061/9780784485231.053.

[9] A. Hajare and E. Elwakil, "Integration of life cycle cost analysis and energy simulation for building energy-efficient strategies assessment," *Sustain Cities Soc*, vol. 61, Oct. 2020, doi: 10.1016/j.scs.2020.102293.

[10] Y. W. Al-Saeed and A. Ahmed, "Evaluating Design Strategies for Nearly Zero Energy Buildings in the Middle East and North Africa Regions," *Designs 2018, Vol. 2, Page 35*, vol. 2, no. 4, p. 35, Sep. 2018, doi: 10.3390/DESIGNS2040035.

[11] A. Shetty, K. S. Samarth, S. K. Naik, H. T. M. Setty, and K. Hannah, "Retrofit Guidelines Towards Achievement of Net Zero Energy in a University Building," *Lecture Notes in Civil Engineering*, vol. 256, pp. 231–243, 2023, doi: 10.1007/978-981-19-1862-9_15/FIGURES/7.

[12] H. Elba, H. Hegazy, J. Zhang, I. M. Mahdi, I. Abdel Rashid, and H. M. Hassan, "Exploring critical barriers towards the uptake of renewable energy usage in Egypt," *Innovative Infrastructure Solutions*, vol. 9, no. 7, pp. 1–17, Jul. 2024, doi: 10.1007/S41062-024-01578-3/TABLES/6.

[13] S. H. Alyami, A. A. Almutlaqa, A. M. Alqahtany, and N. Ashraf, "Likelihood of Reaching Zero Energy Building Design in Hot Dry Climate: Saudi Arabia," *IEEE Access*, vol. 9, pp. 167054–167066, 2021, doi: 10.1109/ACCESS.2021.3134365.

[14] B. Lin and Z. Chen, "Net zero energy building evaluation, validation and reflection – A successful project application," *Energy Build*, vol. 261, Apr. 2022, doi: 10.1016/j.enbuild.2022.111946.

[15] E. Salem and E. Elwakil, "Innovative methods to optimize the integration of passive solar design principles into buildings," *Innovative Infrastructure Solutions*, vol. 9, no. 9, pp. 1–15, Sep. 2024, doi: 10.1007/S41062-024-01635-X/FIGURES/10.

[16] T. Spiegelhalter, "ACHIEVING THE NET-ZERO-ENERGY-BUILDINGS '2020 AND 2030 TARGETS' WITH THE SUPPORT OF PARAMETRIC 3-D/4-D BIM DESIGN TOOLS," *Journal of Green Building*, vol. 7, no. 2, pp. 74–86, Apr. 2012, doi: 10.3992/JGB.7.2.74.

[17] S. Kaewunruen, P. Rungskunroch, and J. Welsh, "A Digital-Twin Evaluation of Net Zero Energy Building for Existing Buildings," *Sustainability 2019, Vol. 11, Page 159*, vol. 11, no. 1, p. 159, Dec. 2018, doi: 10.3390/SU11010159.

[18] E. Salem and E. Elwakil, "The Challenges of Transition from Traditional to Net Zero Energy Buildings," *2023 IEEE Conference on Technologies for Sustainability, SusTech 2023*, pp. 104–110, 2023, doi: 10.1109/SUSTECH57309.2023.10129611.

[19] E. Salem and E. Elwakil, "Minimizing the Cost Gap between Net Zero Energy and Conventional Buildings," in *2023 IEEE Conference on Technologies for Sustainability, SusTech 2023*, 2023. doi: 10.1109/SusTech57309.2023.10129570.

[20] F. F. Taha, W. A. Hatem, and N. A. Jasim, "Effectivity of BIM technology in using green energy strategies for construction projects," *Asian Journal of Civil Engineering*, vol. 21, no. 6, pp. 995–1003, Sep. 2020, doi: 10.1007/S42107-020-00256-W.

[21] S. Shahi, P. Wozniczk, T. Truyensb, I. Trudeaub, and C. Haasa, "Energy performance and lca-driven computational design methodology for integrating modular construction in adaptation of concrete residential towers in cold climates," *Proceedings of the 37th International Symposium on Automation and Robotics in Construction, ISARC 2020: From Demonstration to Practical Use - To New Stage of Construction Robot*, pp. 1061–1068, 2020, doi: 10.22260/ISARC2020/0146.

[22] Z. Wang, M. Bergés, and B. Akinci, "Pre-trained language model based method for building information model to building energy model transformation at metamodel level," *Proceedings of the International Symposium on Automation*

and Robotics in Construction, pp. 17–25, 2024, doi: 10.22260/ISARC2024/0004.

[23] E. Salem and E. Elwakil, "On-Site Zero Energy by Integrating Photovoltaic Technologies into Buildings," pp. 281–288, Jun. 2024, doi: 10.1109/SUSTECH60925.2024.10553621.

[24] E. Salem, E. Elwakil, and M. Hegab, "Unlocking the Path to Zero Energy Buildings: Exploring Key Determinants for Success," *Sustainable Civil Infrastructures*, pp. 298–306, 2024, doi: 10.1007/978-3-031-76557-5_23.

[25] B. Polly *et al.*, "From Zero Energy Buildings to Zero Energy Districts", Accessed: Jan. 12, 2025. [Online]. Available: http://www.compactofmayors.org/

[26] C. A. Antonopoulos, T. L. Gilbride, E. R. Margiotta, and C. E. Kaltreider, "Guide to Determining Climate Zone by County: Building America and IECC 2021 Updates," Sep. 2022, doi: 10.2172/1893981.

[27] "Popular models | 3D CAD Model Collection | GrabCAD Community Library." Accessed: Dec. 23, 2024. [Online]. Available: https://grabcad.com/library

Signals vs. Videos: Advancing Motion Intention Recognition for Human-Robot Collaboration in Construction

Charan Gajjala Chenchu[1], Kinam Kim[2*], Gao Lu[2], and Zia Ud Din[2]

[1]Department of Computer Science, University of Houston, Texas, USA
[2]Department of Construction Management, University of Houston, Texas, USA
[*]Corresponding Author

cgajjala@cougarnet.uh.edu, kkim48@central.uh.edu, lgao5@central.uh.edu, uziauddi@central.uh.edu

Abstract -

Human-robot collaboration (HRC) in the construction industry depends on precise and prompt recognition of human motion intentions and actions by robots to maximize safety and workflow efficiency. There is a research gap in comparing data modalities—signals and videos—for motion intention recognition. To address this, the study leverages deep learning to assess two different modalities in recognizing workers' motion intention at the early stage of movement in drywall installation tasks. The Convolutional Neural Network–Long Short-Term Memory (CNN-LSTM) model utilizing surface electromyography (sEMG) data achieved an accuracy of around 87% with an average time of 0.04 seconds to perform prediction on a sample input. Meanwhile, the pre-trained Video Swin Transformer combined with the transfer learning technique, harnessed video sequences as input to recognize motion intention and attained an accuracy of 94% but with a longer average time of 0.15 seconds for a similar prediction. This study emphasizes the unique strengths and trade-offs of both data formats, directing their systematic deployments to enhance HRC in real-world construction projects.

Keywords -

Human-Robot Collaboration (HRC); Motion Intention Recognition; Surface Electromyography (sEMG); Convolutional Neural Network–Long Short-Term Memory (CNN-LSTM); Video Swin Transformer

1 Introduction

Human-robot collaboration (HRC) refers to the mutual coordination between humans and robots in a collaborative workspace, where robots can understand the environmental context, and human actions and can make decisions to supplement human tasks. Due to recent advancements in hardware, software, and deep learning algorithms, the performance of construction robots has increased in construction projects in terms of safety, efficiency, and productivity of construction workers [1]. Despite the advancements, construction robots still face challenges due to the dynamic and rapidly changing environments where understanding human behaviors with precise timing is essential [2]. Hence, it is critical to recognize workers' motion intention to enable HRC in construction [3].

Human motion intention recognition is being researched and implemented by modeling machine learning algorithms on two primary modalities: sensor and video data. Among sensor-based approaches, surface electromyography (sEMG) sensors have been utilized to capture the muscle activity of human because they represent muscle movements and their acquisition is easier due to their non-intrusive nature [4]. Meanwhile, videos are also being extensively utilized to analyze human motion with emerging computer vision and image processing technologies [5]. Videos can be used to predict the motion trajectories of workers, equipment, and harmful actions to enable safety monitoring in construction sites [6, 7].

However, these two modalities have their limitations. EMG signals are prone to noise and variability across different subjects which make generalizing patterns difficult [4]. It may be uncomfortable for workers to carry sensor with them. On the other hand, videos require high computational power for model training and inference. Video-based approaches are vulnerable to occlusions, varying viewpoints, background clutters, etc. There have been studies exploring these approaches separately, but the existing studies lack a solid comparison of these modalities in this recognition task, especially in construction.

To address this research gap, this study proposes a comprehensive comparison of the applications of sEMG and video-based modalities as input to deep learning algorithms in human motion intention recognition in construction. This research implements a Convolutional Neural Network–Long Short-Term Memory (CNN-LSTM) model to analyze sEMG signals and utilizes a pre-trained Video Swin (Shifted Window) Transformer model to train on video input. Finally, the performance of these two models is evaluated and compared. This proposed system could provide insights into tradeoffs between their performance and drawbacks, which can help to choose an appropriate approach among the two options for HRC in construction.

2 Literature Review

To enable adaptive and seamless cooperation between human workers and robotic systems, the construction industry has centered on integrating human motion recognition and behavioral analysis into HRC. This included sensors, wearable devices, computer vision, and machine learning models, to capture and interpret human movements in real-time [1, 8]. This section reviews the state-of-the-art studies in human motion recognition using two primary modalities and their applications in construction.

2.1 Signal-based Methods

sEMG and other signal-based methods are extensively adopted in industrial and construction applications [3]. These methods can predict human movements before they occur which can enhance safety and operational efficiency of exoskeleton robots [9]. sEMG signals were also used to recognize hand gestures and control industrial robots in the context of a smart factory [10].

Shah and Kim [3] proposed an sEMG-based motion intention recognition system at an early stage by analyzing the first 1-second muscle activity from four muscle sensors. They implemented a CNN-LSTM model and achieved an accuracy of 98.5% in classifying four construction-related activities. Combining convolutional layers with gated recurrent unit layers (GRU), a hybrid model was trained on sEMG signals [10]. This model was tested on 11 subjects and obtained around 96% accuracy in detecting 10 types of hand gestures. A CNN-LSTM-based system was developed to predict various motion intentions in a squat activity which represents lifting motion in industrial environments [9]. This methodology can catch the movement initiations 0.3 seconds ahead using sEMG signals with an accuracy of 84.02%.

While sEMG plays a key role in human motion detection, several studies show that these signals are prone to variability created by factors such as electrode placement, muscle fatigue, skin impedance, environmental weather conditions, and subject-level bias in terms of gender and age, which leads to noise that can hamper accuracy in action detection [4, 11].

2.2 Video-based Methods

Tang et al. [7] presented a methodology for utilizing the LSTM model's encoder-decoder architecture combined with multi-head prediction modules to predict the future paths of workers and equipment on construction sites using video data. A transformer-based model [6] was proposed for recognizing unsafe gestures on construction sites which applied transformer and 3D CNN to extract spatial and temporal features. It achieved a precision of 88.7% on a customized dataset outperforming baseline methods. For swimming action recognition, a novel multimodal approach called Swintrans Net [12] was presented, building upon the Swin-Transformer's [13] superior feature extraction and adaptability.

Video-based motion recognition systems on construction sites also encounter several challenges. Blocked camera views by frequent obstructions from materials, machinery, and workers, result in incomplete or inaccurate motion data. With constantly changing layouts, equipment, and the dynamic nature of these environments, it is difficult to develop robust systems capable of adapting to unpredictable conditions [2]. Real-time motion detection can be a technical problem due to the processing time of high-resolution videos. Additionally, varied lighting conditions, such as low light, shadows, and glare, can further degrade the performance [14].

2.3 Research Gap

Comparative analysis between sEMG and video data sources in motion intention recognition is relatively scarce, specifically in the context of HRC in construction. One potential reason is due to the limitation of publicly available datasets containing both sensor and video data types that represent similar experimental settings [5]. Two data types can be utilized in analyzing human motion in their own way, but they also have some shortcomings. This presents an opportunity to compare the two approaches in terms of dataset dimensions, deep learning models, accuracy, performance, and prediction latency. This systematic comparison can aid researchers in developing more effective motion intention recognition systems tailored to specific needs and tasks. For the sEMG, we chose a CNN-LSTM model since it outperforms non-recurrent models in time series analysis [15]. Video Swin Transformer model was chosen for its advantage over 3D CNNs in capturing long-range spatial and temporal dependencies in video sequences using self-attention mechanisms [6, 12].

3 Proposed Methodology

The proposed methodology starts with data acquisition of raw sEMG signals and videos which are then preprocessed as inputs to the models. The obtained training curves and test evaluation metrics are interpreted for a comparative analysis. The system workflow is depicted in Figure 1.

3.1 Signal-based Approach

3.1.1 Data Acquisition and Preprocessing

The sEMG signal data was captured from 3 subjects performing four drywall installation activities multiple times. This data consists of four channels representing signals

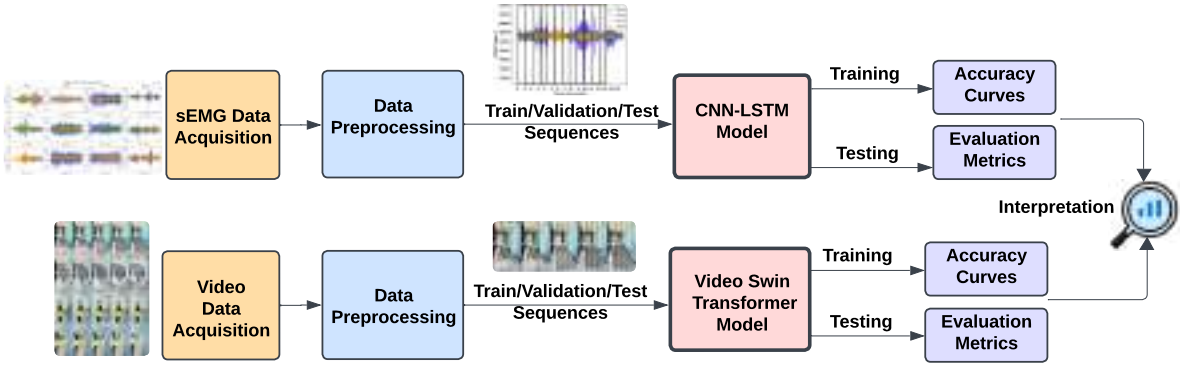


Figure 1. Overall System Architecture

coming from four EMG electrodes, which were attached to the biceps brachii, triceps brachii, shoulder, and forearm muscles of the dominant hand, respectively. These sensor placements are effective in capturing the muscle activity of these actions [3]. A baseline reference node was attached to the subjects on top of the wrist of the same hand. This data was stored in a text file in a tabular format in which rows are time series data points with each point containing 4 columns corresponding to the 4 channels. Hence, one such text file represents the muscle signals captured during one subject doing one of the drywall activities one time.

The activities were divided into "Intention" group, which constitutes the first 1 second of the activity, and the rest as "Actual" group. The time series signals were divided into fixed-size sequences. A sequence represents a small fraction of continuous movements. Consequently, the motion detection task was translated into a classification task with 2 classes (corresponding to the two groups) per each activity, totaling 8 labels. In this way, the whole dataset of text files was divided into overlapping chunks with a sliding window approach. Due to the small length of the intention group, the number of chunks that belong to this intention group became significantly smaller compared to chunks in the actual group which can lead to a major class imbalance within the groups. To tackle this, minority intention classes were randomly oversampled by duplicating the existing samples by 2 times ensuring uniform distribution across activities. This dataset was then divided into train, validation, and test splits ensuring stratification for a uniform and representative distribution of each activity and subject type. A random Gaussian noise was added to the training set consistently to each data point in a sequence as part of the data augmentation. Consequently, these three sets were normalized using a standard scaler that transforms the data to have a mean of 0 and a standard deviation of 1.

3.1.2 Deep Learning Model

The proposed model architecture consists of two 1D CNN blocks followed by two LSTM layers and two fully connected dense layers at the end as shown in Figure 2. This hybrid model can process both spatial and temporal features inherent in sEMG data. The initial 1D CNN layers are designed for the extraction of local patterns in the form of spatial features from the multichannel signals. Next, the output activations are normalized across the batch using the batch normalization (Batch Norm) layer. After this, max-pooling layers are employed to retain only the most useful patterns. To mitigate overfitting, dropout layers are integrated to randomly nullify the contributions of a fraction of input units during training. This can make the model generalize well to the training data by preventing it from overly relying on specific hidden units. The second CNN block is added with an increase in filter size from 64 to 128 (see Figure 2). Following this second block, LSTM units are deployed to exploit their inherent feedback connections in the form of memory gates and cells [3]. This enables the analysis of temporal dependencies in the sEMG sequences without the problem of vanishing gradients. The multidimensional output from the LSTM layers is transformed into a 1D vector using a flatten layer before passing them to fully connected dense layers. Finally, these learned attributes are mapped to the 8 desired output classes using a dense layer with softmax activation.

3.2 Video-based Approach

3.2.1 Data Acquisition and Preprocessing

While the subjects performed the activities equipped with the sEMG sensors, videos were recorded using a typical smartphone at the same time. A single data point in this context refers to a single image frame. One video file corresponds to the action sequence generated by one

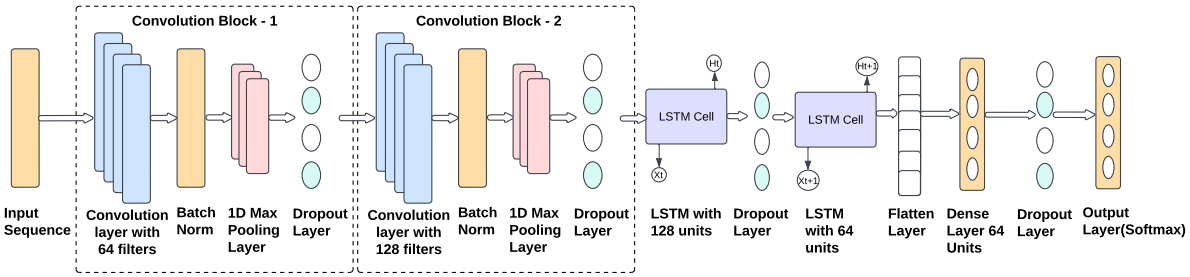


Figure 2. Proposed CNN-LSTM Model Architecture

subject performing one type of activity at a time. Each video was also partitioned into Intention and Actual activity groups with the first 1 second interval labeled as Intention classes. All the videos were resized to 224×244. The pixel values were normalized to have a range between 0 and 1 to improve model convergence and stability during training. These frames were further segmented into overlapping sequences that represent small fragments of continuous movements. The created chunks were divided into train, validation, and test splits such that they contained a representative distribution of sequences in terms of actions and subjects. The labeling strategy used in the signal-based approach was deployed in the video-based approach. Common image data augmentations were applied to the training dataset, ensuring all the frames in a sequence receive the same transformation. This sequence-level augmentation virtually increases training data size, so that models can generalize over varying noise, occlusions, and other real-world conditions.

3.2.2 Deep Learning Model

In this study, the Video Swin Transformer architecture was implemented for the classification task from videos [16]. Initial tokenization followed by a patch merging layer maintains the hierarchical structure of the model. The Video Swin Transformer block [16] starts with a 3D window-based multi-head self-attention (MSA) module, which processes tokens within independent 3D windows to capture specific spatiotemporal relationships in the video. This block also employs a 3D shifted window approach between layers which integrates context from adjacent areas and frames. This lets it recognize long-running complex interactions and actions throughout the video input. To learn more detailed patterns, multilayer perceptron (MLP) is applied after the MSA module to transform features non-linearly.

The proposed approach uses the existing base version of the Video Swin model (Video Swin-B) which was pre-trained on a dataset called Kinetics 600 [16]. We employed a transfer learning strategy, starting with these pre-trained weights frozen for training. The final MSA module, layer normalization, and dense layers of the last transformer block in Stage 4 are unfrozen to update their weights (see Figure 3). However, the MLP layer was retained to be frozen to preserve the high-level non-linear feature transformations learned during pretraining. This makes the model fine-tune to only the attention-driven feature extraction process, reducing complexity and overfitting. Finally, a custom classification head is added that contains a dense output layer with 8 neurons (corresponding to the target classes) at the end, as shown in Figure 3. This head is added to scale complex representations extracted by the transformer and finally map the reduced features to class probabilities using softmax activation.

4 Experiment

4.1 Data Acquisition

We conducted an experiment that involved 3 subjects performing four drywall installation tasks as displayed in Figure 4. These tasks [3] were selected because of their potential to be supported by construction robots. Each subject performed these tasks four times with varying durations while carrying four sEMG muscle sensors and a reference electrode. Simultaneously, these activities were recorded using a smartphone camera. Using OpenSignals software, the sEMG signals (see Figure 4) were collected and stored in a ".txt" format at a frequency of 500 Hertz (Hz), implying that for one second 500 data rows, each with four channel columns would be produced. Videos were filmed in ".mp4" format at a resolution of 1920×1080 pixels with 60 frames per second (FPS). In this way, the original raw dataset of sensor signals contains 16 text files for each subject with 4 files per activity per subject, totaling 48 files for the 3 subjects. The video dataset also contains the same number of video files. Table 1 provides a comprehensive overview of the experimental setup including the raw data point counts.

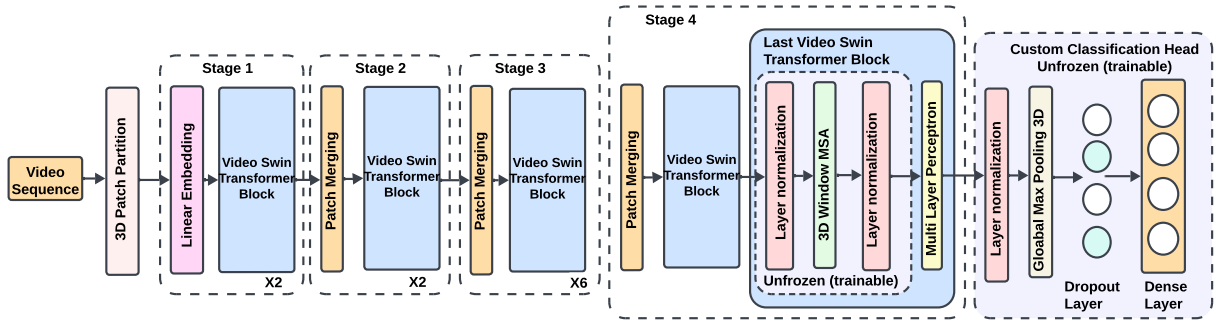


Figure 3. Video Swin Transformer (Base version) Architecture with Custom Classification Head

Table 1. Data Comparison Between Signal-Based and Video-Based Approaches

Method	Data	File Format	Frequency	File Count	Data Point Count
Signal	sEMG signals	.txt	500 Hz	48	195,450
Video	RGB Videos	.mp4	60 fps	48	20,690

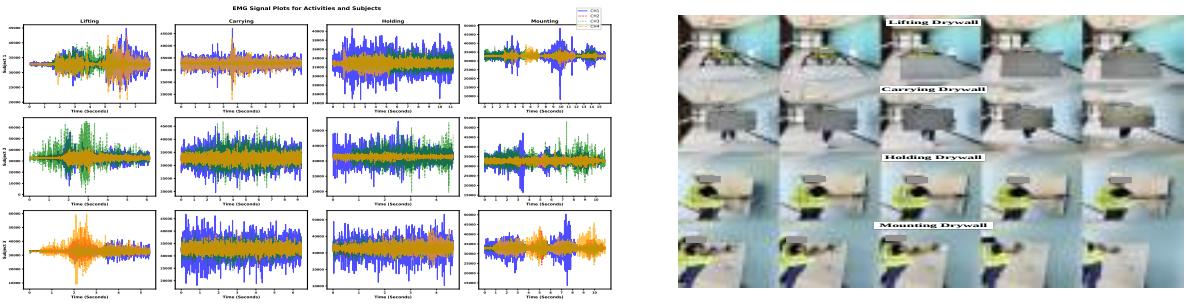


Figure 4. Sample sEMG Signals (Left) and Video Frames (Right) of activities performed by 3 subjects

4.2 Data Preprocessing

Following Shah and Kim [3], the first 1 second of each activity was regarded as the Intention group and the rest as the Actual group. As a result, 8 target classes, namely Lifting Intention, Actual Lifting, Carrying Intention, Actual Carrying, Holding Intention, Actual Holding, Mounting Intention, and Actual Mounting, were formed. This can enable robust recognition of motion triggers and their transition to actual motion. For implementation purposes, the intention groups from the video dataset were divided into separate video clips, producing a total of 96 videos. Due to the size of the sensor data points being significantly higher compared to videos, the video files were increased by 2 times using data augmentations, yielding a final dataset of 288 videos. The signal data was processed into sequences in each of 100 data rows, while each sequence extracted from videos contains 32 frames. Thus, a sequence of signals equates to 0.2 seconds of movement and a video sequence encloses 0.5 seconds of motion. Such an individual sequence becomes the single unit of input to the models. To increase the classification accuracy, both datasets were fragmented into overlapping sequences with a ratio of 50%. To overcome the imbalance between "Intention" and "Actual" classes, the intention sequences were oversampled 3 times for signal data, but the video sequences were kept unaltered since they were already increased.

The EMG sequences were split into train, validation and test splits in the ratio of 70:15:15 respectively. The individual video files were first divided into train, validation, and test splits in the ratio of around 67%, 17%, and 17% respectively, maintaining the representative distribution of activity and subject types. Random Gaussian noise was applied to signal training examples. For the video training dataset, along with Gaussian noise, image, augmentations like random horizontal flip, hue, brightness, saturation, and contrast were employed. Finally, the sensor data was scaled using a standard scaler whereas the image frames were resized to a resolution of 224×244 and then scaled by dividing pixel values by 255.

Table 2. Comparison of Data Preprocessing: Signal-Based vs. Video-Based Approaches

Method	Intention Group	Resolution	Seq. Length	Overlap	Oversampling
Signal	1 second	1 row, 4 channels	100	50%	3x
Video	1 second	1 Frame (224x224, RGB)	32	50%	None

4.3 Training and Hyperparameter Tuning

The CNN-LSTM model was trained on the signals training dataset and the pre-trained Video Swin-B model was partially fine-tuned on the training video sequences. A dedicated validation dataset was used for guiding the model selection process due to its capability of offering an unbiased estimate of model performance. For both models, sparse categorical cross-entropy loss was used to handle the integer encoded true labels and Adam optimizer with fixed learning rate (LR) was employed to leverage the adaptive learning rate. Class weights were incorporated during training to tackle the imbalance across the activity classes. The majority classes were assigned lesser weights compared to the minority labels based on the inverse frequency of the classes and the loss was scaled dynamically using these computed weights. The sensor-based and video-based models were trained for 1000 and 100 epochs with batch sizes of 128 and 32 respectively. The intermediate dropout layers and the final dropout layer in the proposed CNN-LSTM model were given a dropout rate of 0.4. Ridge (L2) regularization was applied to the 1D convolutional and LSTM layers in the form of the kernel and recurrent regularizers, respectively. Analogously, an L2 rate of 0.1 and a dropout value of 0.5 were utilized for the last output and dropout layers in the custom head of the proposed video model. A grid search method was deployed to identify the best combinations of L2 rate and dropout values, with L2 ranging from 0.0001 to 0.1 in steps of 0.05 and dropouts from 0.1 to 0.5 in the same increments. The best hyper-parameter values are outlined in Table 3.

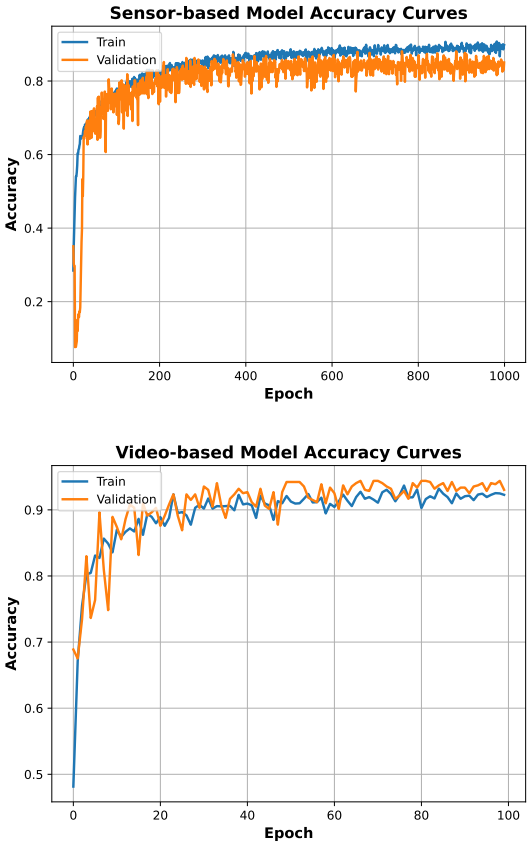


Figure 5. Training and Validation Accuracy curves for signal-based (top) and video-based models (bottom).

5 Results

The accuracies of training and validation data over the iterations were recorded and plotted as indicated in Figure 5. The video-based model achieved consistently higher accuracy, with training and validation curves stabilizing above 90%, indicating robust generalization and minimal performance fluctuations. On the other hand, the CNN-LSTM model exhibited a slower rise in accuracy, stabilizing around 85%, with higher variability in the validation accuracy curve. Overall, both the models were trained without overfitting.

After the training was completed, test sets were used to evaluate the models' performances. Based on test predictions, confusion matrices (Figure 6) were computed to assess the performance in individual classes. Due to the imbalance in the classes, overall weighted F1 scores were also determined. The Video Swin-B model demonstrates high accuracies for categories such as 'Actual Mounting' (95.72%), 'Actual Lifting' (98.1%), and 'Actual Holding' (100%). It also distinguished intentions effectively, correctly predicting all 18 instances of 'Carrying Intention' and 'Lifting Intention'. In contrast, the CNN-LSTM model shows greater variability in performance, with higher misclassifications. 'Actual Carrying' showed more confusion with 98 out of 129 instances being correct and 11 'Actual Lifting' instances being mislabeled as Lifting Intention. The model faced challenges in distinguishing intentions, as shown by the lower performances on 'Carrying Inten-

Table 3. Training and Testing: Signal-Based vs. Video-Based Approaches

Method	Model	LR	Epochs	Batch Size	Dropout	L2	Accuracy	F1 Score
Signal	CNN-LSTM	0.001	1000	128	0.4	0.05	86.82%	86.79%
Video	Video Swin-B	0.001	100	32	0.5	0.1	94.05%	94.22%

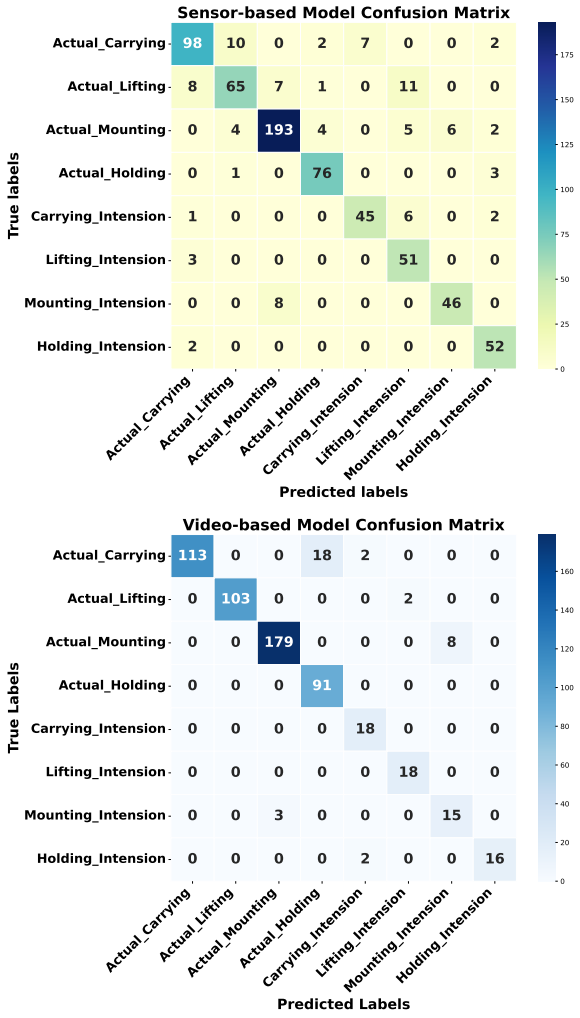


Figure 6. Confusion matrices for signal-based (top) and video-based models (bottom).

tion' and 'Mounting Intention' classes.

This demonstrates that the video-based model with 94.05% accuracy outperformed the sensor model with 86.82% accuracy, in terms of stability and classification performance, as mentioned in Table 3. Given identical hardware resources, the CNN-LSTM model took around 40 milliseconds, and the Video Swin-B model took approximately 150 milliseconds on average to carry out a prediction on one test sample.

6 Discussions

The proposed study implemented motion intention recognition systems using two kinds of modalities (sEMG and Videos). There was an inherent imbalance between the 8 classes and the "Actual" and "Intention" groups caused due to the varying durations and nature of the tasks. The two proposed models performed well in classifying human activities despite this significant imbalance. This shows the robustness and generalization of these trained models across multiple subjects. The technique of oversampling of intention classes and the application of class weights to the loss function majorly improved the classification ability of the models.

The performance scores from Table 3 clearly state that video-based data could be preferred to sensor signals. However, the signal-based model was around 3.7 times faster in making predictions on test data. This suggests that for time-sensitive applications, e.g., safety hazard monitoring, muscle signals can be a better choice. Unlike video systems, sensors provide consistent input signals even in varying environmental conditions like lighting, visibility, and obstructions. sEMG sensors are attached directly to the skin, making them highly portable compared to a stable camera setup. Hence, in use cases like material transportation, and pavement construction that involve dynamic surrounding factors and long-distance movements of workers, muscle signals can be more efficient. However, with advancements in transformers, video-based methodologies are attaining higher accuracy in action detection with relatively reduced latencies. Videos can capture the full visual context of surroundings, unlike EMG sensors which only collect motion cues from specific target muscles. Therefore, for operations that involve interactions with multiple workers, and complex behaviors, videos could be a more suitable approach.

7 Conclusion

In conclusion, the construction industry can be revolutionized by the integration of mutual collaboration between workers and robots. The proposed system conducted experiments with 3 subjects involving sEMG and video data collection followed by data processing. Finally, a hybrid CNN-LSTM model and a fine-tuned Video Swin Transformer model were implemented to handle muscle signal and video sequences respectively. This study includes comprehensive comparisons and discussions on

these modalities which conclude that both are competent in their unique ways and there cannot be a universal approach.

Future advancements can involve experiments with a larger subject pool and broader construction tasks to provide more comprehensive validation of the results obtained. This study can be expanded to build a prototype of HRC in construction sites by deploying models on robots with control interfaces. Inference optimization techniques can be integrated to reduce model latency when deployed on edge devices. Fusion model architectures that can exploit the complementary strengths of both modalities can also be considered.

References

[1] C.-J. Liang, X. Wang, V. R. Kamat, and C. C. Menassa. Human–robot collaboration in construction: Classification and research trends. *Journal of Construction Engineering and Management*, 147(10):03121006–1–03121006–23, 2021. doi:10.1061/(ASCE)CO.1943-7862.0002154.

[2] C. Brosque, E. Galbally, O. Khatib, and M. Fischer. Human-robot collaboration in construction: Opportunities and challenges. In *Proceedings of the 2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, pages 1–8, Ankara, Turkey, 2020. doi:10.1109/HORA49412.2020.9152888.

[3] J. D. Shah and K. Kim. Motion intention recognition of construction workers for human-robot collaboration in construction. In *Proceedings of the Construction Research Congress 2024*, pages 717–726, 2024. doi:10.1061/9780784485262.073.

[4] L. Zhang, G. Liu, B. Han, Z. Wang, and T. Zhang. semg based human motion intention recognition. *Journal of Robotics*, 2019:1–12, 2019. doi:10.1155/2019/3679174.

[5] J. Yang, Z. Shi, and Z. Wu. Vision-based action recognition of construction workers using dense trajectories. *Advanced Engineering Informatics*, 30: 327–336, 2016. doi:10.1016/j.aei.2016.04.009.

[6] M. Yang, C. Wu, Y. Guo, R. Jiang, F. Zhou, J. Zhang, and Z. Yang. Transformer-based deep learning model and video dataset for unsafe action identification in construction projects. *Automation in Construction*, 146:104703, 2023. doi:10.1016/j.autcon.2022.104703.

[7] S. Tang, M. Golparvar-Fard, M. Naphade, and M. M. Gopalakrishna. Video-based motion trajectory forecasting method for proactive construction safety monitoring systems. *Journal of Computing in Civil Engineering*, 34(6):04020041–1–04020041–14, 2020. doi:10.1061/(ASCE)CP.1943-5487.0000923.

[8] M. Zhang, R. Xu, H. Wu, J. Pan, and X. Luo. Human–robot collaboration for on-site construction. *Automation in Construction*, 150:104812, 2023. doi:10.1016/j.autcon.2023.104812.

[9] M.-S. Gwon, J.-H. Woo, K. K. Sahithi, and S.-H. Kim. Continuous intention prediction of lifting motions using emg-based cnn-lstm. *IEEE Access*, 12:42453–42464, 2024. doi:10.1109/ACCESS.2024.3378249.

[10] E. Kim, J. Shin, Y. Kwon, and B. Park. Emg-based dynamic hand gesture recognition using edge ai for human–robot interaction. *Electronics*, 12: 1541, 2023. doi:10.3390/electronics12071541.

[11] H. Xu and A. Xiong. Advances and disturbances in semg-based intentions and movements recognition: A review. *IEEE Sensors Journal*, 21(12):13019–13028, 2021. doi:10.1109/JSEN.2021.3068521.

[12] H. Chen and X. Yue. Swimtrans net: A multimodal robotic system for swimming action recognition driven via swin-transformer. *Frontiers in Neurorobotics*, 18:1452019, 2024. doi:10.3389/fnbot.2024.1452019.

[13] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9992–10002, 2021. doi:10.1109/ICCV48922.2021.00986.

[14] Y. Jang, I. Jeong, M. Younesi Heravi, S. Sarkar, H. Shin, and Y. Ahn. Multi-camera-based human activity recognition for human–robot collaboration in construction. *Sensors*, 23:6997, 2023. doi:10.3390/s23156997.

[15] T. Bao, S. A. R. Zaidi, S. Xie, P. Yang, and Z.-Q. Zhang. A cnn-lstm hybrid model for wrist kinematics estimation using surface electromyography. *IEEE Transactions on Instrumentation and Measurement*, 70:2503809, 2021. doi:10.1109/TIM.2020.3036654.

[16] Z. Liu, J. Ning, Y. Cao, Y. Wei, Z. Zhang, S. Lin, and H. Hu. Video swin transformer. In *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3192–3201, 2022. doi:10.1109/CVPR52688.2022.00320.

Transforming Construction Robotics Education with Virtual Reality: Analyzing Student Presence and Self-Efficacy

Amit Ojha[1], Yuming Zhang[1], and Houtan Jebelli[1]

[1]Department of Civil and Environmental Engineering, University of Illinois Urbana Champaign, United States
amito2@illinois.edu, yumingz4@illinois.edu, hjebelli@illinois.edu

Abstract

The integration of robotics into construction is transforming industry, necessitating that Construction Engineering and Management (CEM) education programs adapt to prepare future professionals for increasingly technologically driven workplaces. Traditional lecture-based methods often lack interactivity and fall short in imparting practical skills needed for emerging technologies. This study explores the potential of virtual reality (VR) as a teaching tool to enhance student presence and self-efficacy in human-robot collaboration (HRC) education within the construction sector. A VR-based immersive environment was developed, including modules for robot exploration, application tutorials, and hands-on operational training. Forty students from diverse backgrounds participated in the study, experiencing the VR environment and assessing their sense of presence and self-efficacy through validated instruments: the ITC Sense of Presence Inventory (ITC-SOPI) and the New General Self-Efficacy Scale (NGSE). Results indicated high spatial presence (M = 4.2, SD = 0.6) and ecological validity (M = 4.1, SD = 0.4), with minimal negative effects (M = 2.0, SD = 0.7), validating the environment's realism and comfort. Self-efficacy scores averaged 4.1, suggesting strong confidence among students in managing HRC tasks in construction sites. These findings validate VR as a tool that enhances student presence and self-efficacy in learning HRC in construction setting.

Keywords –

Virtual Reality; Human-Robot Collaboration; CEM; Student Presence; Self-Efficacy

1 Introduction

The landscape of technology is rapidly transforming, with robotics playing a pivotal role in revolutionizing various sectors. [1,2]. Notably, the construction industry is poised for major changes, driven by robotic technologies. Research suggests that within the next two decades, 81% of construction firms globally anticipate integrating robotics into their workflows [3]. This shift underscores the need for construction engineering and management (CEM) education programs to adapt by embedding essential technological competencies in their curricula to prepare future professionals for the evolving demands of the industry.

In conventional CEM education, passive instructional methods like lecture-based sessions are commonly used, but these approaches often fall short in effectively teaching emerging technologies [4]. Lecture-based learning typically lacks the necessary interactivity, which is crucial for students to truly understand and engage with complex technological concepts. Passive learning tends to present information in a linear and structured manner, which does not capture that nuanced, real-world challenges and ambiguities that students will face when working with cutting-edge tools and methodologies. This can lead to a superficial understanding of technological advancements, where students may grasp theoretical knowledge but lack the ability to translate that into real-world problem-solving. Emerging technologies, particularly in the construction sector, are dynamic and often require a flexible, hands-on learning approach to develop critical thinking and adaptability. In contrast, integrating virtual environments that mimic real-world construction scenarios can offer a more immersive and engaging way to teach these technologies [5–8]. In virtual environments, learners can interact with virtual tools, robotics, and other advanced systems, gaining valuable experience without the safety risks and costs associated with physical experimentation. This "learning by doing" approach allows students to explore complex concepts in a safe, controlled setting, where they can make mistakes, troubleshoot, and learn from their experiences.

By actively engaging with the virtual environment, students are better equipped to understand the intricacies of construction technologies and how they apply to real-world situations. This method not only fosters deeper learning but also encourages exploration and creativity, as students are given the freedom to experiment, problem-solve, and adapt in ways that passive learning methods do not typically allow. Ultimately, such hands-on experiences prepare students for the practical challenges they will face in the construction industry, equipping them with the skills needed to confidently handle emerging technologies.

Immersive learning environments are increasingly central to innovation in construction education, yet their success hinges significantly on two factors: student presence and self-efficacy within these virtual settings. Student presence, or the degree to which learners feel actively involved and engaged in the virtual learning process, is crucial for the effectiveness of educational technologies. This sense of presence can greatly influence engagement levels, learning outcomes, and overall satisfaction with the educational experience. Self-efficacy, defined as a students' belief in their ability to accomplish specific tasks, also plays a critical role in the dynamics of learning environments. When students perceive themselves as capable, it not only boosts their motivation but also enhances their engagement, leading to more effective learning. Environments that foster high self-efficacy help students take initiative, face challenges confidently, and persist in solving complex problems, which are essential skills in the construction industry

Towards this end, the aim of the research to assess the effectiveness of virtual reality (VR) as a teaching tool in enhancing student presence and self-efficacy within HRC education in the construction sector. As robots are getting popular in construction operations, the use of safety training module will equip the students with pragmatic safety knowledge for working alongside these robots. The research intends to explore how VR-based learning impact student presence and self-efficacy. Insights from the research will provide valuable guidance on designing and implementing VR-based educational tools, enhancing their effectiveness and usability for construction curricula and beyond.

2 Impending Adoption of Robotics Technology in Construction Sector

The construction industry is experiencing an unprecedented transformation through the integration of advanced robotic technologies. This shift represents a strategic response to several deeply entrenched challenges that have long hindered the progress of construction sector, including stagnant low productivity, an increasingly severe shortage of skilled labor, and ongoing safety concerns that drive up the operational costs [9–13]. While the inherently chaotic and unstructured nature of the construction site, combined with the constant interplay of heavy machinery and human workers poses significant barriers to implementing fully autonomous robotic systems, innovative solutions are emerging to address such complexities.

The construction industry had undergone a paradigmatic shift, moving away from the pursuit of complete automation to embrace a more nuanced approach that emphasizes HRC over complete automation. Such a strategic pivot acknowledges that human expertise- including problem-solving abilities, and adaptability remains irreplaceable in construction operations. The development of collaborative robots, or cobots, represents a breakthrough in this direction, offering systems that harmoniously combine cognitive abilities with the mechanical advantages of robotics [13–15]. These collaborative systems excel in precise execution of repetitive tasks, enhanced physical endurance for demanding operations, consistent performance in standardized procedures [16–19]. Importantly, these systems are specifically engineered to function as supportive tools rather than replacements for human workers, creating a symbiotic relationship that enhances overall project efficiency and safety.

This technological evolution brings new challenges and requirements for CEM professionals. The integration of the advanced robotic systems demands the development of new technical competencies, enhanced understanding of HRI protocols, cultivation of trust in automated systems, adaptation to technology-driven work environments, and mastery of new interface systems and control mechanisms. The construction site of the future is increasingly envisioned as an environment where human workers and robotic systems work in tandem, each contributing their unique strength to project success. The construction education landscape must evolve substantially. CEM curricula need to incorporate comprehensive training in human-robot interaction principles, ensuring that future professionals are well-prepared to navigate complex HRC environments, implement effective safety protocols for mixed human-robot teams, and optimize workflow processes that involve both human and robotic elements.

3 VR-based Learning of Human Robotic Collaboration in Construction Curriculum

Traditional CEM education has long relied on conventional teaching methods that often fall short in preparing students for the complex realities of construction environments. This gap between classroom learning and practical application has prompted educators to explore more effective pedagogical approaches, with immersive technologies emerging as a promising solution. VR technology stands out as a transformative tool in CEM education, offering students the opportunity to gain hands-on experience in simulated construction environments without exposure to real-world risks. Students can practice programming and operating collaborative robots, learn safety protocols specific to HRI, and understand the capabilities and limitations of various robotic systems. VR simulations can recreate diverse construction scenarios, from simple

repetitive tasks to complex operations requiring sophisticated human-robot coordination. These virtual environments can also simulate potential hazards and emergency situations, allowing students to develop appropriate response strategies without real-world risks. Through detailed visualizations of robot components and functionalities, VR helps students gain a deeper understanding of robot capabilities, limitations, and potential applications in construction. The simulated environment also allows students to learn how robots integrate with other construction systems and processes, providing a holistic view of modern construction environments. VR-based HRC training can present students with challenging scenarios that require quick thinking and problem solving, allowing them to practice making decisions and manage unexpected issues that may arise when working with robots on construction sites. Implementing VR-based HRC training can also be more cost-effective than traditional methods, reducing the need for physical robot prototypes and simulating a wide range of construction environments without the expense of setting up real-world training sites. Furthermore, VR-based learning modules can be designed to progressively increase in complexity, matching students' growing proficiency in HRC principles. This structured approach helps develop comprehensive understanding and practical skills essential for managing future construction projects where humans and robots work together. The integration of such technology-driven learning methods in CEM curricula ensures that graduates are well-prepared to lead and manage the increasingly automated construction sites of tomorrow.

Immersive learning environments are increasingly central to innovation in construction education, yet their success hinges significantly on two factors: student presence and self-efficacy within these virtual settings. Presence, as defined by Constructivist Learning Theory [20,21], refers to the extent to which students feel connected and engaged with their learning environment and content. This theory emphasizes that learning is most effective when students are actively involved in constructing their own understanding through meaningful interactions. A strong sense of presence helps students perceive their education as relevant and closely tied to real-world applications. When students perceive a direct connection between what they learn in the classroom and what they observe on actual construction sites, their understanding of theoretical concepts is enhanced. This connection ensures that the knowledge they acquire is not abstract but seen as directly applicable and impactful in their future careers. On the other hand, self-efficacy in CEM education is crucial for student motivation and success. Self-efficacy involves students' beliefs in their abilities to successfully perform tasks and solve problems they will encounter in their professional lives. According to Social Cognitive Theory [22], self-efficacy is developed through mastery experiences, vicarious experiences, verbal persuasion, and physiological feedback. In CEM, where projects often involve significant challenges-from logistical issues to managing diverse teams-self-efficacy determines a student's perseverance and resilience. Students who believe in their capabilities are more likely to take on challenging projects, persist through setbacks, and innovate under pressure. Students are also more inclined to pursue leadership roles and advocate for necessary changes in safety practices, sustainability measures, and technological adaptations. By leveraging Constructivist Learning Theory, educators can design VR-based modules that enhance presence through immersive, scenario-based learning. Simultaneously, applying Social Cognitive Theory principles within these modules—such as enabling practice in a safe environment—can significantly boost self-efficacy. This dual approach ensures that students are not only deeply engaged with their learning but also equipped with the confidence to apply their knowledge decisively and effectively in real-world situations.

4 VR-Based Learning of Human Robotic Collaboration and its Impacts on Student Presence and Self-Efficacy

The primary objective of this research is to assess the effectiveness of VR as a teaching tool in enhancing student presence and self-efficacy within HRC education in the construction sector. The study is divided into two phases (Figure 1): Development of an Immersive Learning Environment for Teaching Human-Robot Collaboration and Evaluation of the Developed System on Students' Sense of Presence and Self-efficacy. Initially, a virtual learning environment was created to offer comprehensive insights into the human-robot collaboration in the construction sector. Subsequently, the immersive learning environment was analyzed to determine its impact on students' sense of presence and self-efficacy. To conduct a comparative analysis between the VR-based approach and conventional methods, a cohort of 40 students from diverse educational backgrounds was recruited for this study. The participants represented a mix of undergraduate and graduate students, with varying levels of experience in the construction sector, ranging from no experience to up to two years of practical experience through internships or academic projects. However, none of the participants had prior experience with VR technologies. All the recruited students took turns in experiencing the developed learning environment, respectively. Then, a self-assessment technique was leveraged to measure the students' sense of presence and self-efficacy across the

two learning scenarios. Student sense of presence was evaluated using ITC Sense of Presence Inventory (ITC-SOPI) [23]. Likewise, self-efficacy (i.e., beliefs in an individual's capability to organize and execute the actions required to complete a particular task successfully [24]) will be evaluated by using a New General Self-Efficacy Scale (NGSE Prior to commencing the study, all participants provided informed consent and were briefed on the confidentiality of data and their rights as participants. Health-related information was also collected, with all participants reporting normal or corrected vision, and no indications of oculomotor or neurological issues.

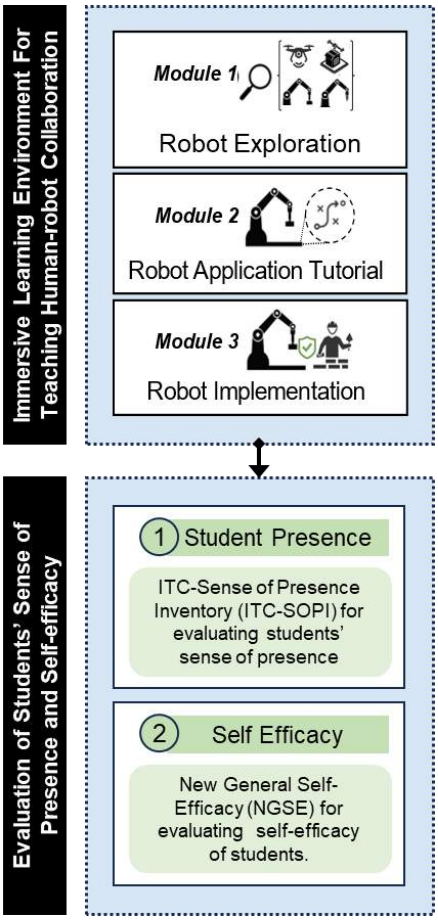


Figure 1: Proposed Steps for Investigating Immersive Learning Environment for Teaching Construction Robotics on Users' Sense of Presence and Self-efficacy

4.1 Development of Immersive Learning Environment for Teaching Human-Robot Collaboration

The developed learning environment included a detailed simulation of a real construction site, complete with educational modules that highlighted various aspects of working alongside construction robots.

For this, a construction site was meticulously designed using Autodesk Revit and then transferred into the Unity game engine for further development. Within Unity, various elements of the immersive environment were designed as Game Objects equipped with collider and Rigidbody components to simulate realistic physical interactions. The various elements within the virtual space were developed using a combination of 3D modeling software like Blender and Unity's native creation tools, allowing for precise control over the environmental details and interactive features. The developed learning environment consists of three modules, as shown in Figure 2.

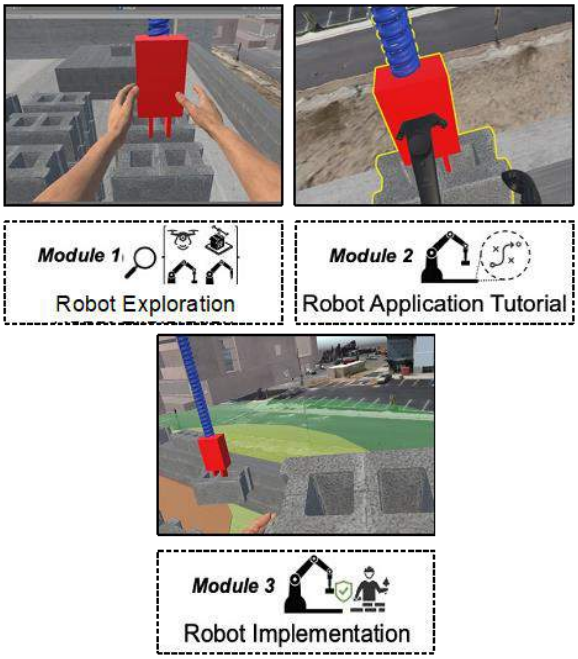


Figure 2: Modules of the Developed Immersive Learning Environment

1. ***Module 1- Robot Exploration***: This module served as an interactive introduction to various construction automation technologies. Students could navigate through a comprehensive catalog of construction robots, including advanced systems like concrete printing machines, automated bricklaying units, and aerial drones. The interface featured an intuitive menu system allowing learners to select and study different robotic systems. When examining each robot, students could interact with specific components - from robotic manipulators to control interfaces and actuator systems - with detailed technical specifications appearing in contextual text displays upon component selection. Interaction was facilitated through MANUS haptic gloves, allowing

users to intuitively grab, manipulate, and explore virtual objects. The gloves provided haptic feedback to simulate tactile sensations, enhancing realism by mimicking the feel of surfaces and resistance when interacting with robotic components.

2. ***Module 2- Robot Application Tutorial****:* This module delivered in-depth guidance on utilizing robotic systems to enhance construction safety and efficiency. A key feature of this module was the integration of a virtual instructor, represented by an animated 3D human figure capable of demonstrating proper robot deployment techniques. This digital mentor could perform realistic movements while providing verbal explanations and demonstrations of correct robotic implementation procedures, helping students understand proper placement and operational protocols based on established safety criteria and best practices. Students mirrored the instructor's actions using MANUS gloves, with haptic feedback providing physical cues to guide correct hand positioning and movement.
3. ***Module 3– Robotic Implementation:*** This module emphasized hands-on operational experience through practical task completion. This section was designed as an interactive training ground where students could develop proficiency in robot control and operation. The module provided step-by-step instruction in fundamental robot control principles, followed by practical exercises requiring students to complete specific construction-related tasks. The MANUS gloves enabled precise hand-based control of virtual joysticks, buttons, and control panels, offering haptic feedback to replicate the physical resistance and tactile response of real-world equipment. The iterative nature of these exercises allowed learners to refine their skills through repeated practice and progressive improvement.

4.2 Inferences on Student Presence and Self-Efficacy

To understand if the developed immersive setting makes the student feel present in the environment as intended: the sense of the presence of the users was evaluated via the ITC-SOPI questionnaire [25]. Further, self-efficacy (i.e., beliefs in an individual's capability to organize and execute the actions required to complete a particular task successfully [24]) will be evaluated by using a refined self-efficacy scale [26], modified by the experts from industrial collaborators for construction-specific purposes.

To assess participant's sense of presence, a validated 44 item 5-point Likert scale (ranging from 1 to 5) questionnaire originated from Lessiter et al. [23], ITC-SOPI was selected due to its cross-media comparison capabilities. The instrument is divided into four factors: Spatial Presence (19 items), Engagement (13 items), Ecological Validity/ Naturalness (5 items), and Negative Effects (6 items). Spatial presence refers to the ability to physically control and manipulate aspects of the displayed environment. Ecological Validity refers to the believability, realism of the content, and naturalness of the environment. Negative effects reflect headache, eye strain, tiredness, and other negative effects that may be associated with the virtual environment. To validate the consistency of the student presence evaluation, the author performed a reliability analysis on the ITC-SOPI using the Cronbach's alpha test. The analysis results in Cronbach's alpha of 0.83, indicating a good level of internal consistency.

Likewise, self-efficacy will be evaluated by using NGSE. NGSE assesses an individual's belief in their ability to achieve goals and overcome challenges across various situations. NGSE consists of 8 statements that respondent's rate on how well each reflects their self-belief. The statements were rated on a 5-point Likert scale, where 1 indicate "Strongly Disagree," 2 indicate "Disagree," 3 indicate "Neither Agree nor Disagree," 4 indicate "Agree," and 5 indicate "Strongly Agree." To derive a general self-efficacy score, the responses to each of the eight items are averaged, yielding a single score that ranges from 1 to 5. Higher scores indicate a stronger belief in one's ability to manage and succeed in a variety of situations. To validate the consistency of the self-efficacy evaluation, the authors performed a reliability analysis on the NGSE using Cronbach's alpha test. The analysis resulted in Cronbach's alpha of 0.86, indicating a good level of internal consistency.

5 Results and Discussion

The ITC-SOPI questionnaire results demonstrated significant levels of presence across multiple dimensions, as detailed in Table 1. Spatial presence exhibited the highest mean score (M = 4.2, SD = 0.6) on the 5-point Likert scale, indicating that participants experienced a strong sense of physical interaction within the virtual construction environment. This was followed by ecological validity (M = 4.1, SD = 0.4) and engagement (M = 3.8, SD = 0.5). Notably, negative effects demonstrated the lowest mean score (M = 2.0, SD = 0.7), suggesting minimal adverse physiological or psychological reactions to the virtual environment. The high spatial presence score (M = 4.2) is particularly significant as it surpasses previous studies in construction-based virtual environments, where mean scores typically range from 3.5 to 3.9 [citation needed]. The ecological validity results (M = 4.1) suggest strong environmental verisimilitude, crucial for transferring skills from virtual to real construction scenarios. The relatively low standard deviations across presence

metrics (ranging from 0.4 to 0.7) indicate consistent participant experiences, supporting the robustness of the implementation across different user profiles.

Table 1. Results of metrics for ITC-SOPI

Metrics	Mean	Standard Deviation
Spatial Presence	4.2	0.6
Engagement	3.8	0.5
Ecological Validity	4.1	0.4
Negative Effects	2	0.7

The analysis of the NGSE data, presented in Table 2, yielded an overall mean self-efficacy score of 4.1 (SD = 0.5), with individual statement scores ranging from 3.8 to 4.3. The detailed examination of NGSE components revealed that participants exhibited highest confidence in goal achievement and task performance capabilities. The statement *"I will be able to achieve most of the goals that I set for myself"* received one of the highest ratings (M = 4.3, SD = 0.6), along with *"I am confident that I can perform effectively on many different tasks"* (M = 4.3, SD = 0.6). These strong endorsements suggest that exposure to the virtual environment particularly enhanced participants' goal-oriented self-efficacy, a crucial finding for construction education where clear goal visualization and task completion are essential components of professional development. The data further indicated robust self-efficacy in managing construction-related challenges. Participants expressed high confidence through their responses to *"When facing difficult tasks, I am certain that I will accomplish them"* (M = 4.2, SD = 0.7) and *"I will be able to successfully overcome many challenges"* (M = 4.2, SD = 0.7). These closely aligned scores suggest that the virtual environment effectively cultivated resilience and problem-solving confidence among participants.

Table 2. Results of NGSE metrics for assessing self-efficacy of students in developed immersive environment

Statements for NGSE	Mean $\pm$ SD
I will be able to achieve most of the goals that I set for myself.	4.3 $\pm$ 0.6
When facing difficult tasks, I am certain that I will accomplish them	4.2 $\pm$ 0.7
In general, I think that I can obtain outcomes that are important to me	3.8 $\pm$ 0.8
I believe I can succeed at most any endeavor to which I set my mind.	4.1 $\pm$ 0.6
I will be able to successfully overcome many challenges.	4.2 $\pm$ 0.7
I am confident that I can perform effectively on many different tasks.	4.3 $\pm$ 0.6
Compared to other people, I can do most tasks very well.	4 $\pm$ 0.8
Even when things are tough, I can perform quite well	4.1 $\pm$ 0.7
Overall Self Efficacy Score	4.1 $\pm$ 0.5

Participants also demonstrated strong self-efficacy in comparative performance and resilience metrics. The statement *"Compared to other people, I can do most tasks very well"* yielded positive results (M = 4.0, SD = 0.8), while "Even when things are tough, I can perform quite well" showed similarly strong outcomes (M = 4.1, SD = 0.7). The statement *"I believe I can succeed at most any endeavor to which I set my mind"* also received favorable ratings (M = 4.1, SD = 0.6). These findings suggest that the virtual environment not only enhanced individual capability beliefs but also strengthened participants' confidence in their relative competence within the field. Notably, while all NGSE components received positive evaluations, the statement *"In general, I think that I can obtain outcomes that are important to me"* showed relatively lower scores (M = 3.8, SD = 0.8), though still represents a positive self-efficacy level.

The combined results from both instruments suggest that the immersive environment successfully creates an engaging, realistic learning experience while building student confidence in construction-related capabilities. The high spatial presence and ecological validity scores from ITC-SOPI align well with the strong self-efficacy results, indicating that the realistic environment contributes to students' confidence in their abilities. These findings align with previous research suggesting that immersive environments can enhance learning outcomes through increased presence and engagement []. The particularly strong scores in spatial presence and ecological validity, combined with minimal negative effects, suggest that the technical implementation of the environment effectively balances realism with user comfort.

While the results indicated that the developed immersive environment the results indicated that the developed immersive environment effectively supports a sense of presence and bolsters self-efficacy among students, there are limitations to consider that could be addressed in future studies. First, the study was conducted with a relatively small sample size of 40 students, all from a specific academic background, potentially limiting the generalizability of the findings. A larger, more diverse sample could yield insights that are

more broadly applicable across different educational and professional settings. The study also relied on short-duration sessions in the virtual environment, which might not reveal the full spectrum of user engagement and potential negative effects such as fatigue or cognitive overload. Prolonged exposure assessments could provide a more comprehensive understanding of how extended use impacts both presence and self-efficacy. Furthermore, while the simulation focuses on human-robot collaboration, the lack of physical robot interactions limits students' opportunities to develop hands-on skills that are essential in real-world settings. Incorporating augmented or mixed reality elements that allow physical interaction with robots could improve skill transferability.

6 Conclusion

This study evaluated the effectiveness of VR as a teaching tool in enhancing student presence and self-efficacy within HRC education in the construction sector. The findings demonstrate that the virtual environment provides a robust sense of presence, particularly in spatial presence and ecological validity, which are essential for realistic and engaging learning experiences. Additionally, the high self-efficacy scores indicate that the environment supports students' confidence in their ability to achieve goals and tackle challenges related to construction tasks, which are key attributes for success in construction engineering and management. The VR-based modules, including robot exploration, application tutorials, and hands-on operational exercises, appear to effectively prepare students for real-world HRC applications by promoting engagement and fostering practical skills in a safe, simulated setting. The low negative effects observed further underscore the usability and comfort of the environment, suggesting that students can engage in immersive learning for extended periods without adverse physiological or psychological impacts.

However, the study has some limitations. The sample was drawn primarily from CEM students, which may not fully represent the broader range of students or professionals who may benefit from VR-based learning. Additionally, the exposure to the virtual environment was limited in duration, and future studies could explore the effects of extended exposure to VR learning environments to assess the long-term retention of skills and the sustainability of enhanced self-efficacy. Furthermore, the study did not include physical robot interaction, which could enhance the realism of certain tasks. Future research could integrate augmented or mixed reality (AR/MR) elements or provide opportunities for physical robot interaction to improve the hands-on nature of the training.

Future directions for research could also include expanding the sample to include participants from various academic disciplines and professional backgrounds to better understand the broader applicability of VR training. Investigating the long-term impact of VR training on skill retention and career readiness would be valuable, as would the integration of AR/MR technologies offer more immersive and physically interactive learning experiences.

Overall, this research provides valuable insights into the design and implementation of VR-based tools in construction education. By integrating such immersive learning environments into CEM curricula, educational institutions can better equip future professionals with the skills and confidence needed to navigate the increasingly automated construction industry. This study underscores the potential of VR-based training to bridge the gap between theoretical learning and real-world application, paving the way for more comprehensive and effective construction education.

Acknowledgements

This material is based upon work supported by the National Science Foundation (under Grant No. DUE-2402008, “Investigating the Impact of an Immersive VR-based Learning Environment for Learning Human-Robot Collaboration in Construction Robotics Education”). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] Sweet R. The contractor who invented a construction robot. Construction Research and Innovation 2018;9:9–12. https://doi.org/10.1080/20450249.2018.1442702.

[2] Zhao X, Jebelli H. A computational method for real-time roof defect segmentation in robotic inspection. Computer-Aided Civil and Infrastructure Engineering 2025. https://doi.org/10.1111/MICE.13471.

[3] ABB Construction Industry Survey. 2021.

[4] Scott LM. Theory and research in construction education: the case for pragmatism. Construction Management and Economics 2016;34:552–60. https://doi.org/10.1080/01446193.2016.1151539.

[5] Le QT, Pedro A, Park CS. A Social Virtual Reality Based Construction Safety Education System for Experiential Learning. Journal of Intelligent and Robotic Systems: Theory and Applications 2015;79:487–506. https://doi.org/10.1007/s10846-014-0112-z.

[6] Shayesteh S, Ojha A, Liu Y, Jebelli H. Human-robot teaming in construction: Evaluative safety

training through the integration of immersive technologies and wearable physiological sensing. Saf Sci 2023;159:106019. https://doi.org/10.1016/J.SSCI.2022.106019.

[7] Ojha A, Shayesteh S, Jebelli H, Akanmu A, Miller S, Sharma P. Revitalizing Students' Perception and Knowledge of Construction Safety: Leveraging Virtual Reality as an Experiential Learning Tool for Construction Robotic Safety Education. Construction Research Congress 2024, CRC 2024 2024;4:129–38. https://doi.org/10.1061/9780784485293.014.

[8] Ojha A, Zhang Y, Shayesteh S, Sharifironizi A, Jebelli H, Akanmu A. Advancing Construction Robotics Learning Through Virtual Reality: A User Experience and Usability Study. Computing in Civil Engineering , Pittsburgh: 2025.

[9] Liu Y, Habibnezhad M, Jebelli H. Brain-computer interface for hands-free teleoperation of construction robots. Autom Constr 2021;123:103523.

[10] Liu Y, Habibnezhad M, Jebelli H. Brainwave-driven human-robot collaboration in construction. Autom Constr 2021;124:103556. https://doi.org/10.1016/j.autcon.2021.103556.

[11] Haas C, Skibniewski M, Budny E. Robotics in Civil Engineering. Computer-Aided Civil and Infrastructure Engineering 1995;10:371–81. https://doi.org/10.1111/j.1467-8667.1995.tb00298.x.

[12] Han C. "Human-Robot Cooperation Technology" An Ideal Midway Solution Heading Toward the Future of Robotics and Automation in Construction. 2011 Proceedings of the 28th ISARC 2011:13–8. https://doi.org/10.22260/ISARC2011/0297.

[13] Bock T, Linner T. Single-Task Construction Robots by Category. Construction Robots, Cambridge: Cambridge University Press; 2016, p. 14–290. https://doi.org/10.1017/CBO9781139872041.002.

[14] Petters S, Belden R. SAM, the robotic bricklayer. SMART/Dynamics of Masonry 2014;1:10–4.

[15] Afsari K, Gupta S, Afkhamiaghda M, Lu Z. Applications of Collaborative Industrial Robots in Building Construction. 54th ASC Annual International Conference , 2018, p. 472–9.

[16] Thomaz A, Hoffman G, Cakmak M. Computational Human-Robot Interaction. Foundations and Trends in Robotics 2016;4:104–223. https://doi.org/10.1561/2300000049.

[17] BAUER A, WOLLHERR D, BUSS M. Human-Robot Collaboration: a Survey. International Journal of Humanoid Robotics 2008;05:47–66. https://doi.org/10.1142/S0219843608001303.

[18] Ajoudani A, Zanchettin AM, Ivaldi S, Albu-Schäffer A, Kosuge K, Khatib O. Progress and prospects of the human–robot collaboration. Auton Robots 2018;42:957–75. https://doi.org/10.1007/s10514-017-9677-2.

[19] Villani V, Pini F, Leali F, Secchi C. Survey on human–robot collaboration in industrial settings: Safety, intuitive interfaces and applications. Mechatronics 2018;55:248–66. https://doi.org/10.1016/j.mechatronics.2018.02.009.

[20] VYGOTSKY LS. Mind in Society. Harvard University Press; 2019. https://doi.org/10.2307/J.CTVJF9VZ4.

[21] Duffy TM., Jonassen DH., Cole Peggy. Theory into Practice : How do we Link? Constructivism and the Technology of Instruction, Routledge; 2013, p. 17–34. https://doi.org/10.4324/9780203461976-4.

[22] Bandura A. Social Foundations of Thought and Action. The Health Psychology Reader 1986:94–106. https://doi.org/10.4135/9781446221129.N6.

[23] Lessiter J, Freeman J, Keogh E, Davidoff J. A Cross-Media Presence Questionnaire: The ITC-Sense of Presence Inventory. Presence: Teleoperators and Virtual Environments 2001;10:282–97. https://doi.org/10.1162/105474601300343612.

[24] Bandura A. Self-efficacy: Toward a unifying theory of behavioral change. Psychol Rev 1977;84:191–215. https://doi.org/10.1037/0033-295X.84.2.191.

[25] Witmer BG, Singer MJ. Measuring Presence in Virtual Environments: A Presence Questionnaire. Presence: Teleoperators and Virtual Environments 1998;7:225–40. https://doi.org/10.1162/105474698565686.

[26] Chen G, Gully SM, Eden D. Validation of a New General Self-Efficacy Scale. Organ Res Methods 2001;4:62–83. https://doi.org/10.1177/109442810141004.

A Framework for Real-Time Estimation of Human Activity Intensity in Indoor Environments

Moein Younesi Heravi[1], Youjin Jang[2], Inbae Jeong[3], and Israt Sharmin Dola[4]

[1,2] Dept. of Civil, Construction and Environmental Engineering, North Dakota State Univ., Fargo, USA
[3,4] Dept. of Mechanical Engineering, North Dakota State Univ., Fargo, USA
moein.younesiheravi@ndsu.edu, y.jang@ndsu.edu, inbae.jeong@ndsu.edu, israt.dola@ndsu.edu

Abstract

The real-time estimation of human activity level in indoor environments is crucial for thermal comfort prediction and optimizing HVAC systems, as well as supporting health monitoring and ergonomic assessments. Existing methods typically categorize activities into predefined types and therefore are not efficient for new activities. Furthermore, the transition between activities is not captured in these classification methods. To address these limitations, we propose a novel Activity Intensity Score (AIS) framework that provides a continuous, non-intrusive assessment of human activity intensity. The proposed AIS framework uses kinematic parameters derived from video-based pose estimation to calculate a dynamic intensity score. Key kinematic parameters, such as angular speed, angular acceleration, range of motion, movement frequency, and rotational energy, are extracted from pose landmarks using MediaPipe. These parameters are then normalized and combined to generate a continuous AIS that reflects real-time variations in movement intensity. The proposed approach was validated through controlled experiments with participants performing activities of varying intensities, including low (e.g., sitting), moderate (e.g., walking), and high (e.g., jumping jacks). Results showed that the AIS effectively distinguishes between different activity intensities, with higher AIS values corresponding to more intensive activities. The proposed method addresses the need for a more granular understanding of activity levels beyond simple categorical classification. This research contributes to real-time activity estimation, which has applications in optimizing indoor environmental conditions, health monitoring, and dynamic ergonomic assessments.

Keywords – Activity Intensity Score; Real-Time Estimation; Kinematic Chains; Computer Vision

1 Introduction

In indoor environments, the real-time estimation of human activity intensity plays a crucial role in optimizing environmental control systems, such as HVAC, as well as supporting health monitoring and ergonomic assessments [1], [2], [3], [4], [5]. Traditional HVAC systems operate with predefined control settings that do not respond to changes in occupant behavior or environmental conditions [6]. Dynamic HVAC systems that adapt to real-time changes in occupant activities can significantly enhance both comfort and energy efficiency. Therefore, such systems are essential for addressing the dynamic requirements of occupants, since they provide intelligent, responsive adjustments that improve overall indoor conditions. With recent progress in AI-driven models, the autonomous HVAC control systems have been developed to adjust the settings by considering energy efficiency and thermal comfort of occupants. However, typically in these models the role of human activity in thermal comfort is simplified by assuming comfort as a function of environmental parameters such as temperature [7], [8], [9], [10], which often leads to either excessive energy consumption or inadequate thermal comfort for the occupants.

The activity intensity is a key determinant of a building occupant's thermal comfort perception, as it directly influences the body's energy expenditure, similar to how metabolic rate (MET) is used in thermal comfort models like PMV [11]. This energy expenditure impacts the thermal conditions required for maintaining comfort, making activity intensity an essential parameter for dynamic HVAC adjustments. Since dynamic HVAC systems require real-time adjustments to occupant behavior, understanding activity levels is crucial for enabling these systems to respond effectively to varying thermal demands [12], [13]. For example, an individual engaging in physical activity generates more body heat compared to someone resting, requiring different temperature regulation to maintain comfort [14]. By integrating activity intensity information, environmental

control systems can achieve a better balance between energy consumption and occupant comfort. Additionally, human activity level estimation can support health monitoring, as sudden changes in activity patterns may indicate health issues that need attention [3], [14].

Although understanding occupant activity intensity is essential for dynamic HVAC adjustments, most existing approaches focus on activity recognition, which involves categorizing activities into distinct types such as walking, sitting, or running [15], [16]. These activity recognition methods face limitations in addressing the needs of adaptive environmental control, as they typically focus on discrete classifications rather than providing a continuous measure of activity intensity [1]. This not only limits their functionality for certain defined activities, but also fails to provide the nuanced understanding required for optimizing HVAC systems. For HVAC control, a more granular understanding of activity intensity is beneficial, as it allows for finer adjustments to the environment in real time. Furthermore, many existing approaches rely on wearable sensors, which present practical limitations due to being intrusive and uncomfortable for users [12], [17], [18], [19]. As a result, there is a need for a non-intrusive, real-time solution that provides a continuous measure of activity intensity, overcoming the limitations of predefined activity categories, and effectively linking activity levels to their impact on energy expenditure and thermal comfort.

To address these challenges, this research proposes a novel Activity Intensity Score (AIS) framework that uses movement data to provide a continuous measure of activity intensity. The AIS framework is designed to overcome the limitations of existing activity recognition methods by providing a detailed and dynamic assessment of activity intensity rather than discrete activity classes. This approach integrates various kinematic parameters, such as angular speed, angular acceleration, range of motion (ROM), and cumulative effort index (CEI), to generate a continuous intensity score that reflects real-time movement variations across different parts of the body. By focusing on these parameters, the AIS framework provides a more detailed and dynamic assessment of activity intensity, which is crucial for optimizing HVAC systems and supporting health monitoring while addressing the need for accuracy and adaptability. Importantly, AIS can be calculated using any image data, including thermal imaging or visible-light cameras, allowing flexibility in its application. The key contributions of this work include the development of a novel AIS framework that estimates human activity intensity based on thermal imaging data. The proposed approach has been validated through real-world experiments, demonstrating its capability to assign a proper score for differentiating various activity levels.

2 Methodology

The proposed activity intensity estimation framework, as depicted in Figure 1, aims to continuously assess the activity intensity of building occupants by analyzing movement data derived from image frames. The framework consists of five phases that together provide a comprehensive approach for estimating activity intensity. First, human body landmarks are detected using pose estimation to create a time series representation of body movements. Next, specific body segments, known as kinematic chains, are selected, and joint angles are computed to analyze movement patterns. Then, key movement parameters such as angular speed, acceleration, and range of motion are extracted and normalized. These normalized parameters are combined to compute the Activity Intensity Score (AIS) for each kinematic chain. Finally, AIS values from all kinematic chains are aggregated to produce an overall measure of activity intensity. Each phase is explained in detail in the following sections.

2.1 Pose Estimation and Landmark Detection

The first phase focuses on extracting accurate human body landmarks from video frames to enable kinematic analysis. This phase uses the MediaPipe Pose, an advanced pose estimation library developed by Google [20]. MediaPipe Pose is a machine learning-based framework that detects and tracks human body landmarks from video frames, providing real-time and high-confidence estimations. MediaPipe utilizes a combination of deep learning and computer vision techniques to effectively identify 33 key human body landmarks, including major joints such as shoulders, elbows, wrists, hips, knees, and ankles.

The pose estimation algorithm processes each video frame and extracts the 3D coordinates (x, y, z) of these relevant body joints. These coordinates are then used to construct a time series representation of the subject's body landmark positions, capturing the dynamic movements over time. The outcome of this phase is a comprehensive dataset of body landmark positions, which provides the foundation for subsequent kinematic analyses, such as joint angle computation and movement parameter extraction.

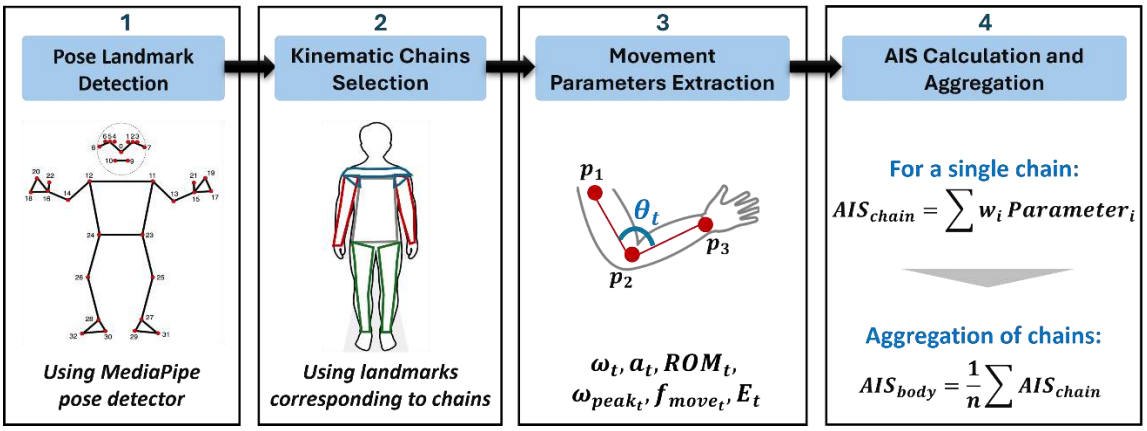


Figure 1. Proposed activity intensity estimation framework

2.2 Kinematic Chain Selection and Joint Angle Computation

In this phase, specific body segments, or kinematic chains, are defined, and joint angles are computed to analyze movement patterns. Kinematic chains represent interconnected body segments that work together to perform complex movements [21], [22]. In this study, the kinematic chains considered include the arms, legs, spinal segment, and other significant body segments, such as the neck and shoulder girdle. These chains are selected because they effectively capture the primary movements and joint dynamics of the human body, which are most relevant for assessing activity intensity in indoor environments. Table 1 provides an overview of the kinematic chains used in this study and the corresponding landmarks selected for each chain.

Each kinematic chain is defined by selecting three key landmarks that form two segments meeting at a joint. For example, the arm kinematic chain is defined by the shoulder, elbow, and wrist (see Figure 2), where the angle at the elbow joint is computed to represent arm movement. Joint angles are calculated using vector mathematics, with vectors formed from the selected landmarks. The joint angle (θ) is computed using Equation (1).

$$\theta = arccos\left(\frac{(p_1 - p_2).(p_3 - p_2)}{|p_1 - p_2| \times |p_3 - p_2|}\right) \quad (1)$$

Where p_1 , p_2 and p_3 are the joints in the corresponding chain. This calculation yields the angle at the intersection of these segments, which provides insight into the movement pattern of the body.

Table 1. Kinematic chains and corresponding landmarks used in the study

Kinematic chain	Landmarks selected
Arm (Left)	Left Shoulder, Left Elbow, Left Wrist
Arm (Right)	Right Shoulder, Right Elbow, Right Wrist
Hand-Wrist (Left)	Left Elbow, Left Wrist, Left Index
Hand-Wrist (Right)	Right Elbow, Right Wrist, Right Index
Neck-Head	Shoulder Center, Neck, Head
Shoulder Girdle	Left Shoulder, Spine, Right Shoulder
Torso	Shoulder Line, Hip Line
Spinal	Hip Center, Shoulder Center, Neck
Leg (Left)	Left Hip, Left Knee, Left Ankle
Leg (Right)	Right Hip, Right Knee, Right Ankle
Lower Leg (Left)	Left Knee, Left Ankle, Left Foot Index
Lower Leg (Right)	Right Knee, Right Ankle, Right Foot Index
Full Body	Head, Spine, Feet

The rationale behind using kinematic chains is that they allow for a more detailed understanding of human movement by focusing on specific body segments. By analyzing the joint angles of these chains, we can effectively capture the intensity and dynamics of movement, which is crucial for applications like HVAC control that depend on accurate assessments of occupant activity levels. This phase yields time series of joint angles for each selected kinematic chain, which will be used in the subsequent phases to extract movement parameters and compute the AIS.

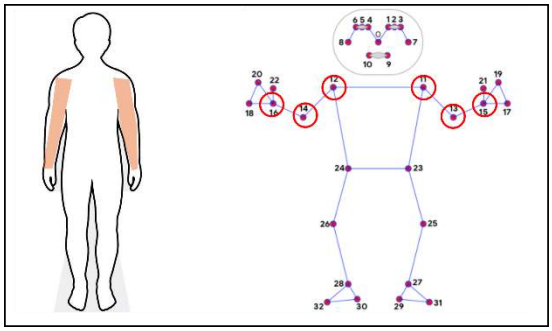


Figure 2. Illustration of sample kinematic chains and the selected landmarks from MediaPipe

2.3 Movement Parameter Extraction and Normalization

The third phase involves the extraction of key movement parameters from the joint angle data and their normalization for AIS computation. The goal is to derive meaningful metrics that reflect the movement intensity of different body segments. The following key parameters are derived, each representing a particular aspect of the activity intensity.

2.3.1 Angular Speed

Angular speed is computed as the first derivative of joint angles with respect to time, representing how quickly a joint angle changes. Since we have discrete angle measurements at each frame, numerical differentiation is used (see Equation (2)).

$$\omega_t = \left|\frac{\theta_t - \theta_{t-1}}{\Delta t}\right| \quad (2)$$

Angular speed captures the velocity of rotational movement, providing insight into the dynamic nature of the activity. High angular speed indicates rapid movement, contributing significantly to the overall activity intensity.

2.3.2 Angular Acceleration

Angular acceleration is the rate of change of angular speed, computed as the second derivative of joint angles with respect to time. Again, since we have discrete data, numerical differentiation is used (see Equation (3)).

$$\alpha_t = \left|\frac{\omega_t - \omega_{t-1}}{\Delta t}\right| \quad (3)$$

Angular acceleration measures sudden changes in movement speed, highlighting periods of rapid acceleration or deceleration. This parameter helps to identify bursts of activity that contribute to increased intensity.

2.3.3 Range of Motion

Range of motion (ROM) is defined as the difference between the maximum and minimum joint angles within a sliding window, as defined in Equation (4).

$$ROM_t = max(\theta_{t-\Delta t:t}) - min(\theta_{t-\Delta t:t}) \quad (4)$$

ROM represents the extent of movement for a specific joint. It indicates how much a joint moves during an activity. A higher ROM suggests more extensive movement, which is directly linked to higher activity intensity.

2.3.4 Peak Angular Speed

Peak angular speed is the maximum angular speed within a sliding window, presented in Equation (5).

$$\omega_{peak_t} = max(\omega_{t-\Delta t:t}) \quad (5)$$

Peak angular speed helps capture rapid bursts of movement that may not be reflected in the average speed. It is an important parameter for understanding the highest intensity levels within a given time frame.

2.3.5 Movement Frequency

Movement frequency is defined as the number of distinct peaks in the joint angle data within a sliding window. Equation (6) shows the movement frequency definition, calculated using peak detection algorithms.

$$f_{move_t} = Number\ of\ peaks\ in\ \theta_t\ within\ [t - \Delta t: t] \quad (6)$$

Higher movement frequency indicates more frequent changes in joint angles, which is often associated with increased physical activity levels.

2.3.6 Rotational Energy

Rotational energy is computed as the integral of the squared angular speed over a given time period. Since the data is discrete, it is calculated as the sum of the squared angular speeds, multiplied by the time interval, as presented in Equation (7).

$$E_t = \sum_{i=t-\Delta t}^{t} (\omega_i^2)\Delta t \quad (7)$$

Rotational energy provides a measure of the sustained effort involved in movement. It accounts for both the speed and duration of movement, making it a valuable indicator of overall activity intensity.

Each of these parameters is normalized to a scale between 0 and 1 using predefined expected ranges based on prior data. This normalization makes it possible for all the parameters to have equal contribution to the AIS irrespective of the range of their scales.

2.4 AIS Computation

In this phase, the AIS is computed by combining the

normalized movement parameters. Each parameter contributes differently to the overall activity intensity, and as such, weights are assigned to each parameter based on their relative importance. These weights can be determined through expert consultation, sensitivity analysis, or empirical studies. The AIS for each kinematic chain and at each time frame is computed using the weighted sum of the normalized parameters, as defined in Equation (8), to provide a quantitative measure of activity intensity that reflects the combined effects of speed, acceleration, ROM, movement frequency, and rotational energy. $w_1, \dots, w_6$ are weighting factors.

$$AIS(t)_{chain_i} = w_1(\omega_t)_i + w_2(\alpha_t)_i + w_3(ROM_t)_i + w_4(\omega_{peak_t})_i + w_5(f_{move_t})_i + w_6(E_t)_i \quad (8)$$

2.5 AIS Aggregation Across Kinematic Chains

The final phase integrates the AIS values from multiple kinematic chains to obtain an overall activity intensity measure. As defined in Equation (9), AIS values computed for all selected kinematic chains are aggregated for each time window, in this research using an average to compute the overall AIS. However, alternatively, a weighted average can be used if certain chains are deemed more influential. If certain kinematic chains are not detected, missed, or unavailable in the frame, their AIS values are excluded from the aggregation and average calculation. Alternatively, a weighted average can be used if certain chains are deemed more influential. The outcome of this phase is a unified AIS that represents the overall activity intensity at frame time, which can be used for thermal comfort prediction, dynamic HVAC control or other real-time applications. Here, N is the number of kinematic chains.

$$AIS(t)_{body} = \frac{1}{N}\sum_{i=1}^{N} AIS(t)_{chain_i} \quad (9)$$

3 Implementation

To test and validate the proposed AIS framework, a controlled data collection experiment was conducted to capture various movement intensities using a single standard camera in an indoor setting. The data collection was carried out at the North Dakota State University (NDSU) campus. A phone camera was positioned at waist height to ensure a full-body view, capturing video at 30 frames per second (fps). Three adults (two males, one female) were participated for the experiment.

To capture a range of movement intensities, participants were instructed to perform a set of predefined activities at low, moderate, and high intensities [14]. In this study, low-intensity included sitting, standing still, typing, and light stretching. Activities in moderate-intensity were walking at a regular pace, repeated arm movements, desk organizing, and dusting. Finally, heavy exercises like jogging in place, jumping jacks, push-ups, and squats comprised high-intensity activities [14], [23]. Activities were performed for approximately 80-90 seconds with a 1-minute rest interval in between to avoid fatigue. To enhance generalizability, participants performed different activities within each intensity level; for example, in the moderate-intensity category, one participant performed dusting, while another walked at a regular pace.

The selected activities are frequently observed in indoor settings, from workplaces to healthcare facilities, where real-time movement intensity estimation would be most applicable. They are common movements that reflect typical energy demands in such environments. For instance, a casual indoor pace (walking) contrasts well with intense exercises (jumping jacks). Furthermore, selected activities engage different body parts and movement dynamics. For instance, Sitting and Standing engage primarily the lower body and torso, while Jumping Jacks and Squats involve multi-joint movements. This diversity is beneficial to validate that AIS is responsive not just to movement intensity but to various kinematic patterns.

The AIS is calculated for each second (30 frames) of corresponding footage. For ROM, peak angular speed, movement frequency, and rotational energy, the time window is set on 30 timesteps (one second), while the instantaneous parameters (i.e., angular speed, angular acceleration) are calculated on the last frame of the past 30 frames. Weights in the AIS formula were chosen based on logical considerations of the role each parameter plays in characterizing movement intensity. Angular speed and acceleration are considered core parameters that directly reflect movement dynamics; hence they were given weights of 20 each. The remaining parameters—ROM, peak speed, movement frequency, and rotational energy—were each assigned a weight of 15, resulting in a total weight of hundred. Sample frames of the method implementation are shown in Figure 3.

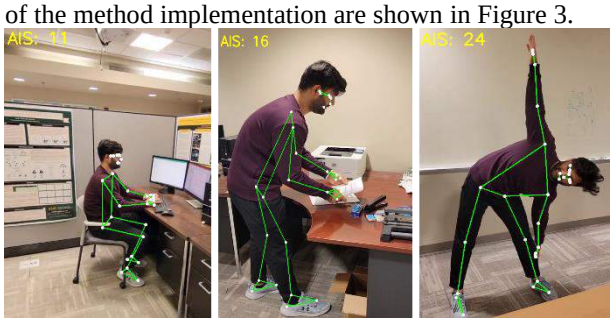


Figure 3. Implementation of the proposed approach on collected data (screenshots taken from videos). From left to right: light, moderate, and high intensity

4 Results

This section presents the findings of applying the proposed AIS estimation method on the collected data. Figure 4 illustrates the time-series plots of AIS values for a range of activities. Each plot shows the AIS variability across activities, highlighting clear differences in AIS between low-intensity, moderate-intensity, and high-intensity activities, for both participants.

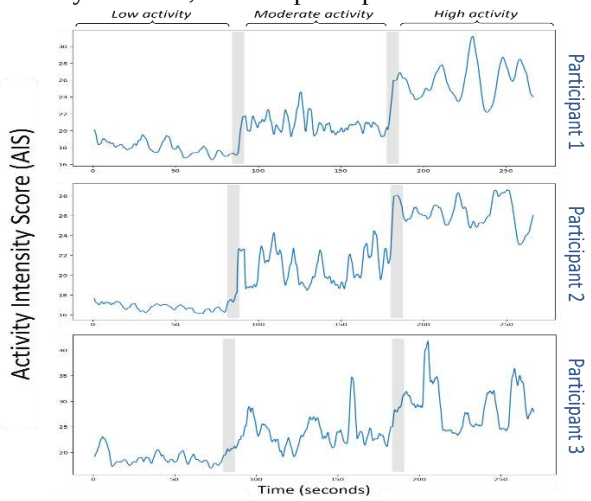


Figure 4. AIS values for different activities over time

These time-series graphs indicate that AIS values increase proportionally with activity intensity, which aligns with the purpose of the framework in capturing the expected differences in movement dynamics. To better confirm that, Figure 5 illustrates the average AIS across three participants for low, moderate, and high activity levels. The AIS increases as the activity level intensifies, confirming the method's ability to distinguish varying movement intensities. Small variations among participants reflect individual differences in movement styles (e.g., limb speed or range of motion). However, each participant shows a clear progression from low to high AIS values in accordance with activity intensity. These findings validate that the proposed AIS framework captures meaningful distinctions in movement intensity.

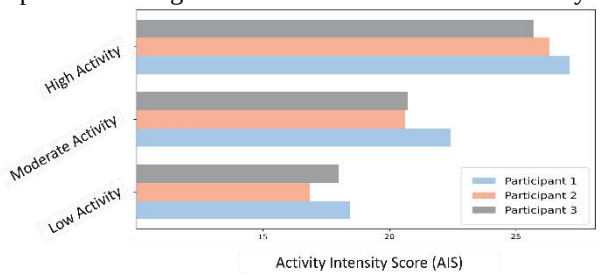


Figure 5. Mean AIS values for different activity levels

In addition, Figure 6 presents boxplots of the AIS constituent movement parameters across the three defined activity levels (for participant 1 as sample). These boxplots indicate a visual summary of the distribution of each parameter within each activity level and could illustrate the differences in movement characteristics associated with each activity level. The clear separation and trends observed among the boxplots suggest that both the AIS and individual movement parameters increase with higher levels of physical activity. This finding supports the validity of the AIS and its components as reliable indicators of activity intensity.

Finally, it should be noted that in some frames, not all kinematic chain landmarks could be detected by the pose estimation algorithm (see Figure 7). This issue primarily occurs when the subject is not fully visible from the camera angle. To address this anticipated limitation, the AIS calculation only includes kinematic chains that are fully detected in the corresponding frame.

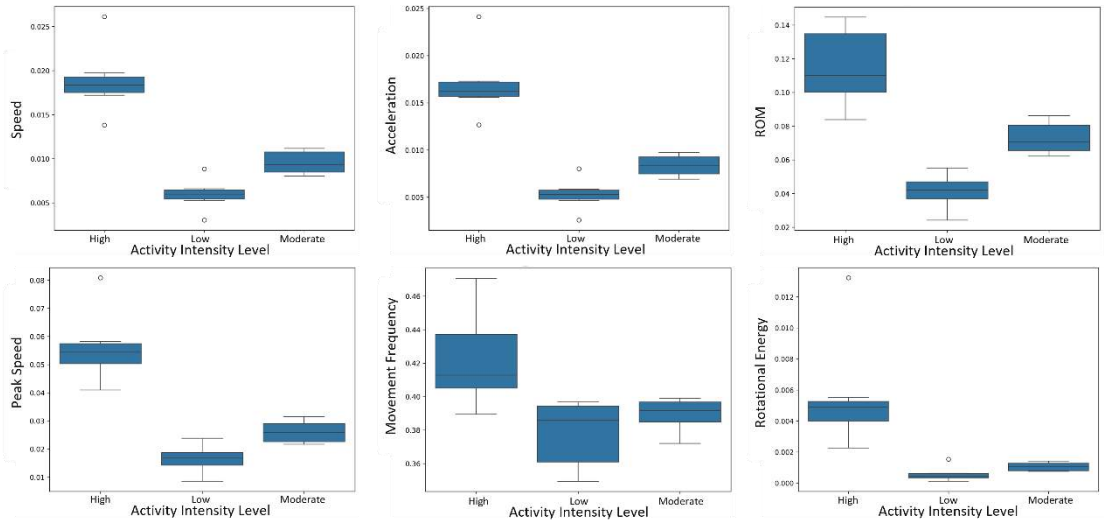


Figure 6. Distribution of AIS parameters within each activity level (normalized values)

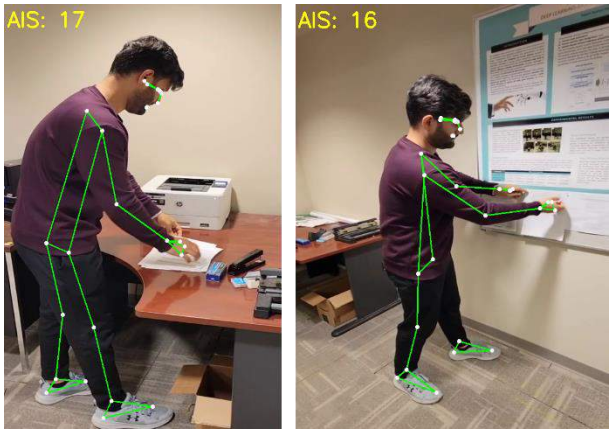


Figure 7. Missed pose landmarks (Left: left hand not detected; Right: left leg not detected)

5 Discussion

The results indicate that the proposed AIS framework effectively captures variations in human activity intensity and distinguishes between low, moderate, and high activity levels across different participants. The estimated AIS provides a real-time measure of occupant movement, which is crucial for optimizing HVAC systems by dynamically adjusting thermal conditions based on occupant activity levels. In the same conditions, higher activity levels generate more body heat, reducing the need for heating, while lower activity levels may require increased heating to maintain comfort. Therefore, AIS could be considered as an input factor when developing a thermal comfort prediction model to be utilized in HVAC control. On the other hand, integrating AIS alone may not be sufficient for HVAC optimization. Other factors, such as the duration of activity (to distinguish brief movements from sustained exertion), occupants' body temperature, and occupants' clothing insulation level should be considered. Future work could incorporate these factors alongside AIS to develop a more adaptive and efficient HVAC control strategy.

Although this study provided significant findings, some challenges remain to be addressed in future research. One limitation is the reliance on a single camera, which can lead to occlusion issues and missing pose landmarks [24], [25]. Additionally, the current framework uses a standard camera, which might raise privacy concerns for some users. Using thermal cameras is a potential solution to these privacy issues. Furthermore, future work could focus on optimizing the weights of AIS parameters.

6 Conclusion

In this study, a novel Activity Intensity Score (AIS) framework was developed to provide a continuous, real-time estimation of human movement intensity in indoor environments. The AIS integrates various kinematic parameters, including angular speed, acceleration, range of motion, movement frequency, and rotational energy, into a single measure that reflects the intensity of movement. Experimental validation confirmed the effectiveness of AIS in differentiating movement intensities, demonstrating its potential for applications such as adaptive HVAC control, ergonomic assessment, and health monitoring. While the framework offers a non-intrusive and scalable approach, future work should focus on improving robustness against occlusion, optimizing AIS parameter weights, and integrating additional physiological signals to further enhance its accuracy and practical implementation.

Acknowledgement

The authors express their gratitude to the funding provided to support this study from National Science Foundation (NSF) EPSCoR RII Track 2 Program under the grant number OIA-2119691. The findings and opinions expressed in this article are those of the authors only and do not necessarily reflect the views of the sponsors.

References

[1] S. Ahmadi-Karvigh, A. Ghahramani, B. Becerik-Gerber, and L. Soibelman, "Real-time activity recognition for energy efficiency in buildings," *Appl Energy*, vol. 211, pp. 146–160, Feb. 2018, doi: 10.1016/J.APENERGY.2017.11.055.

[2] V. A. Arowoiya, A. O. Onososen, R. C. Moehler, and Y. Fang, "Influence of Thermal Comfort on Energy Consumption for Building Occupants: The Current State of the Art," *Buildings, Vol. 14, Page 1310*, vol. 14, no. 5, p. 1310, May 2024.

[3] R. T. Larsen *et al.*, "Effectiveness of physical activity monitors in adults: systematic review and meta-analysis," *BMJ*, vol. 376, Jan. 2022, doi: 10.1136/BMJ-2021-068047.

[4] A. S. M. Moohialdin, F. Lamari, M. Miska, and B. Trigunarsyah, "Proximity Activity Intensity Identification System in Hot and Humid Weather Conditions: Development and Implementation," *J Constr Eng Manag*, vol. 149, no. 12, p. 04023122, Dec. 2023.

[5] C. Uwaoma, "Estimation of Physical Activity Level and Ambient Condition Thresholds for Respiratory Health using Smartphone Sensors," Dec. 2021, Accessed: Nov. 20, 2024. [Online]. Available: https://arxiv.org/abs/2112.09068v1

[6] J. Kim, "LSTM-based Space Occupancy Prediction towards Efficient Building Energy

Management," Dec. 2020, Accessed: Nov. 20, 2024. [Online]. Available: https://arxiv.org/abs/2012.08114v2

[7] Y. Peng *et al.*, "Energy Consumption Optimization for Heating, Ventilation and Air Conditioning Systems Based on Deep Reinforcement Learning," *IEEE Access*, vol. 11, pp. 88265–88277, 2023.

[8] A. Yayla *et al.*, "Artificial Intelligence (AI)-Based Occupant-Centric Heating Ventilation and Air Conditioning (HVAC) Control System for Multi-Zone Commercial Buildings," *Sustainability 2022, Vol. 14, Page 16107*, vol. 14, no. 23, p. 16107, Dec. 2022, doi: 10.3390/SU142316107.

[9] C. Zhong, J. Sun, J. Xie, S. Grijalva, and A. P. S. Meliopoulos, "Real-time human activity-based energy management system using model predictive control," *2018 IEEE International Conference on Consumer Electronics, ICCE 2018*, vol. 2018-January, pp. 1–6, Mar. 2018, doi: 10.1109/ICCE.2018.8326070.

[10] R. Yang and L. Wang, "Development of multi-agent system for building energy and comfort management based on occupant behaviors," *Energy Build*, vol. 56, pp. 1–7, Jan. 2013, doi: 10.1016/J.ENBUILD.2012.10.025.

[11] W. Jung and F. Jazizadeh, "Towards non-intrusive metabolic rate evaluation for HVAC control," in *17th International Conference on Computing in Civil and Building Engineering (ICCCBE 2018)*, Jun. 2018.

[12] J. Lee and Y. Ham, "Physiological sensing-driven personal thermal comfort modelling in consideration of human activity variations," *Building Research & Information*, vol. 49, no. 5, pp. 512–524, Jul. 2021, doi: 10.1080/09613218.2020.1840328.

[13] H. T. Ebuy, H. Bril El Haouzi, R. Benelmir, and R. Pannequin, "Occupant Behavior Impact on Building Sustainability Performance: A Literature Review," *Sustainability 2023, Vol. 15, Page 2440*, vol. 15, no. 3, p. 2440, Jan. 2023.

[14] W. L. Haskell *et al.*, "Physical activity and public health: Updated recommendation for adults from the American College of Sports Medicine and the American Heart Association," *Med Sci Sports Exerc*, vol. 39, no. 8, pp. 1423–1434, Aug. 2007, doi: 10.1249/MSS.0B013E3180616B27.

[15] V. Farrahi and M. Rostami, "Machine learning in physical activity, sedentary, and sleep behavior research," *Journal of Activity, Sedentary and Sleep Behaviors 2024 3:1*, vol. 3, no. 1, pp. 1–17, Jan. 2024, doi: 10.1186/S44167-024-00045-9.

[16] J. Martins, K. Flanigan, and C. McComb, "Skeleton-based Human Action Recognition in a Thermal Comfort Context," *BuildSys 2023 - Proceedings of the10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, pp. 377–384, Nov. 2023, doi: 10.1145/3600100.3626339.

[17] A. Čulić, S. Nižetić, P. Šolić, T. Perković, and V. Čongradac, "Smart monitoring technologies for personal thermal comfort: A review," *J Clean Prod*, vol. 312, p. 127685, Aug. 2021, doi: 10.1016/J.JCLEPRO.2021.127685.

[18] L. Arakawa Martins, V. Soebarto, and T. Williamson, "A systematic review of personal thermal comfort models," *Build Environ*, vol. 207, p. 108502, Jan. 2022, doi: 10.1016/J.BUILDENV.2021.108502.

[19] Z. Deng and Q. Chen, "Development and validation of a smart HVAC control system for multi-occupant offices by using occupants' physiological signals from wristband," *Energy Build*, vol. 214, p. 109872, May 2020.

[20] C. Lugaresi *et al.*, "MediaPipe: A Framework for Building Perception Pipelines," Jun. 2019, Accessed: Nov. 21, 2024. [Online]. Available: https://arxiv.org/abs/1906.08172v1

[21] M. J. Tsai and H. Y. Lung, "Two-phase optimized inverse kinematics for motion replication of real human models," *Journal of the Chinese Institute of Engineers*, vol. 37, no. 7, pp. 899–914, Oct. 2014, doi: 10.1080/02533839.2014.904474.

[22] G. Georgakis, R. Li, S. Karanam, T. Chen, J. Košecká, and Z. Wu, "Hierarchical Kinematic Human Mesh Recovery," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12362 LNCS, pp. 768–784, 2020, doi: 10.1007/978-3-030-58520-4_45.

[23] ASHRAE Standards, *Thermal Environmental Conditions for Human Occupancy*, Addendum G. ASHRAE, 2013.

[24] M. Younesi Heravi, Y. Jang, I. Jeong, and S. Sajib, "Deep learning-based activity-aware 3D human motion trajectory prediction in construction," *Expert Syst Appl*, vol. 239, p. 122423, Apr. 2024, doi: 10.1016/J.ESWA.2023.122423.

[25] Y. Jang, I. Jeong, M. Younesi Heravi, S. Sarkar, H. Shin, and Y. Ahn, "Multi-Camera-Based Human Activity Recognition for Human–Robot Collaboration in Construction," *Sensors 2023, Vol. 23, Page 6997*, vol. 23, no. 15, p. 6997, Aug. 2023, doi: 10.3390/S23156997.

Understanding Cognitive Impacts of Robots in Worker-Robot Collaboration in Modular Construction

Yifan Wang[1], Xiaoyu Hou[1], Bo Xiao[1*], and Shane T. Mueller[2]

[1]Department of Civil, Environmental, and Geospatial Engineering, Michigan Technological University, USA
[2]Department of Psychology and Human Factors, Michigan Technological University, USA
[*] Corresponding author
wyifan@mtu.edu, xiaoyuho@mtu.edu, boxiao@mtu.edu, shanem@mtu.edu

Abstract –

Human-robot collaboration (HRC) holds significant potential to enhance productivity and efficiency in modular construction (MC). While existing studies in construction HRC have predominantly focused on technical advancements, limited attention has been paid to human factors and the cognitive impacts of robots on workers during collaboration. This study addresses this gap through an experimental investigation. A controlled experiment was conducted to compare conditions with and without a collaborative robot (cobot). Work performance metrics (e.g., time efficiency, angular accuracy, and screwing quality) and human factors (e.g., fatigue levels, individual effort, and stress states) were assessed using statistical analysis and self-report questionnaires. Results of work performance indicate that cobot integration significantly increased task completion time (TCT) (Est. = 0.314, t = 5.18, p < 0.001), with variations observed across tasks (p = 0.00452). Results of human factors revealed increased fatigue (mean scores rising from 18.9 to 22.0), mental effort (mean scores rising from 40.2 to 50.1), and distress (mean scores rising from 12.2 to 15.9) under the cobot condition, suggesting additional cognitive demands during collaboration. These findings underscore the importance of addressing task complexity and cognitive challenges when implementing cobots in MC workflows. This research contributes to advancing robotics integration and safeguarding worker well-being in modular construction.

Keywords –

Human-robot Collaboration; Work Performance; Human Factors; Collaborative Robots; Modular Construction

1 Introduction

To address the lack of labors and improve productivity, modular constriction (MC) has gained significant attention and development in the construction industry recently [1]. MC refers to an innovative construction approach that transfers on-site works to highly controlled and standardized environments in MC factories. Robotics have been increasingly applied in MC to enhance both efficiency and safety in prefabrication and assembly processes [2-4]. For instance, this promising technique has been successfully integrated into prefabricated component assembly [5], robotics welding [6] and robotics platforms [7]. Among these advancements, collaborative robots (cobots) have emerged as a transformative tool, offering the benefits of flexible manufacturing and cost-effectiveness. As a result, human-robot collaboration (HRC) is poised to play a significant role in the future development of MC.

Human involvement remains essential in HRC due to unique cognitive and reasoning abilities, while robots excel at performing repetitive, precision-demanding and hazardous tasks [8]. Despite this, existing studies in MC primarily focus on technological advancements, such as robotic programming and automated assembly systems, while giving less attention to the effects of HRC as a novel working paradigm on human workers. Previous research [9, 10] indicated that working with robots may increase workers' cognitive load and mental stress. However, there remains a need for a systematic evaluation of how robots influence workers' performance and well-being in MC workflows.

This study aims to address this gap through an experimental investigation of the effects of robots on construction workers. A controlled experiment was conducted to isolate variables and compare conditions with and without the cobot. Statistical analysis and questionnaire surveys were employed to assess work performance and workers' cognitive and physical responses. The outcomes of this research will provide insights into the impacts of cobots on both work performance and human factors. Furthermore, this study offers practical recommendations for optimizing human-

robot collaboration by managing task complexity and reducing cognitive load, supporting the effective integration of cobots in MC workflows.

2 Literature Review

This section begins with an overview of the applications of cobots in modular construction and then provides a detailed analysis of their impact on human workers.

2.1 Cobots' Application in MC

The integration of cobots in MC has gained increasing attention for their ability to enhance efficiency, safety, and precision in prefabrication and assembly processes. By shifting construction activities from unstructured on-site environments to highly controlled factory settings, MC provides an ideal platform for HRC [8]. Cobots are designed to complement human workers by performing repetitive, precision-demanding, and hazardous tasks, thereby enabling greater flexibility and productivity in production workflows.

In MC factories, cobots have been successfully applied in various tasks, including component assembly [5], 3D printing [11], and automated welding [6]. For example, Kyjanek et al. (2019) investigated HRC in timber prefabrication, where cobots achieved precise positioning of timber components while humans performed tasks requiring dexterity, such as screwing [12]. The use of augmented reality facilitated real-time communication and task sequencing, ensuring seamless collaboration between workers and robots. Similarly, Zhu et al. (2021) proposed a smart component-oriented method for multi-robot coordination in prefabricated housing [13]. By utilizing smart construction objects within a Building Information Modeling (BIM) environment, tasks such as transporting and assembling components were efficiently allocated, improving coordination and streamlining collaborative workflows.

While these applications demonstrate the potential of cobots to streamline MC processes, the integration of cobots also introduces new challenges. Existing research has primarily focused on the technical advancements of cobot systems, such as programming optimization, sensor-based control, and hardware innovations [5, 6, 14]. However, limited attention has been given to understanding how cobots affect human workers, particularly regarding their performance and well-being. This highlights the need for a systematic evaluation of cobot integration from a human-centered perspective.

2.2 Impact Evaluation of HRC

Despite the importance of understanding cobots' impact on construction workers, research in this area remains limited. Existing studies primarily rely on methods such as self-report questionnaires, interviews, and simulations [15]. For example, Bademosi and Issa (2021) used interviews and surveys with industry experts to explore the challenges of implementing automation and robotic technologies in construction [16]. Similarly, Wu et al. (2022) developed a simulation-based framework to assess the impact of HRC on construction productivity [17]. However, controlled experimental methods are relatively scarce, even though experiments offer stronger empirical evidence and better reflect real-world conditions, making them more persuasive for evaluating cobots' practical impacts.

Work performance is one of the key metrics for evaluating human-robot interaction in HRC. Human factors are equally critical in assessing HRC, as the introduction of cobots should remain human-centered. Baltrush et al. (2022) identified four key job quality factors relevant to HRC: cognitive workload, collaboration fluency, trust, and acceptance and satisfaction [19]. However, construction tasks often involve significant physical labor, highlighting the need to assess physical fatigue and exertion. For example, experimental research conducted by Fang et al. (2015) highlighted the importance of fatigue on construction workers performance [20]. Therefore, this study focuses on fatigue levels, individual effort, and stress levels to comprehensively evaluate the human factors associated with cobot integration in construction.

3 Methodology

Experimental analysis method was used in this study to evaluate the impact of cobots on human workers in terms of work performance and human factors. Figure 1 illustrates the overall methodology and process of this study. A typical manufacturing task for a wooden wall panel was designed and decomposed into 9 subtasks. Next, a total of 6 subjects were recruited and a controlled experiment was designed to eliminate the impact of extraneous variables. After setting up an experiment workspace, hardware devices, software system, and tools, each subject was asked to individually conduct experiments in both with/without a cobot for comparison. While the work performance was evaluated from time efficiency, angular accuracy, and screwing quality, human factors were investigated through self-report scales in terms of fatigue levels, individual effort, and stress states.

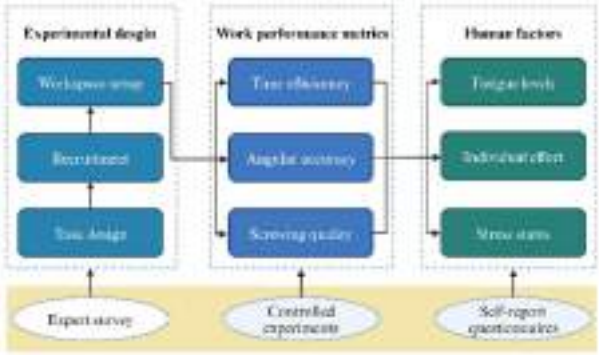


Figure 1. The overall methodology of this study

3.1 Task Design

A typical wooden wall panel manufacturing task in MC was designed for this study. Wood framing is a primary building method in MC for residential, commercial, and industrial structures [21]. Among these, wooden wall panels play a critical role as both structural and functional components. To enhance production efficiency and quality, cobots are increasingly integrated into manufacturing tasks. However, their integration requires close collaboration with human workers, raising concerns about potential impacts on human performance and well-being. Therefore, it is crucial to investigate these impacts in the context of wooden wall panel manufacturing.

The manufacturing task was decomposed into nine subtasks: assembling the framework (T1), screwing to connect the framework (T2), covering with an OSB board (T3), screwing to attach the board to the framework (T4), flipping the panel (T5), adding insulation (T6), covering with another board (T7), screwing again to attach the board (T8), and finally moving the completed product (T9). These tasks were designed to replicate a realistic production process, requiring both physical and cognitive effort from human workers.

To compare the effects of conditions with and without a cobot on human workers, a controlled experiment was designed with two groups. In the baseline group, workers completed the entire task in the same workspace with the cobot powered off. In the cobot group, the cobot was activated and assigned to perform subtask 1, which involved picking and placing wooden strips to assemble the framework. The cobot continuously repeated this task throughout the experiment, while human workers completed the entire sequence of subtasks 1 through 9. Each participant was required to complete both the baseline and cobot conditions on different days to minimize the impact of task repetition.

3.2 Recruitment

A total of six participants, all aged 18 or older, were recruited for the study. The group consisted of three male and three female undergraduate students from the Department of Civil, Environmental, and Geospatial Engineering at Michigan Technological University (MTU). None of the participants had a history of back disorders. This group was selected because of their experience with construction-related activities, making them well-suited to simulate real-world workers while ensuring safety during the experiment.

The recruitment process complied with guidelines with approval from the MTU Institutional Review Board (IRB), under IRB #: IRB-2024-90. All participants signed informed consent forms, voluntarily agreeing to participate with the option to withdraw at any time.

3.3 Workspace Setup

The experimental setup included a UR10e robot equipped with a Robotiq 2F-85 gripper, programmed and controlled using RoboDK software, as shown in Figure 2. The UR10e robot is a versatile and flexible collaborative robot widely used for industrial tasks [6, 22], offering high precision and adaptability for various assembly operations. RoboDK software provides a user-friendly interface for programming and simulating robot tasks [23], making it ideal for ensuring accurate control and program efficiency during experiments.



Figure 2. Experimental workspace setup

A dedicated workspace was enclosed with safety barriers to ensure controlled and interference-free experimental conditions. Wooden wall panel components were scaled down to half their actual size (1:2 ratio) to fit the workspace, with all wooden strips and boards pre-cut and pre-drilled to facilitate assembly. Safety equipment, including helmets, safety glasses, and woodworking gloves, is required within the workspace. Prior to the experiment, safety training and an operation test on the use of an electric screwdriver were conducted. The entire experiment was overseen and safeguarded by an experienced safety assistant.

3.4 Measurement on Work Performance

The impact of cobots on work performance was evaluated in terms of time efficiency, angular accuracy, and screwing quality. Time efficiency was quantified by recording and comparing the task completion time (TCT) for each subtask between the baseline and cobot groups. To stabilize variance and make the data more suitable for linear modeling, a log transformation was applied to the time variable. For further analysis, two linear mixed-effects models (LMMs) were developed using RStudio. LMMs are appropriate for data that include both fixed effects (predictors of interest) and random effects (sources of variability that must be accounted for but are not the primary focus). The first model examined the effects of experimental conditions (baseline and cobot groups), subtasks (T1 to T9), subjects (individual participants), and trial progression (repeats per participant) on time performance. To better understand the interactive effects among these factors, the second model analyzed how the impact of the cobot condition varied across different subtasks.



Figure 3. Distribution angles for the framework

Angular accuracy was assessed by evaluating the total accuracy of the assembled wooden framework and comparing it to the design specifications. Figure 3 shows the distribution of the eight defined angles at each lumber connection within the framework. Raw angle data were collected using a digital angle finder. The accuracy of each joint was calculated using Equations (1) and (2), with the total accuracy of the structure determined using Equation (3) [15]:

$$err = \frac{M_e - M_t}{M_t} \times 100\% \quad (1)$$

$$acc = 1 - \underset{joint\ 8}{err} \quad (2)$$

$$acc_{total} = \prod_{k=joint\ 1} acc_k \quad (3)$$

where err = percentage error; M_e = experimental values measured during assembly; $M_t = 90°$, which is the theoretical design value; acc = percentage accuracy, acc_{total} = the total accuracy of an assembled wooden framework (from joint 1 to joint 8); and acc_k = percentage accuracy of each individual joint.

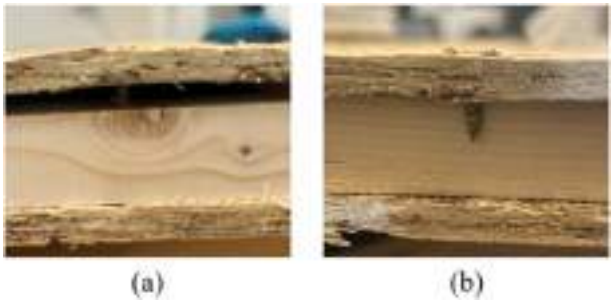


Figure 4. Two types of failed screws

Screwing quality was evaluated as the percentage of failed screws relative to the total number used in the product. Failed screws included two types: those that did not fully penetrate the structure (Figure 4(a)) and those that fully penetrated but failed to connect to the substructure (Figure 4(b)).

3.5 Measurement on Human Factors

The impact of cobots on human factors was evaluated based on fatigue levels, individual effort, and stress states. Data for these dimensions were collected through self-report questionnaires administered after each complete round of nine subtasks. The process was facilitated by the Psychology Experiment Building Language (PEBL) [24], an open-source platform that enables the design and execution of psychological experiments, streamlining the collection and analysis of scale-based data.

Fatigue levels were assessed using the Fatigue Assessment Scale for Construction Workers (FASCW) [25], a 10-item scale where each question is scored on a 5-point scale. The total score is obtained by summing all 10 items. The FASCW comprises two factors: lethargy and bodily ailment. Lethargy, primarily associated with mental fatigue, is calculated as the sum of questions 1, 3, 5, 7, and 9, while bodily ailment, reflecting physical fatigue, is derived from the sum of questions 2, 4, 6, 8, and 10.

Individual effort was measured using Borg's Rating of Perceived Exertion (RPE) [26] for physical effort and Zijlstra's Rating Scale for Mental Effort (RSME) [27] for mental effort. The Borg RPE is a numerical scale ranging from 6 to 20, where participants rate their overall exertion, considering physical stress and fatigue during the task. The Zijlstra RSME, a unidimensional scale, measures subjective mental workload. It consists of a 0–150 scale marked with nine descriptive anchor points representing increasing levels of mental effort.

Stress states were assessed using the Short Stress State Questionnaire (SSSQ) [28], a 24-item scale where each question is rated on a 5-point scale. The SSSQ

comprises three factors—distress, engagement, and worry—each represented by eight questions. Distress reflects negative states such as unpleasant mood, tension, lack of confidence, and perceived loss of control. Engagement captures positive states, including task interest, motivation, energetic arousal, and concentration. Worry focuses on cognitive interference, self-focused attention, and self-esteem.

4 Results and Discussion

A total of 12 experimental groups were conducted, involving six participants under two experimental conditions: baseline and cobot. Each group completed two repetitions of the entire process consisting of nine subtasks. However, four groups were excluded because participants #3 and #4 did not complete the entire process. As a result, 16 data groups (4 participants × 2 conditions × 2 repetitions) were included in the analysis.

4.1 Results of Work Performance

This section examines the impacts of cobots on human work performance from three perspectives: time efficiency, angular accuracy, and screwing quality. Figure 5 illustrates the TCT results under baseline and cobot conditions. Figure 5(a) shows the distribution of TCT in seconds, while Figure 5(b) presents the log transformed TCT distribution for improved clarity and visual comparison. Figure 5(c) displays the mean log transformed TCT across the nine subtasks. In most subtasks, the cobot group exhibited longer completion time compared to the baseline group. However, for subtasks 5 (flipping the panel) and 6 (adding insulation), the differences were minimal or reversed.

To further analyze the effects of experimental conditions, subtasks, subjects, and trial progression on time performance, two LMMs were developed using RStudio. In the first LMM, experimental conditions, subtasks, and trials were included as fixed effects, while subjects were treated as a random effect. The response variable was defined as $LogTCT = \log(time)$ to stabilize variance and improve model fit.

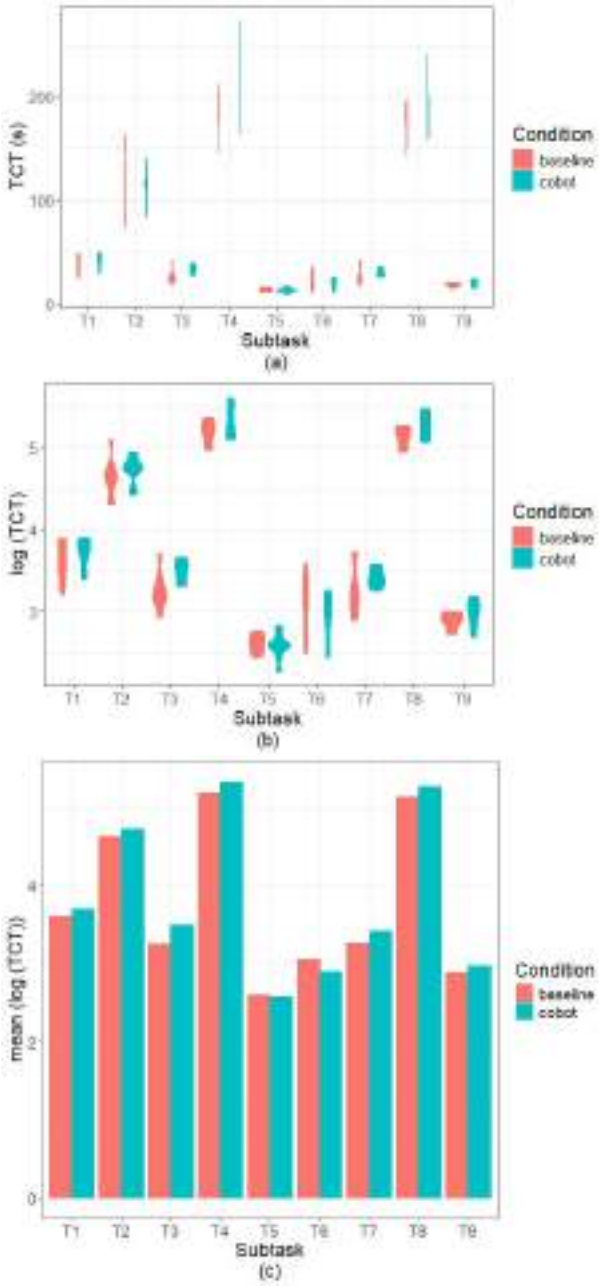


Figure 5. Performances of time efficiency

Table 1. Summary of the LMM coefficients

Groups	Est.	SE	t-value	p-value
Intercept	3.761	0.088	42.94	< 2.2e-16
Condition	0.314	0.061	5.18	< 2.26e-07
Trials	-0.058	0.014	-4.15	< 3.29e-05
Subtasks	df = 8, Significant, p < 2.2e-16			
Subjects	Var = 0.0199, SD = 0.141			
Residual	Var = 0.0213, SD = 0.146			

Table 1 summarizes the LMM results. The "cobot" condition significantly increased $logTCT$ compared to

the baseline data ($Est. = 0.314, t = 5.18, p < 0.001$), suggesting that cobot involvement slowed task completion, potentially due to worker distraction or mistrust of the robot. Trail progression also had a significant effect on time performance ($Est. = -0.058, t = -4.15,\ p < 0.001$), indicating that task completion times decreased with repetition, likely reflecting learning and adaptation. Subtasks significantly influenced time performance ($p < 0.001$), with varying effects observed across different tasks. Additionally, the inclusion of a random effect for subjects ($Var = 0.0199, SD = 0.141$) accounted for individual differences in time efficiency, such as variations in skill or experience. Residual variance of the model was 0.0213 ($SD = 0.146$), reflecting some unexplained variability in the data.

Given the significant influence of both condition and subtasks, the interaction between them was further examined using a second LMM. Results revealed a significant "condition * subtasks" interaction ($p = 0.00452$), indicating that the cobot condition's impact on TCT varied across subtasks. Specifically, the main differences observed in subtasks 5 and 6 may be attributed to the nature of these tasks, which require less direct interaction with the cobot. These findings suggest that the cobot's influence on time efficiency is task-dependent, highlighting the need for further investigation into the underlying mechanisms.

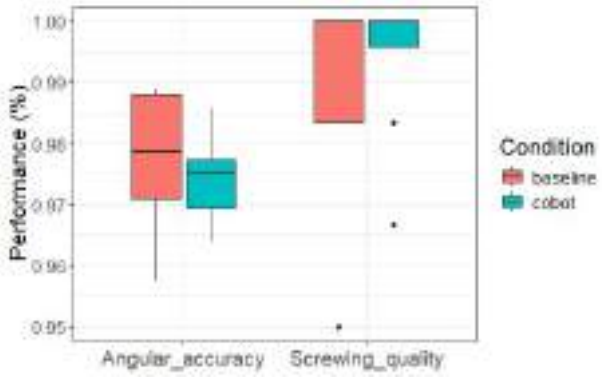


Figure 6. Performances of angular accuracy and screwing quality

Figure 6 presents the results for angular accuracy and screwing quality under the two experimental conditions. Angular accuracy was evaluated by comparing the total structural accuracy of the assembled wooden framework with the design specifications. The results indicate a slight reduction in angular accuracy in the cobot condition. This minimal effect can be attributed to the framework's structural properties, where the extrusion between wooden strips naturally constrains angular deviations, reducing observed variability.

Screwing quality, measured as the percentage of failed screws, showed negligible differences between the two conditions. The simplicity of the screwing task likely contributed to the lack of observable differences. These results suggest that for straightforward tasks, such as screwing, collaborative robots have a limited impact on the performance of screwing quality. Future studies should consider increasing task complexity to better assess the influence of cobots on such outcomes.

4.2 Results of Human Factors

This section examines the impact of cobots on human workers from three human factor perspectives: fatigue levels, individual effort, and stress states. Figure 7 presents the results of fatigue levels measured by the FASCW. The total average FASCW score shows an increase from 18.9 to 22, suggesting that participants experienced greater fatigue under the cobot condition. Specifically, both FASCW factors—lethargy (associated with mental fatigue) and bodily ailment (associated with physical fatigue)—demonstrated upward trends, with lethargy showing a more pronounced increase.

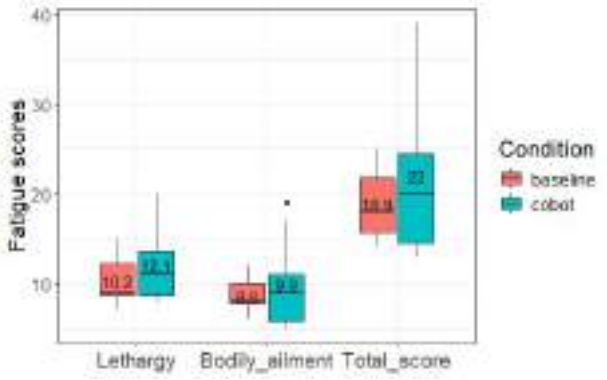


Figure 7. Results of fatigue levels

Figure 8 illustrates the results of physical and mental effort, measured using Borg's RPE scale and Zijlstra's RSME, respectively. The results reveal a slight increase in RPE scores (mean rising from 11.1 to 11.9) and a more substantial increase in RSME scores (mean rising from 40.2 to 50.1) under the cobot condition. This indicates that collaboration with robots resulted in a greater increase in mental effort compared to physical exertion.

These findings on individual effort align with the fatigue results, as both measures highlight increases in cognitive or mental dimensions (i.e., lethargy and mental effort). The observed effects may be attributed to the increased cognitive load and potential distractions introduced by the cobot.

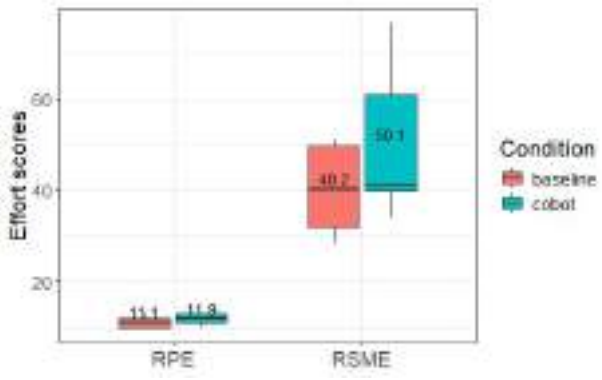


Figure 8. Results of individual effort

Figure 9 presents the results of the SSSQ for assessing the stress levels. The findings reveal a significant increase in distress scores (mean rising from 12.2 to 15.9) under the cobot condition, while engagement and worry exhibited only slight changes. This suggests that collaboration with cobots increases emotional strain and discomfort, distinguishing it from the psychological interference associated with worry and the motivational focus represented by task engagement.

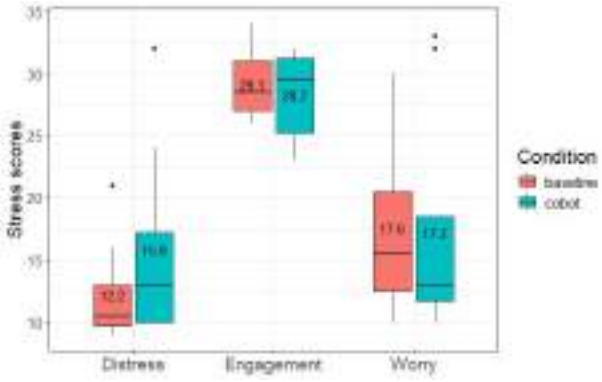


Figure 9. Results of stress states

4.3 Limitations and Future Works

This study has certain limitations. The sample size was relatively small (N=6), and the participants were undergraduate students rather than experienced construction workers. While these participants had relevant construction-related knowledge, their responses and performance may not fully represent those of seasoned professionals. Future research should involve a larger and more diverse participant pool, including experienced construction workers, to enhance the generalizability of the findings.

While cobots offer significant potential for enhancing productivity, their cognitive impact on human performance and well-being must be carefully managed. Future research should explore more complex tasks and diverse working scenarios to gain deeper insights into the dynamics of human-robot collaboration. By incorporating tasks that more closely mimic real-world MC workflows, future research can provide more practical and actionable insights for optimizing cobot integration in industrial settings.

5 Conclusion

The study assessed the impacts of collaborative robots on construction workers' work performance and human factors. Work performance was assessed in terms of time efficiency, angular accuracy, and screwing quality, while human factors were examined through fatigue levels, individual effort, and stress states. The results showed that cobot involvement significantly increased task completion time (Est. = 0.314, t = 5.18, $p < 0.001$), with variability observed across different task types ($p = 0.00452$). Angular accuracy showed a slight decline, though structural constraints reduced deviations, while screwing quality remained largely unaffected due to the task's simplicity. Human factor results revealed increased fatigue levels, with mean scores rising from 18.9 to 22.0, mental effort scores increasing from 40.2 to 50.1, and distress scores rising from 12.2 to 15.9 under the cobot condition. These findings suggest that collaborating with robots increases cognitive load and emotional strain among workers, highlighting the need for careful consideration of task complexity and worker cognitive demands when integrating cobots into MC workflows.

References

[1] C.-H. Yang, T.-H. Wu, B. Xiao, and S.-C. Kang. Design of a Robotic Software Package for Modular Home Builder. In *Proceedings of the 36th international symposium on automation and robotics in construction, ISARC*, pages 1217-1222, Banff, Canada, 2019.

[2] B. Xiao, C. Chen, and X. Yin. Recent advancements of robotics in construction. *Automation in Construction*, 144:104591, 2022.

[3] C.-J. Liang, X. Wang, V. R. Kamat, and C. C. Menassa. Human–robot collaboration in construction: classification and research trends. *Journal of Construction Engineering and Management*, 147(10):03121006, 2021.

[4] B. Xiao, Y. Wang, Y. Zhang, C. Chen, and A. Darko. Automated daily report generation from construction videos using ChatGPT and computer vision. *Automation in Construction*, 168:105874, 2024.

[5] C. Liu, J. Wu, X. Jiang, Y. Gu, L. Xie, and Z. Huang. Automatic assembly of prefabricated components

based on vision-guided robot. *Automation in Construction*, 162:105385, 2024.

[6] D. Lee and K. Han. Vision-based construction robot for real-time automated welding with human-robot interaction. *Automation in Construction*, 168:105782, 2024.

[7] H. J. Wagner, M. Alvarez, O. Kyjanek, Z. Bhiri, M. Buck, and A. Menges. Flexible and transportable robotic timber construction platform – TIM. *Automation in Construction*, 120:103400, 2020.

[8] Y. Fu, J. Chen, and W. Lu. Human-robot collaboration for modular construction manufacturing: Review of academic research. *Automation in Construction*, 158:105196, 2024.

[9] S. Shayesteh and H. Jebelli. Investigating the impact of construction robots autonomy level on workers' cognitive load. In *Proceedings of the Canadian society of civil engineering annual conference 2021*, pages 255-267, Singapore, 2022.

[10] F. Baek, D. Kim, G. Lee, B. Choi, and S. H. Lee. Effect of human emotional responses on human-robot team pty in construction. *Automation in Construction*, 166:105608, 2024.

[11] M. Alabbasi, A. Agkathidis, and H. Chen. Robotic 3D printing of concrete building components for residential buildings in Saudi Arabia. *Automation in Construction*, 148:104751, 2023.

[12] O. Kyjanek, B. Al Bahar, L. Vasey, B. Wannemacher, and A. Menges. Implementation of an Augmented Reality AR Workflow for Human Robot Collaboration in Timber Prefabrication. In *Proceedings of the 36th international symposium on automation and robotics in construction, ISARC*, pages 1223-1230, Banff, Canada, 2019.

[13] A. Zhu, P. Pauwels, and B. De Vries. Smart component-oriented method of construction robot coordination for prefabricated housing. *Automation in Construction*, 129:103778, 2021.

[14] B. Belousov et al. Robotic architectural assembly with tactile skills: Simulation and optimization. *Automation in Construction*, 133:104006, 2022.

[15] X. Liang, U. Rasheed, J. Cai, B. Wibranek, and I. Awolusi. Impacts of collaborative robots on construction work performance and worker perception: Experimental analysis of human–robot collaborative wood assembly. *Journal of Construction Engineering and Management*, 150(8):04024087, 2024.

[16] F. Bademosi and R. R. A. Issa. Factors influencing adoption and integration of construction robotics and automation technology in the US. *Journal of Construction Engineering and Management*, 147(8):04021075, 2021.

[17] M. Wu, J. R. Lin, and X. H. Zhang. How human-robot collaboration impacts construction productivity: An agent-based multi-fidelity modeling approach. *Advanced Engineering Informatics*, 52:101589, 2022.

[18] D. Mitterberger, L. Atanasova, K. Dörfler, F. Gramazio, and M. Kohler. Tie a knot: Human–robot cooperative workflow for assembling wooden structures using rope joints. *Construction Robotics* 6(3):277–292, 2022.

[19] S. J. Baltrusch, F. Krause, A. W. de Vries, W. van Dijk, and M. P. de Looze. What about the human in human robot collaboration? *Ergonomics*, 65(5):719–740, 2022.

[20] D. Fang, Z. Jiang, M. Zhang, and H. Wang. An experimental method to study the effect of fatigue on construction workers' safety performance. *Safety Science*, 73:80–91, 2015.

[21] N. Gutiérrez, J. Negrão, A. Dias, and P. Guindos. Bibliometric review of prefabricated and modular timber construction from 1990 to 2023: Evolution, trends, and current challenges. *Sustainability*, 16(5):2134, 2024.

[22] S. Espinosa, M. Slepicka, and A. Borrmann. Integration of laser profiler feedback into FIM for additive manufacturing in construction. In *Proceedings of the 41st international symposium on automation and robotics in construction, ISARC*, pages 89-96, Lille, France, 2024.

[23] I. M. Sarivan, O. Madsen, and B. V. Wæhrens. Automatic welding-robot programming based on product-process-resource models. *International Journal of Advanced Manufacturing Technology*, 132(3):1931–1950, 2024.

[24] S. T. Mueller and B. J. Piper. The psychology experiment building language (PEBL) and PEBL test battery. *Journal of Neuroscience Methods*, 222:250–259, 2014.

[25] M. Zhang et al. Development and validation of a fatigue assessment scale for U.S. construction workers. *American Journal of Industrial Medicine*, 58(2):220–228, 2015.

[26] G. A. V. Borg. Psychophysical bases of perceived exertion. *Medicine and Science in Sports and Exercise*, 14(5):377–381, 1982.

[27] F. R. H. Zijlstra. Efficiency in work behaviour: A design approach for modern tools. PhD diss., Delft University of Technology, Delft: Delft University Press, 1993.

[28] W. S. Helton. Validation of a short stress state questionnaire. In *Proceedings of the human factors and ergonomics society 48th annual meeting*, pages 1238-1242, New Orleans, USA, 2004.

Drone-Induced Distraction on Construction Sites: A Quantitative Analysis of Flight Configurations

Wei Han[1], Liqun Xu[1], Mahfuza Maisha Mouri[2], Wei-Yin Loh[3], Fei Dai[2], Zhenhua Zhu[1]

[1]Department of Civil and Environmental Engineering, University of Wisconsin - Madison, USA
[2]Wadsworth Department of Civil and Environmental Engineering, West Virginia University, USA
[3]Department of Statistics, University of Wisconsin - Madison, USA
whan59@wisc.edu, liqun.xu@wisc.edu, mm00286@mix.wvu.edu, loh@stat.wisc.edu, fei.dai@mail.wvu.edu, zzhu286@wisc.edu

Abstract –

Drones are increasingly used in construction projects to facilitate tasks such as progress tracking, safety monitoring, and quality inspection. Despite their advantages, the presence of drones may introduce distractions that compromise workers' ability to identify hazards, thereby elevating the risk of accidents. While limited studies have conducted quantitative assessments of drone-induced distractions among construction workers, the interplay between specific drone flight configurations and the degree of distraction remains underexplored. This study investigated the influence of varying drone flight configurations on workers' attentional focus through quantitative experiments. Seven virtual construction site scenarios were developed to simulate both the visual and auditory stimuli associated with drones. A total of 17 participants engaged in simulated construction tasks within these scenarios, utilizing virtual reality systems and an omnidirectional treadmill. Their situational awareness was evaluated through a detailed analysis of gaze fixation times and counts. The results indicated that the presence of drones significantly distracted participants, with distraction levels intensifying as the drones approached closer. These findings contributed to the understanding of how drone proximity impacts worker attention and safety and established a basis for refining safety protocols in construction environments where drones are increasingly employed.

Keywords –

Unmanned aerial vehicles (UAVs); Visual attention; Eye-tracking; Virtual reality; Worker safety

1 Introduction

The increasing applications of drones in construction projects have gained significant attention from the industry due to their potential to enhance efficiency and safety. However, concerns have emerged regarding the safety risks and hazards that drone operations may introduce to construction sites. These risks extend beyond the physical hazards of collisions with workers, equipment, or site infrastructure within drone operating zones [1]. The presence of drones has the potential to divert workers' attention, leading to visual, auditory, and cognitive distractions [2]. For example, Union Pacific Corp railroad workers reported being distracted by drones flying over rail yards, causing them to look upwards rather than focusing on their tasks, even in the presence of 200-ton locomotives and railcars in motion [3].

The impact of distraction on workers' focus has been widely documented in other domains. For example, Mark et al. showed that external interruptions in office environments could delay workers' return to full task concentration by up to 25 minutes [4]. Similar interruptions were observed in construction settings, where distractions caused by drones can impair hazard recognition and safety risk assessment [5] [6]. Such distractions increase the likelihood of errors, endangering both the distracted worker and others on-site [7].

The workers at heights are vulnerable to drone-induced distractions since the drones often fly above buildings on construction sites. These distractions may cause workers to overreach or lose balance on elevated platforms cluttered with tools or areas lacking protections such as unguarded edges and openings. Consequently, the risk of falls, slips, and trips increases, often resulting in serious injuries or fatalities. The Bureau of Labor Statistics (BLS) reported that falls, slips, and trips accounted for 38.4% of all fatalities and 47.4% of nonfatal injuries in the construction industry in 2022; and

most fatal falls, slips, and trips happened to those workers at heights who fell to a lower level [8]. Moreover, a survey of over 1,200 safety professionals identified distraction as the leading cause of these incidents [9].

Although previous studies have acknowledged the safety challenges posed by drones in construction, much of the existing research relies on surveys, expert opinions, and conceptual analyses, with limited experimental data to evaluate the extent of drone-induced distractions [10], [11]. To address these limitations, Albeaino et al. provided initial experimental data, demonstrating that drones can cause measurable distractions among construction workers [12]. However, the influence of specific drone flight configurations on distraction levels remains unexplored.

This study aims to address this gap by examining how drone flight height and approach direction affect workers' safety and attention. These variables were chosen because they are key factors influencing the degree to which drones capture workers' attention. For instance, a drone's flight height can alter its visibility and audibility, while the approach direction (e.g., from the front or back) may impact how quickly workers become aware of its presence. To investigate these effects, a few virtual construction environments were developed, and participants were tasked with performing material handling activities. Gaze fixation times and counts were collected to evaluate visual attention toward hazardous areas. By analyzing this data, our study provided a quantitative assessment of how drone flight configurations influence workers' hazard awareness, contributing to a deeper understanding of drone-induced distractions and their implications for construction site safety.

2 Literature Review

2.1 Drone Application in Construction Sites

Drones, or unmanned aerial vehicles (UAVs), are aircraft that can be remotely controlled or autonomously flown to perform various tasks on construction sites. Equipped with sensors such as RGB, multispectral, thermal, or LIDAR, drones can quickly capture substantial aerial data [13]. This data can be transmitted to computers through drone software, enabling users to analyze and interpret it efficiently. These capabilities enhance construction project efficiency, support better project management, and improve the thoroughness of inspections.

Drones have numerous applications in construction, including topographic mapping, equipment tracking, remote monitoring, and worksite inspection. For topographic mapping, drones significantly reduce the time needed to survey a site's landscape by covering large areas quickly. The high-resolution images captured can be converted into 3D models to identify site challenges and assist with layout planning before construction begins. For instance, Jiang et al. [14] advanced image-based modeling techniques for real-time 3D construction site modeling, proving effective for rapid layout planning in hoisting projects. Additionally, drones help track equipment locations on-site, record operational errors, and monitor construction quality. Sun [15] proposed a UAV-based system to track road construction machinery, enhancing paving quality. Furthermore, drones allow companies to visually inspect large structures or those in hard-to-reach areas more efficiently and at a lower cost. For example, de Melo et al. demonstrated that UAV-collected visuals can improve safety inspections by providing better visibility of job site conditions [16].

2.2 Drone Safety Analysis in Construction Sites

The Federal Aviation Administration (FAA) has implemented regulations for remote pilot certification and operational guidelines for small unmanned aircraft in civil applications, aiming to address safety concerns associated with drones in commercial settings [17]. From April 21, 2021, the FAA permits drones to operate over people, contingent upon the level of risk they pose to individuals on the ground [18]. While these regulations provide a foundational standard for safe drone operations in workplace environments, they do not specifically account for the potential distractions drones may cause, particularly for workers working at heights.

Recently, the potential safety risk related to drone applications on construction sites attracted more attention. Based on the survey of 56 and 69 construction experts and professionals, drone-induced distraction is one of the primary safety concerns in construction projects [10][11]. Moreover, Jeelani and Gheisari investigated different drone applications on construction sites and identified their negative impacts on related workers' health and safety [2]. Fan and Saadeghvaziri identified drone-induced distraction as a key challenge in the planning, design, construction, and maintenance of civil infrastructure [19]. Costa et al. also highlighted that drone-induced distraction may cause issues with workers, concentration, safety, and productivity [20]. These earlier qualitative studies showed that drone-induced distraction can impact construction worker safety.

Thus, further studies started to quantify how drone-induced distraction, because it can provide more solid evidence for the decision-making process of drone use on construction sites. For example, Monte Carlo simulations were used to quantify the direct and indirect risks of drones to construction workers [21], [22]. Although the

simulations can provide valuable insights, they did not study the actual reactions of the workers. Thus, Albeaino et al. conducted human study experiments to investigate the impact of drone-induced distractions on workers [12]. The study revealed that participants spent approximately 18 seconds observing the drones during a 2.5-minute task, highlighting the considerable potential for distraction in active work settings.

2.3 Drone Distraction Studies Using Virtual Reality

To study how drones distract construction workers and impact safety, we need to observe worker responses with and without drone distractions. Since real-life testing poses safety risks, we use virtual reality (VR) to simulate high-risk environments safely [23]. VR replicates real-world conditions in a controlled, risk-free manner and has been shown to produce results closely aligning with real-world experiments in both behavioral outcomes and self-reported measures [24]. It has been widely used in transportation and construction safety research to assess drone-induced distractions on drivers and workers.

In the transportation domain, Barlow et al. employed a VR-based driving simulator to investigate the impact of drone flights near roadways on participants' visual attention and their potential contribution to distraction [25]. Their findings led to the recommendation of a policy requiring drone operations to maintain a minimum distance of 25 feet from the edge of a lane. Similarly, the Oregon Department of Transportation utilized a VR simulator to conduct a randomized factorial experiment evaluating the effects of three independent variables—lateral offset, flight path, and lane use—on driver distraction caused by drone operations near roadways. The study revealed that the frequency and duration of glances at drones increased when drones were positioned closer to the roadway, with unsafe glances exceeding 2 seconds observed at lateral offsets of 0 feet, 25 feet, and 50 feet [26]. In another study, Ryan et al. employed a driving simulator featuring an actual driver's seat within a real vehicle and a screen positioned in front of the driver. Their findings indicated that drivers were distracted by drones, with critically long glances observed in 11% of cases [27]. These studies collectively demonstrated that VR and simulator-based approaches are effective tools for examining drone-induced distractions in driver participants. They used monitors to display the virtual environments for the drivers.

In the construction domain, the researchers adopted VR headsets to facilitate human-subject experiments and training. VR headsets allow construction worker participants to engage deeply with virtual construction environments, creating a highly immersive setting for studying distraction effects. Jeelani and Gheisari utilized virtual reality to simulate a construction site for conducting experiments aimed at evaluating the impact of drone presence on workers' attention and stress levels. The findings indicated that unmanned aerial vehicles (UAVs) could distract workers, especially when operating at distances of 12 feet and 25 feet. However, the drones did not have a significant effect on workers' emotional or psychological distress [28]. Albeaino et al. employed virtual reality to create a simulated construction site, enabling the evaluation of the attentional impact of drones on construction professionals at varying distances. The study revealed that drones positioned at greater distances resulted in more frequent and prolonged visual fixations on drones, emphasizing their potential to distract workers and providing insights into safer operational distances for drone use on construction sites [29].

Most of the existing studies focused on how the presence of drones diverts participants' visual attention from their primary tasks. However, they did not explore how this diversion of attention affects participants' ability to recognize and identify hazards.

2.4 Experimental Design for Drone-Induced Distraction

To study the human's visual attention to drone-induced distraction, the approaches of human study experiments from previous studies can be divided into two categories. First, some studies adopted between-subject design [2], [12], [29]. A between-subjects design is a research method in which different groups of participants are exposed to distinct experimental conditions, allowing comparisons to be made between groups without the influence of repeated exposure [30]. Second, other studies adopted within-subject design [25], [26], [27]. A within-subjects design, also known as a repeated-measures design, is a research method in which the same participants are exposed to all experimental conditions, allowing researchers to compare their responses across different treatments while controlling for individual variability.

The studies collected the participants' fixation time and/or count to analyze their visual attention during distractions. Studies using a within-subject design exposed participants to the same set of conditions, which may cause sequence effects, where the order of exposure impacts the outcomes. Thus, usually, participants will experience the same set of conditions but with different sequences. The design helps the researchers to investigate if there is any sequence effect. Some of the studies fit the variables for evaluation into and linear mix model to check if the sequence effect is significant or not. Then, the variables are further analyzed by analysis of variance (ANOVA) to determine the significance of variables across different conditions.

3 Problem Statement, Objective, and Research Question

Existing studies emphasized the potential hazards posed by drone-induced distractions for construction workers. Although they have investigated several aspects of drone-related distraction, a systematic examination of the influence of specific drone flight configurations on workers' situational awareness remains unexplored. Thus, this study aims to analyze workers' visual attention to hazards in response to drones operating at varying heights and directions. The research question of this research is: "How is the amount of distraction related to drone flying height and direction?"

4 Experiment Design and Implementation

4.1 Procedure



Figure 1. A participant performing material handling tasks in a virtual construction environment using a VR headset and omni-directional treadmill.

In the beginning, participants were provided with an IRB-approved consent form, which they were required to read and sign before participation. The consent form detailed the purpose of the study, emphasizing its focus on drone safety in construction site environments. As shown in Figure 1, participants were equipped with a VR headset (Meta Quest Pro) and positioned on an omni-directional treadmill (ODT) (KAT Walk C2+). The VR headset facilitated full immersion in virtual construction environments developed using Unity. The virtual environments provide a realistic and immersive experience by incorporating real background construction noises and drone flying sounds. Using Unity further enhances this realism by enabling spatially accurate drone sounds, allowing participants to hear the drone's flying noise in stereo based on its distance and location. The ODT enabled participants to navigate the virtual environments by physically walking, replicating the physical demands of real-world construction tasks. After a 5-minute training process, the participants then move on to the official experiment scenes.

For our experiment settings, they included three independent variables, namely, drone presence (Yes and No), height of drone (16, 48, and 82 ft.), and direction of drone (Back and Front). Three heights of drones were chosen since they are the working heights of the drones for visual inspections [31]. By these three independent variables, there were a total of six combinations of drone height and direction, along with a seventh condition where no drone was present.

For our experiment design, a repeated measures design was adopted to minimize variability across participants and enhance the reliability of the findings [32]. Thus, each participant was exposed to all seven scenes in a randomized order to minimize learning effects and bias. To be more specific, participants were randomly divided into seven groups. The participants in different groups experienced different sequences of scenes. The design of the sequence in each group follows a Latin Square design [33], which ensured an approximate balance in the sequence of scene presentations.

In each scene, participants were tasked with carrying a 16.5" x 12" x 14" box along a designated route from a starting location to an endpoint twice. During the task, a virtual drone (DJI Mavic 3 Pro) appeared randomly in one of the two trials. The selections of drone type and required task are made in consultation with Mortenson, a construction company, to ensure the relevance of construction practice. Figure 2 illustrates a virtual construction scene, where the participant's route is depicted by the yellow lines (dashed and solid). To ensure the drone appears from either the front or behind the participant, each virtual construction site included a straight segment of the route, represented by the solid yellow line. The drone's flight path is marked by the red dashed line.



Figure 2. A sample of the virtual construction site

4.2 Data Collection, Processing, and Analysis

This study aims to investigate the situation awareness of workers through their visual attention to hazards. Thus, eye gaze fixation time and count on danger areas were chosen as our dependent variable for analysis. The danger areas included openings and edges and the fences around them in the virtual environment.

During the experiment, the participants' gaze points were collected by VR headset with a frequency of 100 Hz. The raw data consisted of gaze points recorded sequentially over time, resulting in a total of 57,600 data points. For each scene, fixation time and count were computed by summing the fixation events occurring within predefined danger areas. This approach enabled the determination of the total fixation time and count on danger areas for each scene.

To analyze the effects of different drone flying configurations on participants' fixation time and count, a repeated measures linear mixed model was applied. Due to the occurrence of zero fixation times and counts, the model was applied to the natural logarithm of fixation time plus one (*ln*(TP1)) and fixation count plus one (*ln*(CP1)) to enhance the normality of residuals. TP1 denotes "time plus one," while CP1 represents "count plus one." The model included six variables, with participant variability treated as a random effect. Fixed effects included the sequence of scenarios experienced, drone flight configurations, their interaction, and a constant term. A Type III ANOVA was used to evaluate the statistical significance of each fixed effect, particularly the influence of drone flight factors. Results were considered statistically significant if the p-value was less than 0.05. All analyses were performed using the lme4 and lmerTest packages in R [34], [35].

5 Preliminary Results and Discussion

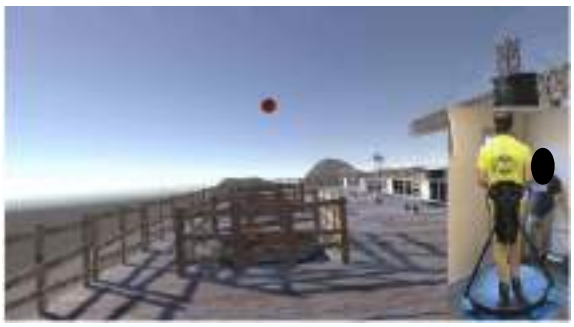


Figure 3. A sample of a construction worker doing our experiment using a VR headset and an ODT.

So far, a total of 17 test participants (12 males and 5 females) were invited, including construction interns, professionals, managers, and workers. Figure 3 illustrates the first-person perspective of a construction worker, represented as our test participant, performing a material handling task within a virtual construction environment. The red dot in the figure indicates the participant's eye gaze point.

From the preliminary results, there were three findings. First, Figure 4 shows the participants' 95 percent confidence intervals for fixation time on danger areas in all 7 different scenes. In the horizontal axis, "F" means the drone flying from the participants' front, and "B" means the drone flying from their back. The number after the latter means the drone flying heights in feet. Since the fixation time in no drone scene is significantly higher than in all the other scenes, drones reduced participants' fixation time on danger areas. Second, Figure 5 compares the participants' 95 percent confidence intervals for fixation time on danger areas in different drone flying heights. The fixation time on danger areas of a drone flying at 82 feet is much higher than a drone flying at 16 feet. Drones flying at lower heights (16 feet) significantly reduced fixation time on danger areas. Figure 6 shows the comparison of the 95 percent confidence intervals for fixation time on the sky in different drone flying directions. The presence of drones approaching from the front results in a significant increase in fixation time on the sky. Overall, the findings suggest that participants experienced greater distraction when the drone approached from the front below 48 feet.

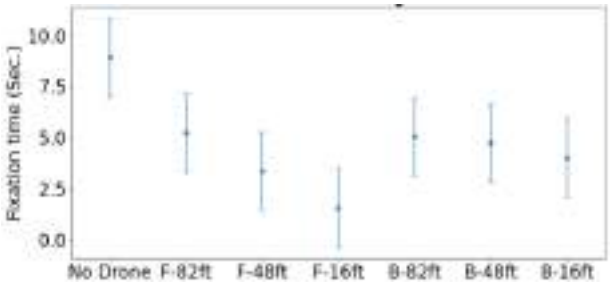


Figure 4. Fixation time on danger areas in different scenes.

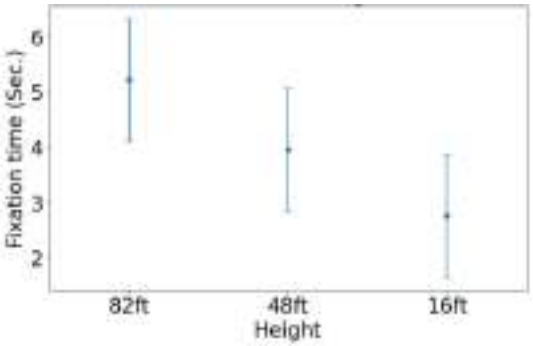


Figure 5. Fixation time on danger areas in different flying heights.

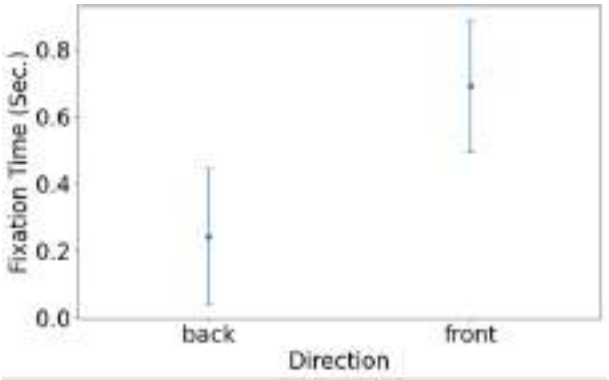


Figure 6. Fixation time on the sky in different flying directions.

6 Discussion

The findings of this study contributed to understanding the influence of drone flight height and direction on the visual attention of construction workers. Specifically, our results indicated that participants experienced greater distraction when drones approached from the front, particularly at lower altitudes (16 ft and 48 ft). These findings aligned with prior research demonstrating that drone interventions can disrupt workers' visual focus, while also extending the literature by examining how specific drone conditions modulate the degree of distraction.

When drones approached at heights of 16 feet and 48 feet, the participants' fixation time on danger areas significantly decreased, potentially impairing their ability to maintain attention on hazardous zones. This has critical implications for construction site safety, as diminished focus on danger areas increases the risk of accidents such as slips, falls, and trips. Furthermore, post-experiment interviews revealed that half of the participants reported feeling distracted by the presence of drones, reinforcing the need to address drone-related safety concerns in dynamic construction environments.

Despite these important contributions, this study has few limitations. First, the experiment only used a single type of virtual drone in the experiment. Drones differ in attributes such as sound, size, and intended purpose, all of which could influence distraction levels. For instance, larger drones with louder noise might induce greater distraction compared to smaller, quieter drones. Future research should investigate the effects of various drone types to assess their specific contributions to worker distraction.

Second, the study focused on a material handling task. However, construction work encompasses a wide range of activities, such as operating vehicles or heavy machinery, where the consequences of drone-induced distractions may vary. For example, the impact of distractions on excavator operators or truck drivers may differ significantly from their effects on workers performing material handling tasks. Future studies should explore a broader array of construction tasks to provide a more comprehensive understanding of how drone interventions affect worker safety across diverse job roles.

Third, to maintain experimental control and isolate the effects of drone distractions, we designed the virtual environment with a simplified construction setting. A more complex and dynamic environment could introduce additional variables, such as moving equipment or other workers, which may confound the results. By minimizing unnecessary distractions, we ensured that participants' visual attention responses were primarily influenced by drone presence rather than other environmental factors. However, real construction sites might be more complex, and future studies should incorporate more realistic elements to enhance the ecological validity of the findings.

7 Conclusion

This study provided valuable insights into how drone flight configurations affect construction workers' visual attention, particularly their focus on hazard areas. Using a VR-based experiment, the research showed that drones significantly distract workers, especially when flying at lower heights (16 ft and 48 ft) or approaching from the front. These conditions reduced fixation times on danger areas, impairing hazard recognition and increasing accident risks. The results underscores the need for carefully planned drone operations and defined safety parameters in construction environments.

Despite its contributions, this study has several limitations that highlight opportunities for future research. The experiment utilized a single type of virtual drone and focused solely on material handling tasks, limiting the generalizability of the findings to other drone models and construction activities. Additionally, to isolate the effects of drone distraction under different flight configurations, the virtual environment was intentionally simplified, excluding other potential distractions such as moving machinery and additional workers. Future studies should examine how different drone attributes impact workers across various construction tasks within more complex and dynamic site conditions to enhance the applicability of the findings.

The findings highlight the need for safety protocols to mitigate drone-induced distractions, such as restricting drone altitude, flight paths, and directions, especially in high-risk areas. While workers may adapt to drones over time, new or inexperienced workers may still face distraction risks from drone. These insights can inform training programs that expose workers to drone distractions before they start to work on a real

construction site. Which can help them adapt in advance and reduce the risk of safety incidents from unexpected attention shifts.

8 Data Statement

All data collected from our test participants during the study are proprietary or confidential in nature and may only be provided with restrictions. Specifically, the de-identified data can only be requested by the National Institute for Occupational Safety and Health, UW-Madison regulatory and research oversight boards and offices. Some or all code generated or used during the study are available from the corresponding author by request.

9 Acknowledgments

This research was supported by the National Institute for Occupational Safety and Health (NIOSH), and the authors would like to express their gratitude for NIOSH's valuable contribution. The research was also supported by projects 1R21OH012455-01A1 and 5R21OH012455-02. The findings and conclusions presented in this paper are solely those of the authors and do not necessarily reflect the views of NIOSH. The authors are also grateful to M. A. Mortenson Company and their workers for participating in our experiment. Any mention of specific companies or products is for informational purposes only and does not imply endorsement by NIOSH, Mortenson, or the authors.

References

[1] John Howard, Vladimir Murashov, and Christine Branche, "Can Drones Make Construction Safer?" Accessed: Dec. 03, 2023. [Online]. Available: https://blogs-origin.cdc.gov/niosh-science-blog/2017/10/23/drones-construction/

[2] I. Jeelani and M. Gheisari, "Safety challenges of UAV integration in construction: Conceptual analysis and future research roadmap," *Saf Sci*, vol. 144, p. 105473, 2021, doi: https://doi.org/10.1016/j.ssci.2021.105473.

[3] The Wall Street Journal, "Drones Are Watching: Railroad Irks Workers With Unmanned Aircraft." Accessed: Dec. 03, 2023. [Online]. Available: https://www.wsj.com/articles/drones-are-watching-railroad-irks-workers-with-unmanned-aircraft-1521028800

[4] G. Mark, V. M. González, and J. Harris, "No task left behind?: examining the nature of fragmented work," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2005, [Online]. Available: https://api.semanticscholar.org/CorpusID:3079084

[5] M. Khalid, M. Namian, and C. C. Massarra, "The Dark Side of the Drones: A Review of Emerging Safety Implications in Construction," 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:237875604

[6] N. Mostafa, A. Alex, and F. Jing, "Effect of Distraction on Hazard Recognition and Safety Risk Perception," *J Constr Eng Manag*, vol. 144, no. 4, p. 04018008, Apr. 2018, doi: 10.1061/(ASCE)CO.1943-7862.0001459.

[7] E. M. Altmann, G. Trafton, and D. Z. Hambrick, "Momentary interruptions can derail the train of thought.," *J Exp Psychol Gen*, vol. 143 1, pp. 215–26, 2014, [Online]. Available: https://api.semanticscholar.org/CorpusID:15459133

[8] U.S. Bureau of Labor Statistics, "A look at falls, slips, and trips in the construction industry : The Economics Daily: U.S. Bureau of Labor Statistics," May 06, 2024. [Online]. Available: https://www.bls.gov/opub/ted/2024/a-look-at-falls-slips-and-trips-in-the-construction-industry.htm

[9] EHS Daily Advisor, "Understanding How Human Factors Affect Slips, Trips, and Falls." Accessed: Dec. 03, 2023. [Online]. Available: https://ehsdailyadvisor.blr.com/infographic/infographic-understanding-how-human-factors-affect-slips-trips-and-falls/

[10] M. Namian, M. Khalid, G. Wang, and Y. Turkan, "Revealing Safety Risks of Unmanned Aerial Vehicles in Construction," *Transp Res Rec*, vol. 2675, no. 11, pp. 334–347, Jul. 2021, doi: 10.1177/03611981211017134.

[11] M. C. Tatum and J. Liu, "Unmanned Aircraft System Applications in Construction," *Procedia Eng*, vol. 196, pp. 167–175, 2017, doi: https://doi.org/10.1016/j.proeng.2017.07.187.

[12] G. Albeaino, P. Brophy, I. Jeelani, M. Gheisari, and R. R. A. Issa, "Impact of Drone Presence on Construction Individuals Working at Heights," *J Constr Eng Manag*, vol. 149, no. 11, p. 4023119, 2023, doi: 10.1061/JCEMD4.COENG-13861.

[13] Wingtra, "Why and How to Use Drones in Construction and Infrastructure," 2020. [Online]. Available: https://wingtra.com/drone-mapping-applications/drones-in-construction-and-infrastructure/

[14] W. Jiang, Y. Zhou, L. Ding, C. Zhou, and X. Ning, "UAV-based 3D reconstruction for hoist site mapping and layout planning in

petrochemical construction," *Autom Constr*, 2020, doi: 10.1016/j.autcon.2020.103137.

[15] S. Sun, "A new method for monitoring machinery movement using an Unmanned Aerial Vehicle (UAV) system," University of Twente, 2019. [Online]. Available: http://essay.utwente.nl/79827/

[16] R. R. S. de Melo, D. B. Costa, J. S. Álvares, and J. Irizarry, "Applicability of unmanned aerial system (UAS) for safety inspection on construction sites," *Saf Sci*, 2017, doi: 10.1016/j.ssci.2017.06.008.

[17] Federal Aviation Administration (FAA), "Certificated Remote Pilots including Commercial Operators." Accessed: Dec. 04, 2023. [Online]. Available: https://www.faa.gov/uas/commercial_operators

[18] Federal Aviation Administration (FAA), "Operations Over People General Overview." Accessed: Dec. 04, 2023. [Online]. Available: https://www.faa.gov/uas/commercial_operators/operations_over_people

[19] J. Fan and M. Saadeghvaziri, *Applications of Drones in Infrastructures: Challenges and Opportunities*. 2019. doi: 10.5281/zenodo.3566281.

[20] D. B. Costa, R. R. S. de Melo, J. S. Álvares, and A. A. Bello, "Evaluating the Performance of Unmanned Aerial Vehicles for Safety Inspection," in *24th Annual Conference of the International Group for Lean Construction*, Boston, Massachusetts, USA, 2016. [Online]. Available: http://iglc.net/Papers/Details/1286/pdf

[21] I. M. Hashem, F. Ian, Z. Xun, A. Behzad, R. Peyman, and M. Michelle, "Quantitative Assessment of Proximity Risks Associated with Unmanned Aerial Vehicles in Construction," *Journal of Management in Engineering*, vol. 37, no. 1, p. 04020095, Jan. 2021, doi: 10.1061/(ASCE)ME.1943-5479.0000852.

[22] H. Izadi Moud, A. Shojaei, I. Flood, X. Zhang, and M. Hatami, *Qualitative and Quantitative Risk Analysis of Unmanned Aerial Vehicle Flights over Construction Job Sites*. 2018.

[23] H. Xie, E. Tudoreanu, and W. Shi, "Development of a virtual reality safety-training system for construction workers," *Digital library of construction informatics and information technology in civil engineering and construction*, pp. 1–9, 2006.

[24] J. Plomin, P. Schweidler, and A. Oehme, "Virtual reality check: a comparison of virtual reality, screen-based, and real world settings as research methods for HRI," *Front Robot AI*, vol. 10, p. 1156715, 2023.

[25] Z. Barlow, H. Jashami, A. Sova, D. S. Hurwitz, and M. J. Olsen, "Policy processes and recommendations for Unmanned Aerial System operations near roadways based on visual attention of drivers," *Transp Res Part C Emerg Technol*, vol. 108, pp. 207–222, 2019, doi: https://doi.org/10.1016/j.trc.2019.09.012.

[26] D. S. Hurwitz, M. J. Olsen, Z. Barlow, and others, "Driving Distraction Due to Drones," 2018.

[27] A. Ryan, C. Fitzpatrick, E. Christofa, and M. Knodler, "Driver performance due to small unmanned aerial system applications in the vicinity of roadways," *Transp Res Part F Traffic Psychol Behav*, vol. 68, pp. 118–131, 2020, doi: https://doi.org/10.1016/j.trf.2019.12.006.

[28] I. Jeelani and M. Gheisari, "Safety Challenges of UAV Integration in the Construction Industry: Focusing on Workers at Height," 2022.

[29] Gilles Albeaino, Idris Jeelani, Masoud Gheisari, and Raja R. A. Issa, "Attentional Impact of Drones on Construction Sites," *Proceedings of 59th Annual Associated Schools*, vol. 4, pp. 121–129, 2023.

[30] G. Charness, U. Gneezy, and M. A. Kuhn, "Experimental methods: Between-subject and within-subject design," *J Econ Behav Organ*, vol. 81, no. 1, pp. 1–8, 2012, doi: https://doi.org/10.1016/j.jebo.2011.08.009.

[31] G. Morgenthal and N. Hallermann, "Quality Assessment of Unmanned Aerial Vehicle (UAV) Based Visual Inspection of Structures," *Advances in Structural Engineering*, vol. 17, no. 3, pp. 289–302, Mar. 2014, doi: 10.1260/1369-4332.17.3.289.

[32] Saul Mcleod, "Experimental Design." Accessed: Dec. 04, 2023. [Online]. Available: https://www.simplypsychology.org/experimental-designs.html#:~:text=Repeated%20Measures%20design%20is%20an,%2C%20or%20within%2Dsubjects%20design

[33] D. A. Preece, "Chapter 10 - Latin Squares as Experimental Designs," in *Annals of Discrete Mathematics*, vol. 46, J. Dénes and A. D. Keedwell, Eds., Elsevier, 1991, pp. 317–342. doi: https://doi.org/10.1016/S0167-5060(08)70970-1.

[34] D. Bates, "Fitting linear mixed-effects models using lme4," *arXiv preprint arXiv:1406.5823*, 2014.

[35] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, "lmerTest package: tests in linear mixed effects models," *J Stat Softw*, vol. 82, no. 13, 2017.

Cognitive Load-Aware Virtual Reality Training Platform for Safe Drone Operations in Construction

Yuming Zhang[1] and Houtan Jebelli[1]

[1] Department of Civil and Environmental Engineering, University of Illinois Urbana-Campaign, Champaign, IL, USA
yumingz4@illinois.edu, hjebelli@illinois.edu

Abstract –

Drones have become indispensable in construction projects, performing tasks such as site surveying, 3D mapping, progress monitoring, and inspecting hazardous or hard-to-reach areas. Their integration has significantly enhanced efficiency, precision, and safety. However, effective human-drone interaction remains challenging due to the complexity of construction environments and the necessity for precise drone control, which requires workers to undergo extensive training. Traditional training programs primarily focus on task-specific skills, often neglecting the influence of workers' cognitive and physiological states on learning effectiveness and performance. This oversight can lead to increased cognitive fatigue, reduced learning outcomes, and heightened safety risks. To address these limitations, this study introduces a Virtual Reality (VR)-based adaptive training platform designed to optimize worker-drone interactions by monitoring cognitive load through physiological indicators such as heart rate and skin conductance. Utilizing a Long Short-Term Memory (LSTM) model, the platform accurately predicts cognitive load in real time and dynamically adjusts the training pace to align with individual cognitive states. In experimental evaluations, the adaptive training platform achieved a cognitive load prediction accuracy of 92% and reduced task completion time by 69.9%, outperforming non-adaptive training methods by 27.6% in worker-drone interaction scenarios. By personalizing the training experience based on real-time cognitive assessments, the platform enhances worker preparedness, promoting safer and more efficient drone operations on construction sites. This research underscores the critical role of integrating physiological data into adaptive training systems to meet the multifaceted demands of modern construction environments.

Keywords –

Virtual Reality; Human-Drone Interaction; Physiological signals; Adaptive training

1 Introduction

Drones, or unmanned aerial vehicles (UAVs), have revolutionized the construction industry by performing a wide array of tasks such as site surveying, 3D mapping, progress monitoring, and inspecting hazardous or hard-to-reach areas. Equipped with advanced sensors and high-resolution cameras, drones provide precise data collection from multiple angles, offering invaluable insights into onsite conditions [1]. Their versatility and adaptability to diverse environments have led to widespread adoption, transforming project management and execution into more efficient and innovative processes [2]. Studies indicate that incorporating drones into construction workflows can boost productivity by up to 94.48%, accelerate operations by 35.41%, and reduce costs by as much as 78% compared to traditional methods [3]. These improvements underscore the significant role drones play in enhancing efficiency, precision, and safety on construction sites [4].

Despite advancements in drone automation, effective deployment in construction settings still necessitates substantial human oversight. Construction sites are inherently dynamic and unpredictable, with constantly changing conditions such as weather variations, movement of workers and equipment, and shifting project requirements. While drones can autonomously execute predefined tasks, they often lack the flexibility to adapt to real-time, unforeseen changes. Complex operations like navigating confined spaces or conducting detailed inspections demand precise, context-specific control that only skilled human operators can provide [5]. This active human-drone interaction ensures that drone operations align with project objectives and adapt to the specific needs of the construction site.

The complexity of drone operation, encompassing both hardware and software components, requires operators to possess a robust foundation of theoretical knowledge and practical skills [6]. Moreover, the unpredictable nature of construction environments necessitates that drone operators are highly adaptable and capable of making swift, informed decisions to

effectively address real-time challenges [7]. Consequently, comprehensive training programs are essential to prepare workers for the multifaceted demands of drone operation in construction settings. These programs aim to equip workers with the ability to operate drones safely and effectively under a variety of real-world conditions by training them to process large volumes of situational and operational information, make complex decisions, and maintain sustained concentration.

However, such intensive training requirements can significantly increase the cognitive demands on workers [8]. Cognitive load, defined as the mental effort required to process information and perform tasks, is influenced by both task complexity and individual capacity. Excessive cognitive load can impair workers' ability to process information efficiently, perform tasks accurately, and ultimately diminish the overall effectiveness of the learning process [9]. Current training systems for construction workers typically emphasize the development of task-specific skills, such as maneuvering drones or performing specialized construction-related operations [5]. While these systems have proven effective in skill acquisition for specific tasks, they often inadequately address fluctuations in cognitive load, which can significantly impact learning efficiency and task performance.

To address these limitations, this study develops and evaluates a Virtual Reality (VR)-based adaptive training platform designed to enhance worker-drone collaboration in construction environments. The system utilizes real-time physiological data, including heart rate and skin conductance, to predict users' cognitive load and dynamically adjust the complexity and frequency of training tasks accordingly. By monitoring cognitive load through physiological inputs, the platform can tailor the training experience to individual cognitive states, ensuring that workers are neither overwhelmed nor under-challenged during training. The ultimate goal of this research is to enhance worker performance and ensure safer, more efficient human-drone collaboration on construction sites. By leveraging adaptive immersive technologies, the proposed system addresses the challenges posed by dynamic construction environments, equipping workers with the necessary skills and adaptability to effectively integrate drones into modern workflows and meet the demands of increasingly complex construction projects.

2 The Multifunctional Role of Drones in Transforming the Construction Landscape

Drones, or Unmanned Aerial Vehicles (UAVs), are advanced aircraft systems designed to operate without an onboard human pilot. Equipped with advanced sensors, high-definition cameras, and GPS systems, drones can perform a variety of tasks with remarkable precision, such as capturing real-time data, generating accurate 3D models, and conducting remote inspections [10]. These features have established drones as versatile tools across multiple industries. In construction, preliminary studies and practical applications have demonstrated their potential to offer significant improvements in efficiency, precision, and safety compared to traditional methods [11]. For instance, drones are increasingly utilized in site surveys, mapping, progress monitoring, and safety inspections, indicating their growing relevance in addressing the complex demands of modern construction projects [12].

However, the integration of drones into construction workflows presents significant challenges While drones possess autonomous capabilities for executing predefined tasks, the dynamic and unpredictable nature of construction environments poses substantial safety risks. For instance, simulations recorded 773 collisions over 230 runs under autonomous control [13], underscoring the critical need for skilled operators to provide remote supervision and control. Studies have shown that worker operators trained in navigating complex terrain, managing sophisticated drone functionalities, and adapting to evolving site conditions reduce operational errors, significantly improving safety and effectiveness in drone operations [5]. This underscores the critical importance of comprehensive training programs aimed at enhancing worker-drone cooperation. Such training must encompass both theoretical understanding and practical application, emphasizing drone operation in complex and dynamic environments[14].

However, the inherent complexity of these training requirements can lead to fluctuations in cognitive load, which may affect training outcome [15] s. Current training programs for drone operation in construction often prioritize the development of technical skills, such as basic navigation and task-specific maneuvers [16]. While these foundational competencies are essential, such programs frequently place less emphasis on the cognitive challenges that operators face during the training process. Cognitive fluctuations are common when operators process complex information, multitask, and make real-time decisions, which, if unmanaged, may compromise their ability to make decisions and execute tasks effectively [17]. These limitations highlight the necessity for more advanced, adaptive training systems that not only equip operators with technical proficiency but also enhance their capacity to navigate the complexities and cognitive demands of drone operation in dynamic and unpredictable construction environments.

3 Impact of Cognitive Load on Training for Enhanced Drone Operation

Cognitive load, defined as the mental effort required to process information and perform tasks, plays an important role in influencing the efficiency and accuracy of drone operators, particularly within the dynamic and high-stakes environment of construction [19] . High cognitive load refers to a state where the mental demands of a task exceed an individual's capacity to process information effectively, leading to reduced performance and increased likelihood of errors [18]. Drone operators often need to process complex spatial, technical, and safety-related information in real time, which can amplify cognitive demands [6]. In construction settings, operators often navigate intricate and constantly changing sites, execute precise maneuvers, and respond to unforeseen challenges such as equipment malfunctions, changing environmental conditions, or sudden safety hazards [15]. Excessive cognitive load has been observed during drone operations, where it negatively impacts human performance by inducing mental fatigue, impairing situational awareness, and compromising decision-making processes [19]. Mental overload can hinder the operator's ability to process critical information, leading to reduced efficiency, operational delays, and increased risks of accidents or safety violations. As construction projects increasingly rely on drones for tasks such as surveying, monitoring, and inspection, understanding and mitigating cognitive load becomes essential for enhancing overall system performance and safety.

However, despite its importance, existing literature on drone operation training for construction lacks a comprehensive understanding of how cognitive load influences learning outcomes and task performance [4]. Most current training programs overlook the dynamic variations in cognitive load experienced by users during drone interaction training, failing to account for individual differences and task complexity. This generic training methodology risks overloading operators with excessive cognitive demands, potentially reducing training effectiveness, impairing task performance, and increasing the likelihood of operational errors [20]. Such an approach fails to prepare operators for the cognitive demands posed by real-world construction scenarios, where task complexity and situational variability are significant.

To address this gap, there is an urgent need for an adaptive training environment that dynamically adjusts task complexity based on the trainee's cognitive load. By incorporating real-time cognitive load assessment into training frameworks, task difficulty can be tailored to individual capabilities, fostering optimal learning conditions. Adaptive training has the potential to enhance learning retention, improve task performance, and better prepare operators to handle the cognitive challenges of drone operation in construction. Such a system would ensure that trainees are neither overwhelmed nor insufficiently challenged, thus promoting a balanced cognitive load that facilitates skill acquisition and long-term proficiency.

4 Method: Adaptive Human-Drone Training Platform Based on Cognitive Load For Construction Workers

This study designed and implemented an LSTM-driven adaptive training system for drone operation in a VR environment (Figure 1). Physiological signals for cognitive load assessment are processed using a preprocessing pipeline to enhance signal quality and consistency. The pre-processed physiological data is then fed into the LSTM-based adaptive training system for cognitive load prediction. Finally, the adaptive system is integrated into the VR-based training program to provide users with a dynamic and immersive simulated training environment.

4.1 Physiological signals for cognitive load assessment and data preprocessing

To assess cognitive load during training, this study utilizes two widely recognized physiological signals: heart rate variability (HRV) and galvanic skin response (GSR) [15], [21]. HRV, derived from photoplethysmography (PPG) [22], measures the variation in intervals between consecutive heartbeats, with low HRV indicating a potential state of high cognitive load. GSR, obtained from electrodermal activity (EDA), indicates changes in skin conductivity driven by sweat gland activity, with higher GSR values often associated with emotional arousal or stress. These physiological signals are continuously recorded during the training process to provide an objective assessment of the user's cognitive load. The collected data undergoes preprocessing to standardize and smooth the signals, ensuring consistency and reliability. PPG data were filtered using a fourth-order Chebyshev II filter with a passband range of 0.5–5 Hz and a stopband attenuation of 30 dB to minimize motion artifacts while retaining physiologically relevant HRV components. To align sampling frequencies across signals, EDA data were upsampled using linear interpolation to match the PPG data. Subsequently, EDA signals were processed with a third-order Butterworth low-pass filter with a cutoff frequency of 0.5 Hz.

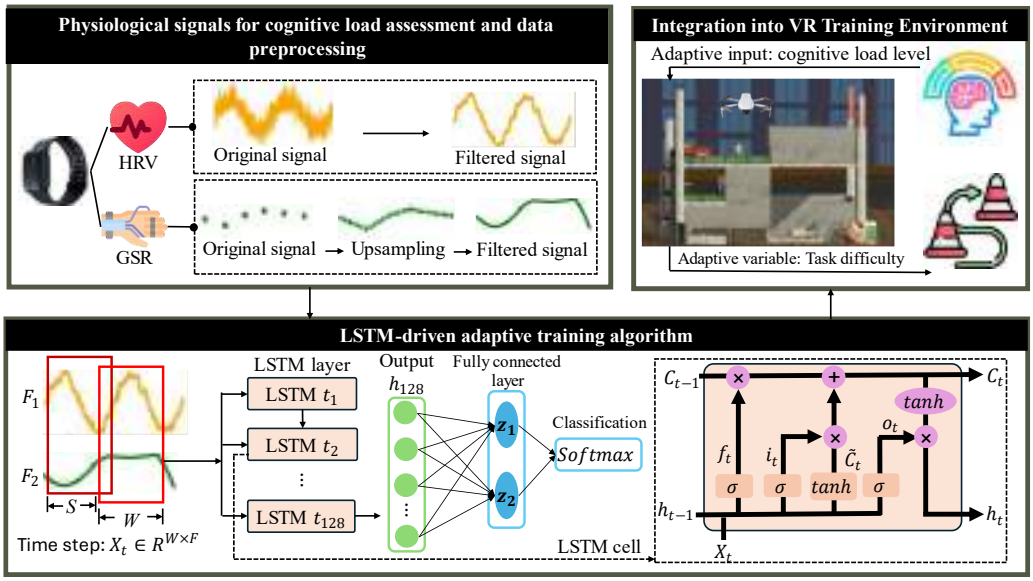


Figure 1. Overview of LSTM-driven adaptive VR training system

4.2 LSTM-driven adaptive training algorithm

The LSTM-driven adaptive training system is designed to dynamically adjust training tasks in real time based on cognitive load predictions. The goal is to maintain users within an optimal cognitive load range, balancing task difficulty and cognitive engagement for effective learning outcomes. The system is underpinned by an LSTM-based cognitive load prediction model, which processes sequential physiological signals to predict cognitive load levels. LSTM networks are particularly suitable for this task due to their ability to capture long-term temporal dependencies while selectively retaining or forgetting information through gating mechanisms.

To train the LSTM model, ground truth cognitive load labels were established by engaging. participants in tasks of systematically increasing complexity. Physiological signals, including HRV derived from PPG and GSR from EDA, were continuously recorded. Ground truth labels were determined using the NASA-TLX, a validated subjective workload assessment tool. Participants rated their perceived workload and aggregated the score into composite scores S_i , normalized as $S_i^{norm} = \frac{S_i - S_{min}}{S_{max} - S_{min}}$, where S_{min} and S_{max} are the dataset's minimum and maximum scores. The normalized scores were discretized into two cognitive load categories, "low" and "high," to serve as the ground truth labels for cognitive load classification.

Pre-processed HRV and GSR data were segmented into 10-second sliding windows with a 3-second update interval for cognitive load monitoring. The data were organized into a feature matrix $X \in R^{W\times F}$, where W is the sequence length and F is the feature dimension. The feature matrix was used as input to the LSTM-driven adaptive algorithm network. The hidden layer followed by the input layer contains 128 units. Each hidden unit processes the input sequence X by maintaining a cell state c_t and producing a hidden state h_t at each time step t. The dynamics of the LSTM are governed by its gating mechanisms, including the input gate i_t, forget gate f_t, output gate o_t, and the candidate cell state $\tilde{c_t}$. These are computed as follows [23]

$$f_t = \sigma\left(W_f[h_{t-1}, x_t] + b_f\right) \quad (1)$$
$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (2)$$
$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (3)$$
$$\tilde{c}_t = \tanh\left(W_c[h_{t-1}, x_t] + b_c\right) \quad (4)$$
$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (5)$$
$$h_t = o_t \odot \tanh\left(c_t\right) \quad (6)$$

where W_f, W_i, W_o, W_c and b_f, b_i, b_o, b_c are the weight matrices and biases for the respective gates, σ represents the sigmoid activation function, $tanh$is the hyperbolic tangent function, and $\odot$ denotes the element-wise multiplication. The final hidden state h_{128}, capturing the temporal dependencies in the input sequence, is passed through a fully connected dense layer where a linear transformation is applied, given by $z = W_z h_{128} + b_z$, where W_z and b_z are the weight matrix and bias for the dense layer, and $z = [z_1, z_2, z_3]$ represents the outputs corresponding to each class. Subsequently, the output z is passed through a softmax activation function to predict the probability distribution over cognitive load levels $\hat{y} = \frac{e^{z_k}}{\sum_{i=1}^{C} e^{z_i}}, k = 1,2,3$, where $\hat{y}$ is the probability distribution over the cognitive load classes, and C corresponds to the number of cognitive load classes.

4.3 Integration into VR Training Environment

A virtual construction environment was developed to simulate realistic and dynamic construction site conditions, providing a platform for users to learn and practice drone operation. The environment was designed to replicate the complexities of high-rise construction sites, incorporating detailed architectural structures such as multi-level buildings, scaffolding, and construction equipment. Dynamic objects, including moving machines such as cranes, excavators, and forklifts, were programmed to operate autonomously, simulating real-time activities. Additionally, virtual workers were included, designed to mimic human movements and interactions within the construction site, creating a realistic sense of activity and unpredictability. To further enhance the fidelity of the environment, Rigidbody components are added to objects and collision detection was incorporated for drone collision error detection. A dual-view setup, featuring both a drone monitor and a drone's first-person view, is provided to enhance situational awareness and support precise navigation.

The LSTM-driven adaptive training system was integrated into the virtual construction environment to adjust task difficulty based on real-time cognitive load predictions. As cognitive load levels fluctuate, the system dynamically modifies elements such as drone flight paths (the number of waypoints n and the distance between waypoints d) and the behaviour of dynamic obstacles (the velocity of dynamic obstacles $v_j(t)$, and their trajectory complexity $r_j(t)$).

When a high cognitive load is detected ($\hat{y} = 3$), the system adjusts to prevent cognitive overload and enhance learning efficiency. Task complexity is reduced by simplifying drone flight paths (reducing n and increasing d), minimizing the number and movement speed of dynamic obstacles $v_j(t)$, and decreasing their trajectory complexity $r_j(t)$. Additionally, the system provides step-by-step guidance or visual cues to assist in task execution. If high cognitive load persists, short rest intervals are introduced to allow cognitive recovery before resuming training. Conversely, when a low cognitive load is detected ($\hat{y} = 1$), the system gradually increases task difficulty to ensure continued engagement and skill development. This is achieved by introducing more intricate drone flight paths (increasing n and decreasing d), introducing Gaussian noise to the velocity of dynamic obstacles ($v_j(t) \sim v_j(t) + N(0, \sigma^2)$), and selecting more complex trajectories $r_j(t)$ that simulate real-world challenges such as wind disturbances. For moderate cognitive load ($\hat{y} = 2$), the system evaluates task completion time (T) as an additional parameter. If $T > T_{threshold}$, task difficulty remains unchanged to allow further adaptation. If $T < T_{threshold}$, task complexity is incrementally increased to maintain an optimal level of challenge without inducing excessive cognitive strain.

5 Case Study: Evaluating the Effectiveness of Cognitive-Based Adaptive VR Training for Construction Worker-Drone Interaction

The experiment was designed to evaluate the effectiveness of the LSTM-driven adaptive training system for drone operation in a VR environment. The virtual environment was developed using Unity and code with Visual Studio with C# and deployed on the SteamVR platform, providing a dynamic and immersive training scenario. Physiological signals including HRV and GSR were measured using the Empatica E4 wristband. An IRB-approved study was conducted, with all participants providing informed consent before data collection. 20 Participants were divided into two groups: one group trained with the adaptive system, while the other followed a non-adaptive VR training protocol for comparison. Each participant underwent multiple training sessions consisting of tasks with systematically varying levels of difficulty. For the adaptive training group (ATG), the task difficulty was dynamically adjusted in real time based on cognitive load predictions generated by the LSTM model. In contrast, the non-adaptive training group (NATG) followed a fixed sequence of tasks with predetermined difficulty levels. During the training sessions, physiological signals, including HRV and GSR of ATG, were continuously recorded to monitor cognitive load. The adaptive system used these signals to adjust task parameters dynamically, aiming to maintain participants within an optimal cognitive load range.

Performance metrics were evaluated using both cognitive load prediction accuracy and task performance indicators. To assess the accuracy of the LSTM-based cognitive load prediction model, an initial validation study was conducted in which participants completed a series of training sessions with progressively increasing difficulty levels. After each session, participants provided self-reported ratings of their perceived cognitive load using the NASA Task Load Index (NASA-TLX), a widely recognized tool for workload assessment. These self-assessments provided a baseline for comparing the model's predictions.

To evaluate the effectiveness of the adaptive training system, pre- and post-training assessments measured task completion times as an indicator of operational efficiency. Task completion time was chosen as an objective metric to quantify the training's impact and facilitate a comparative analysis between adaptive and standard training methodologies. By integrating cognitive load prediction accuracy with performance improvements, the

experiment aimed to validate the hypothesis that adaptive training not only enhances learning outcomes but also significantly improves task efficiency.

6 Results

6.1 Accuracy of the LSTM Model

The LSTM-based cognitive load prediction model demonstrated high accuracy in distinguishing between the three cognitive load levels, achieving an overall accuracy of 92%. Key performance metrics, including precision, recall, and F1-score for each class, are summarized in Table 1. It highlights that the model achieved perfect precision, recall, and F1-score for Class 1. Similarly, Class 3 demonstrated strong performance with an F1-score of 0.92. However, the recall for Class 2 was slightly lower at 0.78, indicating minor misclassification challenges for moderate cognitive load levels.

Table 1 Key Metrics for Cognitive Load Prediction

	Class 1	Class 2	Class 3
Precision	1.00	1.00	0.85
Recall	1.00	0.78	1.00
F1-Score	1.00	0.88	0.92

Figure 2 presents a confusion matrix heatmap, where the color intensity corresponds to the recall percentage for each class, normalized by row totals. Darker shades highlight higher recall values, emphasizing areas of strong predictive accuracy, while lighter shades indicate potential misclassifications or areas requiring further evaluation. The cell annotations display the absolute counts of predicted and actual classifications, providing a detailed overview of model performance.

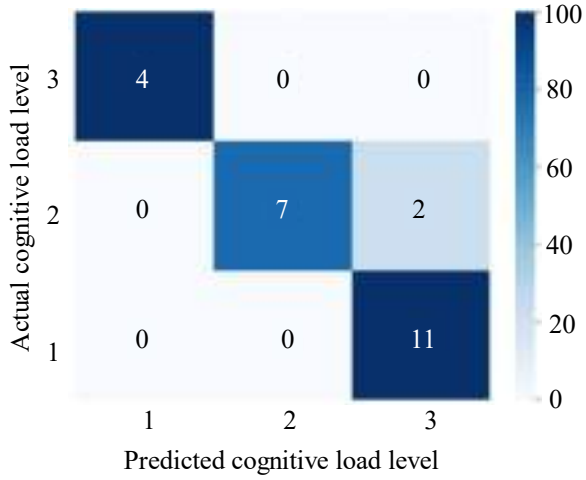


Figure 2. Cognitive Load Prediction Accuracy of LSTM-Based Adaptive Training System

6.2 Task Completion Time

Participants demonstrated significant improvements in task performance following training as Figure 3 shows. In the ATG group, the pre-training task completion time was 266.8 ± 27.5 seconds (mean ± SD), while the post-training time was reduced to 81.6 ± 11.5 seconds. This represents an average reduction of 69.4% in task completion time compared to the pre-training baseline. Similarly, in the NATG group, the pre-training task completion time was 262.1 ± 38.7 seconds, and the post-training time decreased to 110.0 ± 34.0 seconds, reflecting an average reduction of 58%. Furthermore, participants in the ATG group outperformed those in the NATG group, completing tasks 25.8% faster on average during the post-training assessments. These results highlight the effectiveness of the adaptive training system in enhancing task performance and improving operational efficiency by reducing task completion time and maintaining participants within an optimal cognitive load range during training.

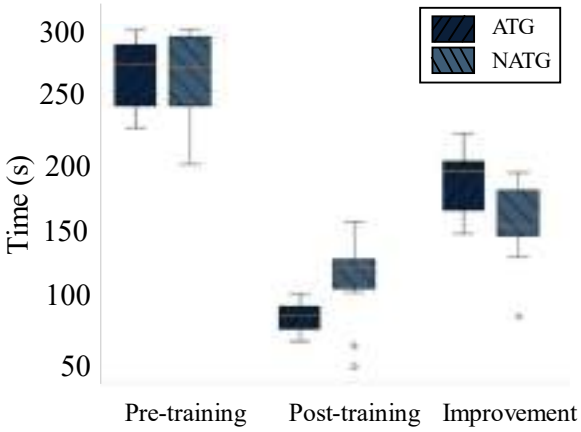


Figure 3. Task Completion Time for ATG and NATG

7 Conclusion

This research successfully developed a cognitive load-based adaptive training system to enhance drone operation training for construction workers. By leveraging physiological indicators, specifically HRV and GSR, the system effectively monitored user's cognitive load and dynamically adjusted training complexity through a LSTM neural network. The adaptive training system significantly outperformed traditional non-adaptive training methods, with participants using the adaptive system completing tasks 69.9% faster compared to their pre-training performance, while the non-adaptive group saw a 58.1% reduction in

task completion time. Additionally, the LSTM-based cognitive load prediction model achieved an impressive 92% accuracy, demonstrating its reliability in assessing cognitive states.

The system's ability to adjust training complexity in real-time ensured that trainees were neither overwhelmed nor under-stimulated, leading to a more engaging and efficient learning experience that fostered better skill acquisition. Additionally, this dynamic approach inherently exposes trainees to varying levels of cognitive demands, helping them develop the ability to operate under pressure and adapt to unpredictable conditions. By adjusting task complexity and introducing real-time environmental variability, the training framework fosters resilience and stress management skills, better preparing workers for real-world drone operations. This personalized approach mitigates cognitive fatigue and improves overall task performance, contributing to more effective human-drone collaboration.

This study has certain limitations that present opportunities for future research. The current cognitive load assessment relies solely on HRV and GSR, which may not fully capture all aspects of cognitive strain. Future research will integrate additional physiological indicators, such as electroencephalography (EEG) and eye-tracking data, to enhance prediction accuracy and system adaptability. Additionally, while the LSTM-based model demonstrated strong predictive accuracy, its reliance on sequential dependencies may introduce latency in responding to sudden cognitive state changes, potentially compromising the real-time adaptability of the training system. Future research exploring hybrid architectures, such as integrating LSTM with attention mechanisms, could mitigate these challenges by enhancing responsiveness and robustness in highly dynamic construction training environments.

By integrating real-time physiological monitoring with adaptive learning algorithms, this study addresses critical gaps in existing training methodologies for drone operation in construction. The proposed system not only improves task performance and operational efficiency but also ensures that training is tailored to the cognitive capacities of individual workers. These advancements pave the way for more resilient and effective human-drone collaborations, ultimately contributing to safer and more productive construction practices.

Acknowledgements

The work presented in this paper was supported financially by the National Science Foundation (Award No. DUE- 2402008 'Investigating the Impact of an Immersive VR-based Learning Environment for Learning Human-Robot Collaboration in Construction Robotics Education'. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] N. Ejaz and S. Choudhury, "Computer vision in drone imagery for infrastructure management," *Autom Constr*, vol. 163, p. 105418, Jul. 2024, doi: 10.1016/J.AUTCON.2024.105418.

[2] H. W. Choi, H. J. Kim, S. K. Kim, and W. S. Na, "An Overview of Drone Applications in the Construction Industry," *Drones 2023, Vol. 7, Page 515*, vol. 7, no. 8, p. 515, Aug. 2023, doi: 10.3390/DRONES7080515.

[3] F. Riyanto, Juliastuti, O. Setyandito, and A. Pramudya, "Realtime monitoring study for highway construction using Unmanned Aerial Vehicle (UAV) technology," *IOP Conf Ser Earth Environ Sci*, vol. 729, no. 1, Apr. 2021, doi: 10.1088/1755-1315/729/1/012040.

[4] J. M. Nwaogu, Y. Yang, A. P. C. Chan, and X. Wang, "Enhancing Drone Operator Competency within the Construction Industry: Assessing Training Needs and Roadmap for Skill Development," *Buildings 2024, Vol. 14, Page 1153*, vol. 14, no. 4, p. 1153, Apr. 2024, doi: 10.3390/BUILDINGS14041153.

[5] G. Albeaino, R. Eiris, M. Gheisari, and R. R. Issa, "DroneSim: a VR-based flight training simulator for drone-mediated building inspections," *Construction Innovation*, vol. 22, no. 4, pp. 831–848, Nov. 2022, doi: 10.1108/CI-03-2021-0049/FULL/XML.

[6] I. Jeelani, M. Gheisari, and M. E. Sr Rinker, "Safety Challenges of Human-Drone Interactions on Construction Jobsites," *Automation and Robotics in the Architecture, Engineering, and Construction Industry*, pp. 143–164, 2022, doi: 10.1007/978-3-030-77163-8_7.

[7] M. Szóstak, T. Nowobilski, A. M. Mahamadu, and D. C. Pérez, "Unmanned aerial vehicles in the construction industry - Towards a protocol for safe preparation and flight of drones," *International Journal of Intelligent Unmanned Systems*, vol. 11, no. 2, pp. 296–316, Apr. 2023, doi: 10.1108/IJIUS-05-2022-0063/FULL/PDF.

[8] S. Shayesteh and H. Jebelli, "Enhanced Situational Awareness in Worker-Robot Interaction in Construction: Assessing the Role of Visual Cues," *Construction Research Congress 2022: Health and Safety, Workforce, and Education - Selected Papers from Construction Research Congress 2022*, vol. 4-D,

pp. 422–430, 2022, doi: 10.1061/9780784483985.043.

[9] Y. Liu, A. Ojha, S. Shayesteh, H. Jebelli, and S. Lee, "Human-centric robotic manipulation in construction: generative adversarial networks based physiological computing mechanism to enable robots to perceive workers' cognitive load," *Canadian Journal of Civil Engineering*, vol. 50, no. 3, pp. 224–238, 2023, doi: 10.1139/CJCE-2021-0646/ASSET/IMAGES/CJCE-2021-0646_F7.JPG.

[10] J. M. Nwaogu, Y. Yang, A. P. C. Chan, and H. lin Chi, "Application of drones in the architecture, engineering, and construction (AEC) industry," *Autom Constr*, vol. 150, p. 104827, Jun. 2023, doi: 10.1016/J.AUTCON.2023.104827.

[11] H. W. Choi, H. J. Kim, S. K. Kim, and W. S. Na, "An Overview of Drone Applications in the Construction Industry," *Drones 2023, Vol. 7, Page 515*, vol. 7, no. 8, p. 515, Aug. 2023, doi: 10.3390/DRONES7080515.

[12] G. Mahajan, "Applications of Drone Technology in Construction Industry: A Study 2012-2021," *Int J Eng Adv Technol*, vol. 11, no. 1, pp. 224–239, Oct. 2021, doi: 10.35940/IJEAT.A3165.1011121.

[13] Z. Zhu, I. Jeelani, and M. M. Gheisari Rinker, "Safety Risk Assessment of Drones on Construction Sites using 4D Simulation".

[14] A. Onososen, I. Musonda, M. Ramabodu, and C. Dzuwa, "Safety and Training Implications of Human-Drone Interaction in Industrialised Construction Sites," *Lecture Notes in Civil Engineering*, vol. 358 LNCE, pp. 281–295, 2023, doi: 10.1007/978-3-031-32515-1_20/FIGURES/12.

[15] S. Shayesteh, A. Ojha, Y. Liu, and H. Jebelli, "Human-robot teaming in construction: Evaluative safety training through the integration of immersive technologies and wearable physiological sensing," *Saf Sci*, vol. 159, p. 106019, Mar. 2023, doi: 10.1016/J.SSCI.2022.106019.

[16] G. Albeaino, P. Brophy, M. Gheisari, R. R. A. Issa, and I. Jeelani, "Working with Drones: Design and Development of a Virtual Reality Safety Training Environment for Construction Workers," *Computing in Civil Engineering 2021 - Selected Papers from the ASCE International Conference on Computing in Civil Engineering 2021*, pp. 1335–1342, 2021, doi: 10.1061/9780784483893.163.

[17] Y. Han, Y. Diao, Z. Yin, R. Jin, J. Kangwa, and O. J. Ebohon, "Immersive technology-driven investigations on influence factors of cognitive load incurred in construction site hazard recognition, analysis and decision making," *Advanced Engineering Informatics*, vol. 48, p. 101298, Apr. 2021, doi: 10.1016/J.AEI.2021.101298.

[18] Y. Liu *et al.*, "Studying the Effects of Back-Support Exoskeletons on Workers' Cognitive Load during Material Handling Tasks," *Construction Research Congress 2024, CRC 2024*, vol. 1, pp. 659–669, 2024, doi: 10.1061/9780784485262.067.

[19] S. Shayesteh and H. Jebelli, "Investigating the Impact of Construction Robots Autonomy Level on Workers' Cognitive Load," *Lecture Notes in Civil Engineering*, vol. 239, pp. 255–267, 2023, doi: 10.1007/978-981-19-0503-2_21/FIGURES/4.

[20] M. Zahabi and A. M. Abdul Razak, "Adaptive virtual reality-based training: a systematic literature review and framework," *Virtual Real*, vol. 24, no. 4, pp. 725–752, Dec. 2020, doi: 10.1007/S10055-020-00434-W/FIGURES/5.

[21] A. Ojha, Y. Liu, H. Jebelli, H. Cheng, and M. Kiani, "Improving Health Monitoring of Construction Workers Using Physiological Data-Driven Techniques: An Ensemble Learning-Based Framework to Address Distributional Shifts," *Computing in Civil Engineering 2023: Resilience, Safety, and Sustainability - Selected Papers from the ASCE International Conference on Computing in Civil Engineering 2023*, pp. 631–638, 2024, doi: 10.1061/9780784485248.076.

[22] Y. Gautam and H. Jebelli, "Design of flexible polyimide-based serpentine EMG sensor for AI-enabled fatigue detection in construction," *Sens Biosensing Res*, vol. 46, p. 100713, Dec. 2024, doi: 10.1016/J.SBSR.2024.100713.

[23] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," *Neural Comput*, vol. 12, no. 10, pp. 2451–2471, 2000, doi: 10.1162/089976600300015015.

Towards AI-based Optimization of human-centered and robot-assisted construction processes

Anne-Sophie Saffert[1], Jonas Wiederer[1], Simon K. Hoeng[1], Thomas Linner[1], Mathias Obergriesser[1] and Patrick Neumann[2]

[1]Faculty of Civil Engineering, Ostbayerische Technische Hochschule Regensburg, Germany
[2]Department of Mechanical, Industrial and Mechatronics Engineering, Toronto Metropolitan University, Canada

annesophie.saffert@oth-regensburg.de, jonas.wiederer@oth-regensburg.de, simon.hoeng@oth-regensburg.de, thomas.linner@oth-regensburg.de, mathias.obergriesser@oth-regensburg.de, pneumann@torontomu.ca

Abstract -

This research explores an innovative AI-driven approach to optimizing construction processes with a focus on human-centered design, addressing key challenges in the construction industry, such as skilled labor shortages and ergonomic risks associated with work-related musculoskeletal disorders. By integrating process design with AI-based algorithms into simulation tools, various construction process layout variants including robot-assisted scenarios can be simulated and evaluated based on user-specific key performance indicators (e.g., ergonomic score, layouting parameters) to identify optimized solutions. A data processing algorithm automates the process, eliminating the need for manual simulation variations and resulting in increased operational productivity. The AI-based system evaluates and optimizes process layouts by adjusting control parameters. A case study on a brick laying process serves as an exemplary use case, highlighting the necessity and impact of adopting process optimization. The findings emphasize the transformative potential of automated process optimization within simulation environments to rethink existing construction practices, enhance worker well-being, and boost operational productivity.

Keywords -

AI-based Learning; Simulation; Process optimization; Human Factors; Parametric Design Automation

1 Introduction

The construction industry is a crucial contributor to global economics [1]. However, it faces arising multifaceted challenges such as economic or demographic change, which worsens existing challenges [2]. In addition to inefficiencies and the slow adoption of digitalization and automation, the construction industry confronts an acute and growing shortage of skilled workers. The declining appeal of construction work as a field of employment exacerbates the physical demands on the existing workforce – who are essential to many construction processes – thereby significantly increasing the risk of work-related musculoskeletal disorders (WRMD) [3]. This widely prevalent health issue can result in premature retirement due to health reasons [4], thus aggravating the existing lack of skilled workers. To sustainably utilize human labor, construction processes must be optimized to minimize the risk of medium- and long-term health issues for workers. Hereby, Computer-Aided Process Planning (CAPP) offers a methodical option, combining a digital twin of the process containing the working tasks, which enables a realistic simulation [5]. The subsequent analysis of the simulated process gains insight to various predefined criteria. Advanced process simulations using the CAPP methodology can identify risks such as musculoskeletal disorders and serve as a foundation for human-centric optimization of construction processes. Human-centric optimized processes may represent a crucial element in addressing the aforementioned increasing shortage of skilled workers. This optimization efforts through simulation currently rely on manually iterations due to insufficient innovative approaches. Thus, each simulative change is implemented and evaluated individually, which involves a high time expenditure and complexity.

In addition to the human factors, a multitude of challenges exist in manual construction processes that can be described parametrically and, depending on the software solution, optimized in a suitable digital environment. With this premise, the number of practicable variants remains limited due to the manual processing of such simulations, which in many cases leads to suboptimal solutions and requires an innovative solution.

2 State of the Art

Process simulation is a digital tool currently implemented in various fields, such as the automotive industry [6]. This approach allows the digital production and process planning by creation of a digital process twin by the utilization of suitable CAPP software [5]. The ema Work Designer by imk Industrial Intelligence or Tecnomatix by Siemens for instance represent such a CAPP tool

including digital human models. Integrating human actions into the digital process enables the evaluation of process ergonomics based on various criteria that represent the physical load on the worker's body, whether caused by motion alone or by external forces acting on it. In the literature, several well-established evaluation criteria are widely recognized. These ergonomic evaluation criteria, which are particularly relevant in the context of this research, include:

- Ergonomic Assessment Worksheet (EAWS) score: Categorized into three groups, the EAWS score within a range of 0 to 25 represents a process without risk to get affected by WRMDs resulting in musculoskeletal damage. The range of 25 to 50 shows middle risk with recommendation to adapt these processes in a human-centered way, whereas values of more than 50 in EAWS score relate to high risk to suffer from WRMDs and therefore indicate the need for urgent process adaptation. Including the process duration, this commonly used criteria considers the posture combined with external loading [6].
- NIOSH: The NIOSH risk index represents a measure for the risk of adopting WRMDs caused by human lifting activities. This approach considers detailed parameters, such as weight of the object or vertical and horizontal location of object and target [7].
- RULA/REBA: The Rapid Upper Limb Assessment (RULA) and Rapid Entire Body Assessment (REBA) scores indicate the risk of getting affected by WRMDs analyzing body and segment angles, and repetitions of movements. RULA and REBA are based on assessment worksheets calculating an index considering single body part postures in addition to loading on the body resulting in a score from 1 (acceptable risk) to 7 in RULA respective 11 and more in REBA (high risk) [8]. In contrast to EAWS, RULA/REBA do not consider body postures and their duration [9].

As the topic of human-centered work process in construction is gaining relevance [10, 11], Kulkani and Devalkar investigated the ergonomic influence on construction workers in various construction tasks (granite cutting, brick work, shuttering, plastering and material transportation) by applying the RULA and REBA index [8]. Resulting in evaluation scores of 6 and 7 in RULA as well as 11 and more in REBA, the investigated construction processes indicate a high risk of the worker adopting WRMDs.

Schmailzl et al. proposed in prior work a framework for decision making regarding the degree of automation of an exemplary brick laying process [12]. To represent a process reproduction as close to reality as possible, motion capturing was used to record the construction worker's movements. Therefore, this approach evaluates the captured brick laying process regarding ergonomics of the worker incorporating the EAWS score as well as task duration within various initial layouts as well as the influence of integration of a robot as assistive technology. The results, showing an EAWS score of up to 154, clearly highlight the need to optimize certain construction processes with a focus on the human worker.

In contrary to standardized production lines, construction sites are characterized by dynamic changing conditions. However, recurring and standardized subprocesses, such as bricklaying, enable the implementation of CAPP software. Currently, even when these types of processes are planned and optimized, they are still manually developed and adjusted by layout designers. A comprehensive parametric design of layouts offers considerable advantages, as changes can be implemented quickly and effectively. Both in production factories and in the construction industry, the ability to make adaptive adjustments is becoming increasingly important, particularly due to the growing demand for customized production and buildings. Despite the great potential of artificial intelligence (AI) and algorithmic approaches in the field of process optimization, there are currently – due to the best knowledge of the authors – a lack of scientific studies that comprehensively address the forementioned process algorithmic approaches combined with ergonomic evaluation scores applied to the construction industry. Parametric design is frequently used in the construction industry, however, mainly in the context of building modelling and less in classical project management. As Abioye et al. showed in their literature review, AI-based optimization seems to be a promising approach for improving processes in the AEC industry [13]. The study by Bergmann shows how parameters simplify the control of an environment and which advantages result from this [14]. The use of generic approaches for optimization can improve the efficiency and performance of production systems significantly. Of particular interest are genetic algorithms (GAs), especially the NSGA-II variant, because they are able to handle numerous parameters and enable multi-objective optimization. Another example from the warehouse sector shows that alternative unconventional layout concepts, such as non-orthogonal shelving arrangements or v-shaped aisle systems, can offer considerable efficiency gains [15]. Investigating different layout options is currently a time-consuming manual task that limits the potential variance of the layout. A tailored algorithm could perform this task in a fraction of the time required by a designer while also identifying irrational but applicable solutions. The use of parametric algorithms in combination with detailed simulations enables the adjustment of critical layout parameters – such as the width of the aisles, the number of racking bays, or the capacities – which optimizes the flow of material and the use of storage space.

Building on the application of AI-driven optimization methods, such as those used in logistics, this research aims to transfer and adapt the concept of automated optimization to the construction industry, focusing on applications where fundamental principles, such as material transportation and storage, are analogous to those in other sectors.

Derivated from, i.a., the findings of Schmailzl et al. [12], this research aims for and proposes a concept for automated generation of process simulation variants, their evaluation as well as AI-based optimization in the field of construction in order to exceed a limited number of simulation cycles as in manual simulation. Additionally, the approach has the potential to reduce the risk of workers developing WRMDs, thereby promoting a more sustainable and attractive utilization of the workforce in the construction industry.

3 Methodology

Given the challenges faced by the construction industry, particularly the high health risks associated with construction work [3] and the limited adoption of digitalization, this research proposes a methodological approach for automated software pipeline encompassing an AI-driven optimization of construction processes with a focus on human-centered design. This section elaborates the concept of parametric process generation, the automation of simulation & evaluation and AI-based optimization.

The presented concept illustrates the intended functionality of a fully automated optimization loop; at this stage, intermediate manual data transfer steps are necessary due to current software capabilities regarding API connectivity, which currently restricts execution to a limited number of optimization cycles. Potential solutions to enable complete automation are discussed further in Section 5.

3.1 Concept

Figure 1 illustrates the proposed methodology. Taking a holistic perspective on construction processes, it presents a workflow comparison between traditional process optimization methods and the enhanced efficiency achievable through AI-based algorithms. With the ongoing digitalization of the construction sector, simulating models and components using advanced software has become a standard practice to optimize building designs prior to physical implementation. This approach extends to factory layout planning, where efficiency improvements are achieved through detailed simulations.

The concept shown in Figure 1 outlines how specific construction tasks, such as bricklaying or concrete pouring, can be integrated into CAPP software. The initial step involves the digital, parametric representation of tasks, incorporating parameters such as object positions and material information. This parametric control can be controlled by a software pipeline and enables a systematic as well as adaptable process. Process layout designers typically create an initial setup based on their expertise, but this manual approach often relies on subjective judgment. Analyzing key performance indicators (KPIs), such as ergonomic factors, material flow efficiency, or pathway criteria, offers a preliminary evaluation of the process. Without a defined baseline incorporating KPIs, comparative assessments regarding the process quality remain limited. The optimization loop in Figure 1 demonstrates two distinct approaches. The dashed line (1) illustrates the traditional method, which is characterized by the manual adjustment of parameters followed by manual repetition of simulations and the comparison of outcomes. Due to the time-intensive nature of this simulation scenario, this restricts the exploration of possible variants. Conversely, the second approach leverages AI-driven algorithms, which autonomously modify parameters and executes numerous simulations. The results are stored systematically and analyzed in relation to the chosen parameters.

By selecting appropriate algorithms, the optimization process achieves superior KPIs and significantly reduces manual work for achieving an optimized solution. AI-based methods are characterized by the recognition of correlations between parameter adjustments and their effects on the results, thus enabling more efficient and precise optimization.

Consequently, this advanced simulation framework provides a robust foundation for optimizing construction sites, enhancing process quality, and improving conditions based on specific KPIs and therefore serves as a basis to prevent the risk of getting affected by WRMDs delineating an approach to reduce the immense lack of skilled workers.

3.2 Structure of the parametric layout

To enable a layout to be fully controlled through parameters, an external structure must first be established that allows such control. This requires assigning unique parameter names to all values within the system, ensuring clear identification and reference. With this parametric structure in place, the entire layout can be adjusted flexibly by simply entering specific values. Moreover, the parametric system should allow the selection of different objects to enhance the realism of the scenario. The selection of specific process components was investigated in prior work [12]. Based on these outcomes, boolean operators must be incorporated to select the desired objects, with all objects existing in parallel within the system to maintain consistent parametric relations.

In case of manual adjustments, these are performed directly within the software and can be tailored to specific needs, offering flexibility, but often at the cost of time and

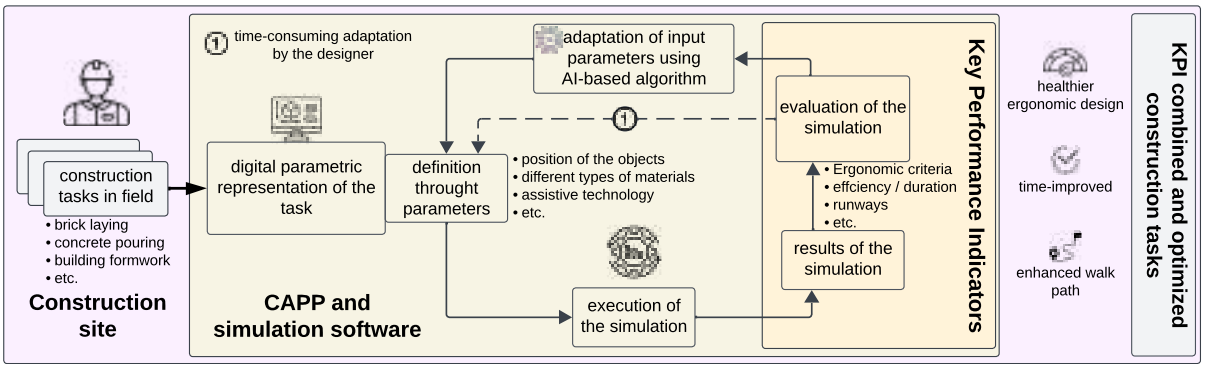


Figure 1. AI-based process optimization method: Is applied to construction tasks (left) optimized by predefined KPIs (right). The green CAPP box represents the automation of the manual workflow, with AI adapting parameters to generate and process multiple variants.

efficiency. To integrate algorithms into this process, an import and export interface for the parameters is essential.

Depending on the software, it may be necessary to export the entire setup or model and reimport it after structural modifications. More advanced software solutions allow the export of parameters alone, which can then be modified and reimported. This leaner approach significantly reduces time intense computational processes and enhances workflow efficiency.

Hereby, a well-developed API interface enables effective collaboration with algorithms and supports streamlined and time-efficient adjustments of parametric structures. Even rudimentary APIs often provide basic import and export functionalities, which are sufficient to implement algorithms into the workflow.

3.3 Validation of the simulation results

The validation is a critical step in process design, it gives value feedback to the planner to improve the setup. For every process, specific KPIs must be defined to serve as measurable benchmarks for future optimization. Efficiency remains a fundamental KPI, often including metrics such as cycle time, resource consumption, and material pathways. In manufacturing, this KPI play a central role in reducing costs and improving production rates. The ergonomic score is meanwhile gaining importance and is increasingly used, especially in manufacturing processes. In Germany, the Load Handling Ordinance requires risk assessments but does not consider details like weight limits or biomechanical impacts on the human [16]. In practice, a 25 kg limit is often used for manual lifting, but factors like carrying distance can greatly affect the biomechanical evaluations. In general, planners must evaluate trade-offs between competing KPIs, as optimizing one parameter may impact other parameters. Storing and documenting process results is essential for thorough validation. This includes capturing input parameters, such as material specifications or positions, alongside output metrics as discussed. Such data build the basis for advanced AI-driven optimization techniques. By analyzing data, AI algorithms can identify patterns and make targeted adjustments, leading to more efficient and streamlined processes.

3.4 Optimization of the results using AI

Optimization tasks require a careful selection of algorithms, each offering distinct advantages depending on the problem and context. For instance, gradient-based algorithms rapidly converge to a single optimal solution by iteratively improving from an initial guess, ideal for clearly defined, differentiable goals. However, for complex problems with conflicting objectives, genetic algorithms (GAs) are particularly effective in finding optimizations due to their inherent capability to manage multiple trade-offs simultaneously. The NSGA-II algorithm applied by Bergmann is a recognized method for multi-objective optimization [14], explicitly chosen in this study for its robust handling of multiple interdependent KPIs and its ability to maintain diverse, balanced solutions.

The NSGA-II algorithm refines a set of solutions step by step using evolutionary principles. Starting with random solutions, each computation repetition represents different combinations of parameters, resulting in specific outcomes. These outcomes are evaluated based on defined KPIs, such as efficiency or human ergonomics. The results are then ranked into groups, so-called Pareto fronts, with the best, non-dominated solutions placed in the first front. To ensure diversity, the algorithm calculates a crowding distance, spreading solutions evenly across the objective space.

The best solutions from the first front are selected and combined to create new combinations of input parameters by recombination. Additional mutations introduce

small variations of the parametric to explore new possibilities and avoid premature convergence. This process of evaluation, selection, recombination, and mutation recurs iteratively until a predefined stopping condition is met, such as reaching a set number of generations or achieving convergence.

The result is a Pareto front, a collection of high-quality solutions that balance conflicting objectives. Decision-makers can choose the most suitable solution from this set, depending on their specific priorities and constraints. This choice typically requires additional context-specific criteria or weighting of KPIs – such as prioritizing ergonomics over productivity. This makes the NSGA-II especially effective for problems involving multiple competing objectives. Additionally, the NSGA-II qualifies as an AI-based approach due to its use of evolutionary mechanisms and heuristic search strategies. Instead of exhaustively evaluating all possibilities, it focuses on exploring the most promising regions of the solution space efficiently.

Concluding, this flexibility and efficiency make the NSGA-II a highly promising choice for multi-objective optimization problems, especially in dynamic and constrained environments.

4 Application example: Brick laying process

The potential applications for this concept are diverse and can be implemented across various construction processes. In this section, we present a specific use case as an example of how the concept can be applied. As the initial scenario, we use a brick laying process that has been previously performed and analyzed in prior research.

Incorporating digital human models, the software is capable to consider human process behavior for a more realistic process representation. Besides the standard movements as provided by the algorithms of the software, human motion can be integrated from motion capturing. The motion data was recorded via an Inertial Measurement Unit (IMU) based Motion Capturing system as the orange sensor attachments in Figure 2 represent. The process was developed within the emaWD simulation software, a powerful tool capable of modeling complex setups and offering a range of advanced analytical functions. emaWD is widely utilized for factory layout simulations, providing comprehensive feedback to designers across multiple KPIs, including all those relevant and previously discussed in earlier sections. These sensor units combine an accelerometer, magnetometer and gyroscope to determine the human movements [17].

In manufacturing processes, robots often play a significant role. The software enables the integration of a vast selection of robot models, each with diverse functionalities and gripping mechanisms. Moreover, the software allows



Figure 2. Motion Capturing of a brick laying use case at the Building Lab of Ostbayerische Technische Hochschule (OTH) Regensburg.



(a) Simulation of a brick laying process with the optimization of the integration of the robot HC10 by Yaskawa.



(b) Realization of simulated and optimized brick laying process integrating the Yaskawa HC10 at the Building Lab of OTH.

Figure 3. Optimized robot-assisted brick laying process towards the human well-being.

customization and restrictions for robots, accommodating specific models that may not be natively included. The ability to incorporate and program these robots within the algorithm significantly broadens the scope of potential solutions, potentially leading to entirely novel configurations with drastically altered KPIs. While simply adding a robot to a process might seem to improve performance, it also introduces additional factors such as flexibility, feasibility, and cost, which must be carefully evaluated. These considerations are critical to ensure that the benefits of robot integration outweigh the associated challenges. Later in this section, the parametric integration of robots into the optimization process will be explored in greater detail. Although the software is not originally designed for construction tasks, designers can creatively adapt it to simulate various building processes, including the one described in

this study.

Besides the application of robotic devices, the earlier study highlights the significant impact of layout changes on key performance indicators (KPIs). In this process, depending on the four manually implemented layout options, the KPIs EAWS score, process duration and walk length were explored. The four layout variants differed as follows:

1. Original process: The original process describes a brick laying process done by a human worker, who transfers bricks from a pallet to a target position.
2. Repositioning of the pallet: Based on the original process, the pallet position was shifted 2 meters to opposite direction of the brick target position (see Figure 4).
3. Adaptation of brick material: Based on the original process, the brick type was changed resulting in a weight reduction from 4.5 kg to 2.8 kg for a same-sized brick.
4. Combined adaptation of brick material and pallet repositioning: This scenario represents both aforementioned repositioning of the pallet and adaptation of brick material.



Figure 4. Repositioning of brick storing pallet in a distance of 2.0 m to the original pallet position, depicted as gray copy.

The outcomes were systematically compared, with KPIs serving as the metrics to evaluate the performance of each variation.

Interestingly, the research revealed that certain irrational layout modifications resulted in better outcomes. All exemplary executed layout configurations were manually designed by planners and executed within the emaWD simulation software. The resulting processes exhibited varying KPI outcomes, highlighting the challenge of identifying the optimal layout. This case study demonstrates the complexity of achieving optimal solutions, even for a straightforward and repetitive process. When additional parameters and variables are introduced, finding layouts that improve all required KPIs simultaneously becomes nearly impossible without systematic optimization. This scenario emphasizes the necessity of integrating the proposed concept into the exemplary brick laying process to address these challenges effectively. The software already supports various export and import formats as essential basis for robust data transfer. An export method is provided by the emaWD Wizard, a data transfer assistant that generates an *xlsx*-file to represent the emaWD setup. This file has the capability to define all tasks, objects, and executing agents – whether human models or robots – completely with coordinate information. For the demonstration of the concept, the brick laying process was exported using the Wizard and described in a structured text format. At this stage, the *xlsx*-file can be accessed via a Python script, allowing extraction, manipulation, and adaptation of all relevant data. In this example, product information such as the weight and dimensions of the bricks was modified within the script and reflected in the *xlsx*-file. As noted earlier, every task and component can be adjusted within the spreadsheet. The algorithm modifies these layout parameters and updates the setup within the *xlsx*-file. Upon importing the updated spreadsheet into the emaWD software, a new simulation is executed. The results, expressed as KPIs, provide immediate feedback to the designer through various representations. These KPI outputs can also be exported in simpler formats, such as *csv*-files, given their reduced complexity compared to the process setup. This exported data is then integrated back into the Python script, supplying the algorithm with essential feedback from the results. The connection between input parameters and KPIs is pivotal, as it forms the foundation of the optimization loop. This single optimization cycle, with changed input parameters, is repeated multiple times to generate a diverse and extensive first dataset with different parametric, enabling the algorithm to identify non-dominated solutions across multiple Pareto fronts.

After the initial round of creating a broad dataset, the algorithm proceeds as outlined in subsection 3.4 selecting the most relevant data to further refine and improve the process. Currently, emaWD does not offer a public API interface, which limits its ability to fully support the intended operation of the optimization loop. The absence of an API interface prevents direct control over the software pipeline, hindering the algorithm's ability to efficiently optimize the parametric configurations and limiting its potential to enhance the parameters in a more effective and automated manner.

5 Discussion

This research explores the implementation of an automated AI-based approach for optimizing construction processes with a focus on the human worker.

The application of proposed research within an exemplary brick laying use case as demonstrated in section 4

Table 1. Overview of the simulation results and KPIs corresponding to the four considered design variants.

Design Variant	EAWS Posture Score	EAWS Load Score	**Sum EAWS Score**	Distance [m]	Duration [s]
1	19.5	140	**159.5**	4.7	47
2	8.5	140	**148.5**	16.5	55
3	37	110.5	**147.5**	4.7	47
4	27	110.5	**137.5**	16.5	50

primarily considers human-related KPIs as a measure to optimize construction processes. In particular, the EAWS score as measure for the worker's ergonomics is utilized in the application example. In general, according to the EAWS scores as depicted in Table 1 it seems that the human experiences lower risk of WRMDs when minimizing the load on the human body, as expected. Furthermore the EAWS increases when reducing the pathway as intuitive assumption would suggest. However, the subcategories load score and posture score of the EAWS cannot be explained by intuitive assumptions. Whereas the material handling score decreases from a value of 140 in the original process to a value of 110.5 in the brick weight reduced process as demonstrated in section 4, the relating posture scores show contradictory behavior. Consequently, the optimal resulting EAWS score can hardly be predicted by intuitive assumption of process design variants. Currently, the emaWD software only allows manual import and export of the KPIs and parametric, creating a bottleneck for efficient algorithm implementation. Integrating an automated AI-based process optimization is likely to enable the finding and application of unconventional and unintuitive parameters in design variants within simulation, which is an essential basis for effective process optimization in general.

The proposed algorithm represents an innovative approach to process optimization, enabling the simultaneous optimization of multiple KPIs. Particularly in optimization scenarios involving human factors alongside efficiency and other metrics, it is essential to identify balanced compromises that address all relevant aspects effectively. However, the effective application of AI-based tools relies on seamless data transfer and access through a robust API that allows algorithms to interface directly with simulation software. Despite the growing capabilities of AI-driven approaches, seemingly suitable simulation tools often lack the necessary API functionalities. This limitation significantly constricts the potential for automating and optimizing processes efficiently. Addressing this gap by developing powerful APIs would unlock new possibilities for improving existing workflows and overcoming current time-induced constraints through leveraging AI-based optimization.

Process simulation tools are already commonly used in various industrial fields, such as the automotive industry [18]. Given their ability to optimize production setups before realization, these simulation tools hold great promise for prefabrication processes within the construction sector. Specifically, the proposed optimization approach is particularly suited to prefabricated building components (e.g., wall elements, modular units) or standardized infrastructure elements (e.g., precast bridge segments, tunnel lining segments), where stable, predictable conditions prevail. Direct implementation on traditional, dynamic construction sites remains challenging due to variability; thus, prefabrication represents the primary and most effective context for applying the proposed simulation-driven optimization method.

This work emphasizes the critical role of the human worker, especially in mitigating issues related to health risks and improving working conditions. By automating process simulations, this approach has the potential to contribute to more efficient and effective human-oriented optimization methods.

While AI-driven optimization offers significant advantages, it also involves trade-offs. Implementation costs – such as computational resources and robotics integration – must be weighed against potential efficiency gains. Additionally, selecting appropriate hyperparameters is essential, as improper tuning can compromise algorithm stability and the quality of optimization results.

6 Conclusion

This research highlights the pressing need for automated, AI-based process optimization in the construction industry, emphasizing the importance of human-centered design. The proposed conceptional holistic approach demonstrates the potential to address critical challenges by identifying and analyzing complex parametric structures, simulating diverse scenarios, and evaluating specific construction tasks to achieve optimized process solutions. The ability of such systems to identify unconventional solutions opens up new possibilities for process improvement, alongside with an increase in simulation efficiency. Crucially, the development of robust APIs is identified as a foundational requirement to fully realize the benefits of this approach. When seamless integration via robust APIs becomes available, the proposed approach has the potential to unlock the full capabilities of AI-based optimization, driving a more efficient, adaptive, and sustainable construction sector while significantly improving human factors and fostering safer and more worker-centric con-

struction processes.

References

[1] Tender M., Couto J.P., Gibb A., Fuller P., and Yeomans S. Emerging technologies for health, safety and well-being in construction industry. In Bolpagni M., Gavina R., and Ribeiro D., editors, *Industry 4.0 for the Built Environment*, volume 20 of *Structural Integrity*, pages 369–390. Springer International Publishing, Cham, 2022. doi:10.1007/978-3-030-82430-3Ω_16.

[2] Agarwal R., Chandrasekaran S., and Sridhar M. Imagining construction's digital future. *McKinsey & Company*, 24.06.2016.

[3] Wang D., Dai F., and Ning X. Risk assessment of work-related musculoskeletal disorders in construction: State-of-the-art review. *Journal of Construction Engineering and Management*, 141(6), 2015. doi:10.1061/(ASCE)CO.1943-7862.0000979.

[4] Shiri R., Heliövaara M., Ahola K., Kaila-Kangas L., and Haukka E. et al. A screening tool for the risk of disability retirement due to musculoskeletal disorders. *Scandinavian Journal of Work, Environment & Health*, 44(1):37–46, 2018.

[5] Alting L. and Zhang H. Computer aided process planning: the state-of-the-art survey. *International Journal of Production Research*, 27(4):553–585, 1989. doi:10.1080/00207548908942569.

[6] Caputo F., Greco A., Fera M., Caiazzo G., and Spada S. Simulation techniques for ergonomic performance evaluation of manual workplaces during preliminary design phase. pages 170–180. Springer, Cham, 2019. doi:10.1007/978-3-319-96077-718.

[7] Dempsey P.G. Usability of the revised niosh lifting equation. *Ergonomics*, 45(12):817–828, 2002. ISSN 0014-0139. doi:10.1080/00140130210159977.

[8] Kulkarni V.S. and Devalkar R.V. Postural analysis of building construction workers using ergonomics. *International Journal of Construction Management*, 19(6):464–471, 2019. doi:10.1080/15623599.2018.1452096.

[9] Konstantinidis D., Dimitropoulos K., and Daras P. Towards real-time generalized ergonomic risk assessment for the prevention of musculoskeletal disorders. In *Proceedings of the 14th PErvasive Technologies Related to Assistive Environments Conference*, pages 466–472, New York, NY, USA, 2021. doi:10.1145/3453892.3461344.

[10] Cheng M.H., Liang C.-J., and Dominguez E.G. Safe operations of construction robots on human-robot collaborative construction sites. *International Symposium on Automation and Robotics in Construction (ISARC) Proceedings*, pages 9–16, 2024. doi:10.22260/ISARC2024/0003.

[11] Fan C., Mei Q., and Li X. Assisting in the identification of ergonomic risks for workers: a large vision-language model approach. *International Symposium on Automation and Robotics in Construction (ISARC) Proceedings*, pages 1010–1017, 2024. doi:10.22260/ISARC2024/0131.

[12] Schmailzl M., Saffert A.-S., Karamara M., Linner T., Eder F. A., Hoeng S. K., and Obergriesser M. Enhancing decision-making for human-centered construction robotics: A methodological framework. 2024. doi:10.22260/ISARC2024/0083.

[13] Abioye S.O., Oyedele L.O., Akanbi L., Ajayi A., and Davila Delgado J.M. et al. Artificial intelligence in the construction industry: A review of present status, opportunities and future challenges. *Journal of Building Engineering*, 44:103299, 2021. doi:10.1016/j.jobe.2021.103299.

[14] Bergmann S. Optimierung des designs modularer montagesysteme. In Breitenecker F., Deatcu C., Durak U., Körner A., and Pawletta T., editors, *ASIM SST 2022 Proceedings Langbeiträge*, pages 15–22. ARGESIM Vienna, July 2022. doi:10.11128/arep.20.a2006.

[15] Saderova J., Poplawski L., Balog Jr M., Michalkova S., and Cvoliga M. Layout design options for warehouse management. *Polish Journal of Management Studies*, 22(2):443–455, 2020. doi:10.17512/pjms.2020.22.2.29.

[16] Ralf Pieper. Verordnung zur umsetzung von eg-arbeitsschutz-richtlinien. *Arbeit und Recht*, 45(1): 21–23, 1997.

[17] Cheuk C.M., Lau T.K., Lin K.W, and Liu Y. Automatic calibration for inertial measurement unit. In *12th International Conference on Control Automation Robotics & Vision (ICARCV)*, pages 1341–1346. IEEE, 2012. doi:10.1109/ICARCV.2012.6485340.

[18] Illmann B., Fritzsche L., Leidholdt W., Bauer S., and Dietrich M. Application and future developments of ema in digital production planning and ergonomics. pages 66–75. Springer, Berlin, Heidelberg, 2013. doi:10.1007/978-3-642-39182-8 8.

Understanding the Impact of an Active Exoskeleton System on Vigilance Level of Workers During Common Repetitive Construction Tasks

Amit Ojha[1], Yogesh Gautam[1], Houtan Jebelli[1], and Abiola Akanmu[2]

[1]Department of Civil and Environmental Engineering, University of Illinois Urbana Champaign, United States
[2]Myers-Lawson School of Construction, Virginia Tech, United States
amito2@illinois.edu, ygautam2@illinois.edu, hjebelli@illinois.edu, abiola@vt.edu

Abstract

The adoption of active exoskeleton systems in the construction industry shows potential for reducing work-related musculoskeletal disorders (WMSDs) by providing physical support to workers during labor-intensive tasks. However, concerns about the potential impact of exoskeletons on workers' vigilance and attention to hazards have not been fully explored. This study examines the effect of active exoskeletons on the vigilance levels of construction workers during common construction tasks. A high-fidelity, immersive virtual reality (VR) environment was developed to realistically simulate material handling tasks and collect eye-tracking data. Eighteen participants completed tasks under two conditions: with and without the use of an active back-support exoskeleton. Eye-tracking metrics, including fixation count and fixation duration on predefined Areas of Interest (AOI) of hazards in the simulated environment, were collected from the participants and analyzed to assess vigilance levels. The results showed no statistically significant difference in vigilance levels between participants, regardless of whether they used the exoskeleton. These findings suggest that the use of active exoskeletons does not negatively affect workers' vigilance or their ability to identify hazards. This study provides an empirical basis for supporting the wider adoption of active exoskeletons on construction sites.

Keywords –

Active Exoskeleton; Vigilance; Psychological Assessment; WMSDs

1 Introduction

The construction industry often requires workers to engage in repetitive, physically demanding tasks while maintaining uncomfortable postures for extended periods [1,2]. This biomechanical stress on the musculoskeletal system increases the likelihood of developing Work-Related Musculoskeletal Disorders (WMSDs) [3,4]. Data from the U.S. Bureau of Labor and Statistics reveals that WMSDs constitute approximately 37% of non-fatal injuries and illnesses among construction workers [5]. Furthermore, studies indicate that the construction industry faces annual workers' compensation costs exceeding $400 million due to WMSDs [6].

Prominent safety organizations like NIOSH and OSHA have advocated proper ergonomic practices to reduce WMSDs in construction [7]. However, the industry continues to be characterized by physically demanding tasks involving repetitive movements and awkward postures. The primary challenge stems from the dynamic and diverse nature of construction work, which resists the application of standardized ergonomic practices across all tasks and sites.

Exoskeletons, also known as wearable robotic systems, are gaining recognition as vital tools in the construction industry for mitigating WMSD risks. These devices are designed to assist workers by providing support during heavy lifting, distributing weight more evenly, and promoting proper posture, thus benefiting various construction tasks. Exoskeletons are generally categorized into two types: active and passive. Active exoskeletons are equipped with actuators, such as electric motors, hydraulics, or pneumatics, which actively amplify the wearer's physical strength [8]. On the other hand, passive exoskeletons do not have actuators; they rely on materials like springs or dampers to store energy from human movement and release it when needed to support the wearer [8]. While passive exoskeletons are often preferred due to their lower cost and ease of use, active exoskeletons offer greater potential for broader adoption in construction due to their capacity to provide more substantial physical assistance. However, as active exoskeletons become more prevalent in construction, new challenges and considerations are likely to arise.

Although powered exoskeletons offer considerable potential to mitigate WMSDs in construction, their

widespread adoption remains hampered by a limited understanding of their broader impacts on workers' psychophysiological state. While much of the current research focuses on understanding the physiological impact of exoskeletons [9–14]—such as muscular fatigue, metabolic cost and ergonomic risks, the psychological dimensions are often overlooked. The current body of knowledge fails to provide a comprehensive understanding of the psychological impacts of exoskeletons which may lead to decreased productivity and safety risks for workers wearing exoskeletons. The deployment of an exoskeleton on a job site may cause additional distractions, negatively impacting their vigilance (situational awareness) [15–18]. Wearing a powered exoskeleton may introduce new distractions for workers, such as the need to adjust the device, manage its operation, or adapt to the altered physical sensation it brings. These distractions can take attention away from the task at hand, reducing vigilance and potentially increasing the likelihood of accidents. For instance, if a worker is overly focused on the functioning of the exoskeleton—such as ensuring it is aligned correctly or adjusting to the support it provides—they may be less attentive to their immediate environment, including hazards or moving equipment. Wearing a powered exoskeleton may introduce new distractions for workers, such as the need to adjust the device, manage its operation, or adapt to the altered physical sensation it brings. These distractions can take attention away from the task at hand, reducing vigilance and potentially increasing the likelihood of accidents [19]. For instance, if a worker is overly focused on the functioning of the exoskeleton—such as ensuring it is aligned correctly or adjusting to the support it provides—they may be less attentive to their immediate environment, including hazards or moving equipment.

To that end, this research aims to investigate how active exoskeleton systems affect the vigilance levels of construction workers during routine tasks. To achieve this, a high-fidelity, immersive virtual reality (VR) environment was created to replicate common construction activities. This simulated environment served as a virtual testbed for gathering eye-tracking data, which was then analyzed to determine the impact of active exoskeleton use on workers' vigilance. The insights gained from this study contribute to understanding the psychological implications of implementing powered exoskeletons on construction sites, particularly regarding how they influence attention and focus during tasks.

2 Vigilance for Worker Safety in Technology-driven Construction Sector

In the construction industry, the integration of emerging technologies such as robotics heightens the need for psychophysiological well-being. Such technologies bring new complexities and learning curves, which can impact workers' psychophysiological well-being [20,21]. Ensuring psychophysiological well-being is essential for the construction workers in adapting to technology-driven workplaces, embracing innovation, and maintaining safe, productive work environment. The introduction of advanced technologies in construction can significantly alter cognitive demands on workers. As construction sites become more technologically advanced, maintaining vigilance about potential hazards is also increasingly and equally important [22]. Workers need to be continuously aware of both safety hazards and new risks introduced by technology. This involves staying vigilant to the operational aspects of technology, understanding its limitation, and being prepared for unexpected scenarios that might arise from technological integration. When safety hazards go unrecognized, workers are prone to engaging in unsafe behaviors, facing unexpected exposure to hazards, and potentially suffering severe injuries. Conversely, identifying safety hazards enables workers to evade such exposures, implement proper safety protocols, and prevent injuries at the workplace. Distracting technology can hinder the performance of construction tasks, impairing workers' ability to detect and address hazards effectively. Such technology not only affects their productivity but also compromises overall safety at the construction site.

3 Impact of Active Exoskeleton on Vigilance Level of Workers

While active exoskeletons in the workplace show great promise, they also introduce a range of new safety and usability concerns, from physical issues to psychological implications. The deployment of an exoskeleton on a job site can inevitably cause additional distractions, negatively impacting their vigilance (situational awareness) [15–18]. The additional sensory inputs from the device, such as vibrations or sounds, can lead to sensory overload, potentially overwhelming a worker's ability to process critical environmental cues. Moreover, the cognitive load required to operate the exoskeleton may divert attention away from the surrounding environment, reducing overall awareness. In addition, there is also a risk of overconfidence and complacency as workers become accustomed to the enhanced physical capabilities provided by exoskeletons. This could lead to a decreased perception of risk and, consequently, reduced vigilance. The altered sense of body awareness (proprioception) caused by exoskeleton use may affect spatial judgment, potentially leading to misjudgments of physical space and increased risk of accidents. Additionally, the pressure to perform with new

technology and anxiety about potential malfunctions can induce stress, further impacting a worker's ability to maintain optimal vigilance levels. Further, active exoskeletons may produce noise that interferes with important auditory cues in the workplace, and some designs might limit peripheral vision, potentially causing workers to miss visual warnings or approaching hazards.

4 Assessment of Active Exoskeletons on Workers' Vigilance Level

This research aims to explore how active exoskeleton systems influence the vigilance levels of construction workers during typical construction activities. Figure 1 outlines a two-phase research methodology.

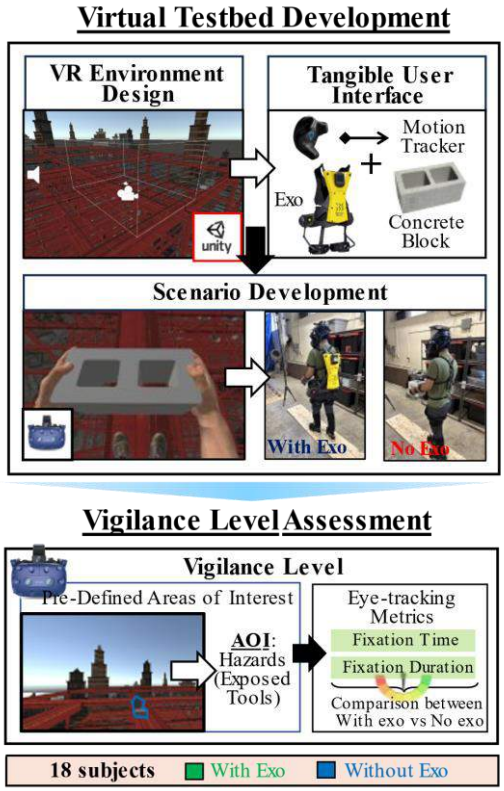


Figure 1. Research Methodology for Assessment of Active Exoskeleton on Workers' Vigilance Level

The study consists of two key phases: Virtual Testbed Development and the Assessment of Vigilance. In the first phase, an immersive and interactive virtual testbed was created to simulate realistic material handling tasks, under two conditions: with and without the use of an active exoskeleton. This environment allowed for the comparison of performance in both scenarios. In the second phase, workers' vigilance levels were assessed by analysing eye-tracking data collected from predefined areas of interest within the virtual environment, corresponding to both conditions. The following subsection will offer a detailed explanation of each step involved in these phases.

4.1 Virtual Testbed Development

In the first module of the proposed framework, a high-fidelity virtual testbed was developed to simulate typical construction tasks. This effort centered on designing a user-centered, simulated workspace to safely assess the impact of active exoskeletons on the vigilance level of workers during common construction tasks. The virtual job site was modeled using Building Information Modeling (BIM) software, specifically Autodesk Revit, and then imported into Unity, a game engine, to create a realistic construction environment, as seen in Figure 2-A. To enhance realism, the virtual space was populated with various elements, including construction machinery, automated systems like drones, personal protective equipment (PPE), and both static and dynamic virtual workers.

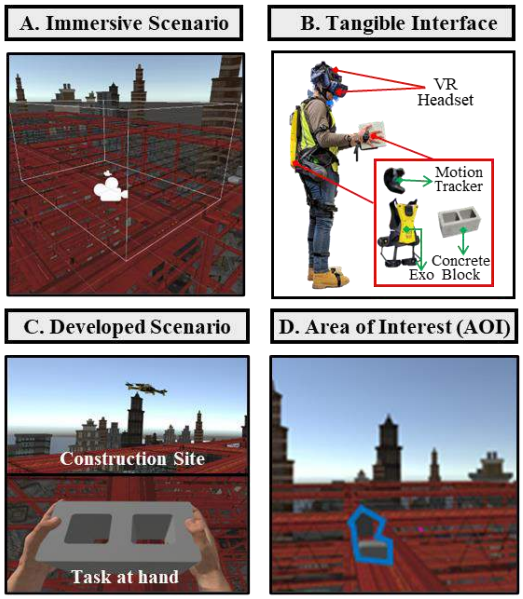


Figure 2. Development of the Virtual Testbed (A) Immersive Scenario Design; (B) Tangible User Interface; (C) Scenario of Material Handling Task at Height; (D) Area of Interest (AOI) in the virtual environment

A virtual representation of an active exoskeleton was integrated into the virtual environment as the primary tool for interaction. To make physical objects interactive in the immersive environment, a motion-tracking sensor was attached to each object, which was then rendered in

the VR setting, as seen in Figure 2-B. The core task simulated in this environment involved material handling, where participants lifted and carried concrete blocks to a designated area, all while working at heights by walking on an iron beam (Figure 2-C). By incorporating the working at height scenario into the simulated material handling tasks, the authors could comprehensively evaluate how the use of powered exoskeletons might influence the vigilance.

To analyze the impact of active exoskeletons on vigilance level of construction workers during common construction tasks, two scenarios were developed: one where participants performed the material handling tasks without an exoskeleton, and another where they used the exoskeleton. In both scenarios, participants interacted with real concrete blocks and an actual powered exoskeleton, presented as physical objects within the virtual environment. The tasks involved picking up the blocks, walking a 4-meter distance along a narrow 40 cm-wide wooden plank, designed to simulate an iron beam at height, and placing the blocks in a specified location. The immersive VR experience, including consistent sounds and visual effects, was the same for all participants, ensuring uniform conditions across both scenarios and eliminating potential biases. In the simulated environment, an Area of Interest (AOI) of hazards was predefined, which represented possible distraction within the virtual environment. Figure 2-D shows the AOI in the simulated environment. The AOIs in the virtual environment were intentionally placed within participants' peripheral vision and along their natural line of sight during task performance. This design ensured that hazards appeared within the operational workspace, mimicking real-world conditions where workers must remain vigilant to dynamic and spatially distributed risks. The immersive VR setting simulated realistic construction scenarios, where hazards typically appear within the operational workspace rather than at extreme angles requiring extensive head rotation.

4.2 Vigilance Level Assessment

Eighteen participants took part in the material-handling task as part of the study. The average age of the participants was 25.2 ± 4 years, with a mean height of 179 ± 5.5 cm, body mass of 74 ± 6.5 kg, and a body mass index (BMI) of 23.1 ± 2.5 kg/m². Before the experiment, all participants were given detailed informed consent forms that outlined the study's purpose, the data collection process, confidentiality measures, and their rights as participants. Written informed consent was obtained from each participant, in accordance with the guidelines approved by the Institutional Review Board (IRB) at the University of Illinois Urbana-Champaign. The researchers ensured that none of the participants had any medical conditions that could interfere with their performance in the immersive virtual environment, ensuring the integrity of the experiment and the safety of all participants.

After agreeing to participate and signing the consent form, the participants were randomly assigned to one of two virtual scenarios: performing the material handling with exoskeleton or without exoskeleton (Figure 3-A). Each participant was equipped with an HMD VR headset (HTC Vive Pro Eye), which included Tobii Pro eye-tracking technology, along with motion tracking sensors (Vive Trackers and Leap Motion), as seen in Figure 3-B. Once the VR system was properly set up, participants were instructed to carry out the material handling task at height within the immersive testbed. The task involved lifting, carrying, and lowering a 10-pound concrete block to replicate typical construction activities. To simulate real-world conditions, participants used an actual concrete block, which was integrated into the virtual environment via a tangible user interface. This setup allowed participants to experience a realistic material handling task at height, ensuring the task closely resembled those found on actual construction sites.

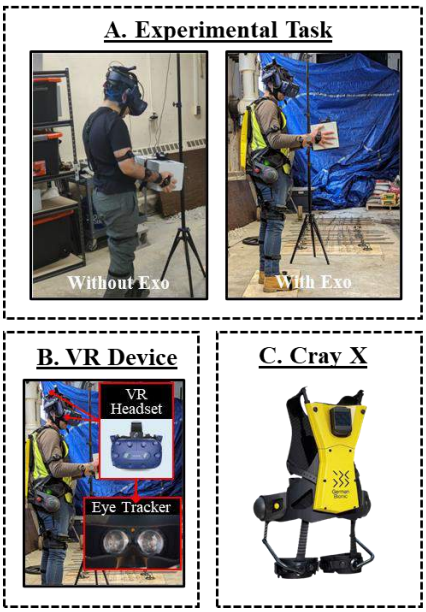


Figure 3. Vigilance Level Assessment Illustration (A) Experimental Task Setup (With and without active exoskeleton); (B) VR Device; and (C) Cray X exoskeleton

In addition, the actual active exoskeleton, Cray X, was integrated into the virtual testbed for the scenario involving material handling tasks with powered exoskeleton assistance. The study employed a crossover design, with participants randomly divided into two groups (nine participants in each group). In the first group, participants began by performing the material handling tasks without the use of the active exoskeleton. After

completing this phase, they crossed over to the second phase, where they repeated the tasks using the Cray X exoskeleton. Conversely, the second group started by using the active exoskeleton and then performed the tasks without it. The Cray X, an active exoskeleton manufactured by German Bionic, weighs approximately 8 kg and was used to provide support during the material handling tasks, as seen in Figure 3-C. The Cray X active exoskeleton, used in this study, is designed to support back and lower body movements without restricting neck mobility. Since the exoskeleton primarily assists with lifting and carrying tasks, its weight distribution does not impose significant constraints on head movements. For both scenarios, participants performed the tasks for 5 minutes, with a 5 -minute break in between sessions. Throughout the experiments, the Tobii eye tracker, integrated with the HMD VR headset, collected data on participants' gaze points, focusing on predefined AOI within the virtual environment. This data was crucial for analyzing the participants' vigilance during the tasks. During the experiment, participants did not report significant discomfort or difficulty in rotating their heads to observe hazards.

The metrics used to assess participants' vigilance while performing tasks in the VR environment include Fixation Count, which measured the number of times participants looked at a predefined AOI; and Fixation Duration, which calculated the total time each participant spent fixating on the AOI. The fixation duration was calculated by subtracting the timestamp when the participant shifted their gaze away from the hazard AOI from the initial timestamp when they first looked at it. A fixation was only recorded if it lasted longer than 100 milliseconds. Both fixation count and fixation duration serve as indicators of vigilance, providing insights into how often and for how long participants were distracted by the hazards in the virtual environment.

To assess the impact of active exoskeleton systems on the vigilance levels of construction workers during typical tasks, the extracted eye-tracking metrics were analyzed statistically. The normality of the data was evaluated using the Shapiro-Wilk test. In this test, a non-significant p-value (generally greater than 0.05) suggests that the data does not significantly deviate from a normal distribution. However, the test yielded a p-value of less than 0.05, indicating that the data significantly deviated from normality. Given that the data did not meet the assumptions of normality, a non-parametric test, specifically the Mann-Whitney U test, was employed to examine the effect of the active exoskeleton on the vigilance levels of the construction workers.

5 Results and Discussion

Table 1 presents the results for the calculated fixation count and fixation duration on AOI during the experiment. When using the active exoskeleton, participants exhibited an average fixation count of 59.85 ± 21.3 and spent an average of 43.8 ± 26.65 seconds fixating on the AOI. This suggests that while performing the material handling tasks, participants shifted their attention away from the task to focus on potential hazards in the environment at least 60 times, for a total of approximately 44 seconds during the 10-minute session in the immersive environment. In contrast, when the tasks were performed without the active exoskeleton, the mean fixation count on AOI was slightly lower at 54.35 ± 16.8, with participants spending an average of 39.85 ± 21.25 seconds focusing on the hazards. These findings indicate that, regardless of whether the exoskeleton was used, participants allocated a significant portion of their attention to environmental hazards, but the exoskeleton may have influenced the frequency and duration of these attentional shifts. The fixation count was slightly higher when participants used the exoskeleton. This suggests that participants may have been more aware of their surroundings or potential hazards while wearing the device. However, the difference is relatively small, and the high standard deviations indicate considerable variability among participants. Similarly, participants spent slightly more time fixating on the hazards when wearing the active exoskeleton. This could indicate that the active exoskeleton might have influenced participants to be more cautious or attentive to potential hazards. Again, the high standard deviations suggest significant individual differences.

Table 1 Results of eye-tracking metrics for AOI

Metrics	With exo	Without exo
Fixation Count	59.85 $\pm$ 21.3	54.35 $\pm$ 16.8
Fixation Duration	43.8 $\pm$ 26.65	39.85 $\pm$ 21.25

To further evaluate the statistical significance of the differences observed between the two conditions (with and without the active exoskeleton), a Mann-Whitney U test was conducted at a significance level of 0.05. The results indicated that the p-values for both the fixation count, and fixation duration metrics were greater than 0.05. This suggests that the differences in attentional allocation between the two conditions, in terms of how frequently participants fixated on the AOI and how long they spent focusing on it, were not statistically significant.

In both conditions, participants demonstrated a high level of attention to potential hazards, with approximately 54-60 fixations on hazards during the 10-minute session. This suggests that participants maintained a good level of safety awareness regardless of exoskeleton use. While the exoskeleton condition showed slightly higher fixation counts and durations, the

lack of statistical significance suggests that the device did not dramatically alter participants' visual attention patterns. The results indicate that the active exoskeleton doesn't significantly affect the vigilance level of construction workers during common construction tasks.

The slight increase in fixation counts and duration when using the exoskeleton could reflect a reallocation of cognitive resources, as suggested by Wickens' Multiple Resource Theory (MRT) [23,24]. According to MRT, cognitive resources are distributed across multiple modalities (visual, auditory) and processing stages (perception, cognition). The exoskeleton introduced additional sensory inputs, such as vibrations and altered proprioception, which could have competed for cognitive resources typically dedicated to visual processing and hazard detection. However, the lack of statistically significant differences in fixation metrics suggests that the exoskeleton's cognitive demands did not exceed participants' available resources, allowing them to maintain vigilance effectively. From the perspective of Endsley's Situation Awareness Model [25], which emphasizes perception, comprehension, and projection of environmental elements, participants demonstrated consistent attention to predefined AOIs representing hazards. This indicates that they effectively maintained all three levels of situation awareness—perceiving, comprehending, and anticipating hazards—despite the exoskeleton's sensory inputs. The slight increase in fixation metrics when using the exoskeleton may reflect heightened caution or attentiveness, but the high standard deviations suggest significant individual differences in how participants adapted to the device.

Despite the insightful findings, this study has several limitations that could be explored in future research endeavors. Firstly, the duration of the study (10 minutes) limits the ability to capture potential long-term effects of exoskeleton use on vigilance levels. Future studies should investigate whether prolonged use of exoskeleton leads to different outcomes in terms of hazard awareness and visual attention. Secondly, the sample size in this study may not accurately represent the full range of the construction workforce, which varies in age, gender, physical fitness, and experience. To improve the relevance of the findings, future research should include a larger and more diverse group of participants. This would help ensure the results are applicable to a broader segment of the construction industry. While the block-handling tasks used in this study are representative of common material handling activities in construction, they may not fully capture the characteristics of other construction tasks with different physical and cognitive demands. Future research should investigate the impact of exoskeleton use on vigilance during a broader range of construction tasks, including those requiring fine motor skills, spatial awareness, or dynamic hazard detection. Likewise, incorporating tasks with varying levels of complexity and collaboration (e.g., working with robotic systems or coordinating with other workers) would provide a more comprehensive understanding of how exoskeletons influence vigilance across different construction scenarios

Additionally, while this study focused on active exoskeletons, future research could compare the effects of active and passive exoskeletons on vigilance levels. Passive exoskeletons, which rely on mechanical components like springs or dampers rather than powered actuators, may introduce different cognitive and ergonomic challenges. A comparative analysis could provide deeper insights into how different types of exoskeletons influence workers' attention and safety awareness, ultimately informing the development of guidelines for their safe and effective use in construction settings. The 10-minute task duration in this study was chosen to establish a controlled and manageable experimental framework for assessing the immediate effects of active exoskeleton use on vigilance levels. While this duration allowed us to minimize participant fatigue and ensure consistency across trials, it lacks in reflecting the conditions of real construction work shifts, where vigilance effects may be cumulative over time due to prolonged exposure to physically and mentally demanding tasks. Future studies should investigate the effects of exoskeleton use over longer durations to provide a more comprehensive understanding of how vigilance evolves during extended work periods.

6 Conclusion

This study evaluated the impact of active exoskeleton systems on the vigilance levels of construction workers during common construction tasks. Eighteen subjects were required to complete material handling tasks under two conditions (with and without using an active back-support exoskeleton) in immersive virtual environments. Using eye-tracking metrics—specifically fixation count and fixation duration on AOI representing potential hazards in the simulated virtual environment, the results demonstrated that while participants showed slightly higher fixation counts and longer fixation durations when using the active exoskeleton, these differences were not statistically significant, as confirmed by the Mann-Whitney U test. Participants consistently maintained a strong level of attention to hazards, with approximately 54-60 fixations during the 10-minute task sessions, regardless of whether they used the exoskeleton. This suggests that the active exoskeleton did not distract workers or negatively impact their vigilance levels. Although there was a small increase in the number and duration of fixations when participants used the exoskeleton, the lack of statistical significance indicates

that the device did not substantially influence workers' vigilance. These findings provide important insights into the psychological effects of active exoskeletons, indicating that their use does not diminish workers' vigilance levels to potential safety risks on construction sites. The findings could help formulate standard guidelines for the adoption of active exoskeletons in construction sites. While this study was conducted with a relatively small sample size, future investigations could recruit a larger pool of subjects to further generalize results. Future studies should include a larger and more diverse group of participants to enhance the reliability of the results and provide more robust insights into the impact of exoskeletons on vigilance levels. Further research is necessary to examine long-term effects and ensure these devices support overall worker performance and safety. While the VR environment provided a controlled setting for studying vigilance, it cannot fully replicate the complexity of real-world construction sites, including factors such as site noise, terrain unevenness, unpredictable distractions, and multi-worker coordination. These variables could significantly affect vigilance levels in practice. Future studies could validate the findings in actual construction settings to provide a more comprehensive understanding of how exoskeletons impact vigilance in dynamic and unpredictable environments. While the VR environment provided a controlled setting for studying vigilance, it cannot fully replicate the complexity of real-world construction sites, including factors such as site noise, terrain unevenness, unpredictable distractions, and multi-worker coordination. These variables could significantly affect vigilance levels in practice. Even though the Mann-Whitney U test was appropriate for this study given the non-normal distribution of the data and the exploratory nature of the research, more sophisticated statistical methods, such as mixed-effects models, could provide deeper insights into the effects of exoskeleton use on vigilance levels.

Acknowledgements

This material is based upon work supported by the National Science Foundation Award (Grant No. IIS-2410255, “Understanding Underlying Risks and Sociotechnical Challenges of Powered Exoskeleton to Construction Workers” and Grant No. DUE- 2402008, “Investigating the Impact of an Immersive VR-based Learning Environment for Learning Human-Robot Collaboration in Construction Robotics Education”). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] Valero E, Sivanathan A, Bosché F, Abdel-Wahab M. Musculoskeletal disorders in construction: A review and a novel system for activity tracking with body area network. Appl Ergon 2016;54:120–30. https://doi.org/10.1016/J.APERGO.2015.11.020.

[2] Wang D, Asce SM, Dai F, Asce M, Ning X. Risk Assessment of Work-Related Musculoskeletal Disorders in Construction: State-of-the-Art Review. J Constr Eng Manag 2015;141:04015008. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000979.

[3] Kazar G, Comu S. Exploring the relations between the physiological factors and the likelihood of accidents on construction sites. Engineering, Construction and Architectural Management 2021. https://doi.org/10.1108/ECAM-11-2020-0958/FULL/XML.

[4] Passmore D, Chae C, Borkovskaya V, Baker R, Yim JH. Severity of U.S. Construction Worker Injuries, 2015-2017. E3S Web of Conferences 2019;97:06038. https://doi.org/10.1051/E3SCONF/20199706038.

[5] Bureau of Labor Statistics. Nonfatal Occupational Injuries and Illnesses Requiring Days Away From Work. 2020. https://doi.org/https://data.bls.gov/PDQWeb/cs.

[6] OSHA Small Business | Safety Pays Program - Estimator | Occupational Safety and Health Administration n.d.

[7] Antwi-Afari MF, Li H, Edwards DJ, Pärn EA, Seo J, Wong AYL. Biomechanical analysis of risk factors for work-related musculoskeletal disorders during repetitive lifting task in construction workers. Autom Constr 2017;83:41–7. https://doi.org/10.1016/j.autcon.2017.07.007.

[8] Shein Esther. Exoskeletons today. Commun ACM 2019;62:14–6. https://doi.org/10.1145/3303851.

[9] Ojha A, Gautam Y, Jebelli H, Akanmu A. Physiological impact of powered back-support exoskeletons in construction: Analyzing muscle fatigue, metabolic cost, ergonomic risks, and stability. Autom Constr 2024;168:105742. https://doi.org/10.1016/J.AUTCON.2024.105742.

[10] Shayesteh S, Ojha A, Liu Y, Jebelli H. Evaluating the Impact of Powered Back-Support Exoskeletons and Virtual Reality Interventions on Gait Stability of Construction Workers at

Heights. Computing in Civil Engineering 2023: Resilience, Safety, and Sustainability - Selected Papers from the ASCE International Conference on Computing in Civil Engineering 2023, American Society of Civil Engineers; 2024, p. 392–9. https://doi.org/10.1061/9780784485248.047.

[11] Ojha A, Guo H, Jebelli H, Martin A, Akanmu A. Assessing the Impact of Active Back Support Exoskeletons on Muscular Activity during Construction Tasks: Insights from Physiological Sensing. Computing in Civil Engineering 2023: Resilience, Safety, and Sustainability - Selected Papers from the ASCE International Conference on Computing in Civil Engineering 2023, American Society of Civil Engineers; 2024, p. 340–7. https://doi.org/10.1061/9780784485248.041.

[12] Shayesteh S, Ojha A, Gautam Y, Jebelli H, Akanmu A. Evaluating the Impact of Back-Support Exoskeletons on Construction Workers' Fall Risk. Computing in Civil Engineering 2024 - Selected Papers from the ASCE International Conference on Computing in Civil Engineering, 2024.

[13] Liu Y, Asce SM, Ojha A, Jebelli H, Asce AM, Candidate PD, et al. Vision-Based Ergonomic Risk Assessment of Back-Support Exoskeleton for Construction Workers in Material Handling Tasks. Computing in Civil Engineering 2023: Resilience, Safety, and Sustainability - Selected Papers from the ASCE International Conference on Computing in Civil Engineering 2023 2024:331–9. https://doi.org/10.1061/9780784485248.040.

[14] Ojha A, Gautam Y, Zhao X, Sharifironizi A, Jebelli H, Matin A. Understanding the Impact of an Exoskeleton System on Metabolic Cost of Workers During Common Repetitive Construction Tasks. Computing in Civil Engineering , Pittsburgh: ASCE; 2025.

[15] Kulyk O, Van Der Veer G, Van Dijk B. Situational Awareness Support to Enhance Teamwork in Collaborative Environments. Proceedings of the 15th European Conference on Cognitive Ergonomics the Ergonomics of Cool Interaction - ECCE '08 n.d. https://doi.org/10.1145/1473018.

[16] Kass SJ, Cole KS, Stanny CJ. Effects of distraction and experience on situation awareness and simulated driving. Transp Res Part F Traffic Psychol Behav 2007;10:321–9. https://doi.org/10.1016/J.TRF.2006.12.002.

[17] Weisberg RW, Reeves LM. Cognition from memory to creativity. Cognition from Memory to Creativity 2013;1:354.

[18] Liao PC, Sun X, Zhang D. A multimodal study to measure the cognitive demands of hazard recognition in construction workplaces. Saf Sci 2021;133:105010. https://doi.org/10.1016/j.ssci.2020.105010.

[19] Hancock PA, Billings DR, Schaefer KE. Can You Trust Your Robot?: Ergonomics in Design: The Quarterly of Human Factors Applications 2011;19:24–9. https://doi.org/10.1177/1064804611415045.

[20] Sgarbossa F, Grosse EH, Neumann WP, Battini D, Glock CH. Human factors in production and logistics systems of the future. Annu Rev Control 2020;49:295–305. https://doi.org/10.1016/J.ARCONTROL.2020.04.007.

[21] Zhao X, Jebelli H. A computational method for real-time roof defect segmentation in robotic inspection. Computer-Aided Civil and Infrastructure Engineering 2025. https://doi.org/10.1111/MICE.13471.

[22] Wang D, Chen J, Zhao D, Dai F, Zheng C, Wu X. Monitoring workers' attention and vigilance in construction activities through a wireless and wearable electroencephalography system. Autom Constr 2017;82:122–37. https://doi.org/10.1016/J.AUTCON.2017.02.001.

[23] Wickens CD. Multiple resources and performance prediction. Theor Issues Ergon Sci 2002;3:159–77. https://doi.org/10.1080/14639220210123806.

[24] Wickens CD. Multiple Resources and Mental Workload. Human Factors: The Journal of the Human Factors and Ergonomics Society 2008;50:449–55. https://doi.org/10.1518/001872008X288394.

[25] Endsley MicaR. Toward a theory of situation awareness in dynamic systems. Human Factors: The Journal of the Human Factors and Ergonomics Society 1995;37:32–64. https://doi.org/10.1518/001872095779049543.

Personalized and Adaptive Virtual Reality Training for Physically Coupled Robots in Construction

Yuming Zhang[1], Amit Ojha[1], Shayan Shayesteh[1] and Houtan Jebelli[1]

[1] Department of Civil and Environmental Engineering, University of Illinois Urbana-Champaign, Champaign, IL, USA

yumingz4@illinois.edu, amito2@illinois.edu, shayan@illinois.edu, hjebelli@illinois.edu

Abstract –

Work-related musculoskeletal disorders (WMSDs) significantly impact worker health and productivity in construction. Physically coupled robotics, such as exoskeletons, offer a promising solution by augmenting human capabilities and reducing strain. However, their effective use requires precise human-robot synchronization, which can be challenging in dynamic construction environments. Traditional training methods often fail to address these complexities, leading to suboptimal performance, misuse, or increased injury risk. To overcome these challenges, this study develops a personalized Virtual Reality (VR) training platform incorporating an Individualized Bayesian Knowledge Tracing (iBKT) framework. The system continuously evaluates user performance in key body regions, waist, knee, and shoulder, using real-time metrics to dynamically adjust task difficulty, training frequency, and instructional feedback. This adaptive approach tailors training content to enhance motor coordination and postural control in exoskeleton-assisted tasks. Results indicate that ATG achieved greater reductions in postural errors across tasks, with total ER decreasing by 6, 3, and 5 points for walking, bending, and squatting, respectively, compared to 4, 2, and 5 points in TTG. These findings underscore the importance of adaptive, personalized training in enhancing human-robot interaction, safety, and productivity. By equipping workers with tailored skills, this study advances workforce-robot collaboration, fostering more resilient and efficient construction practices.

Keywords –

Physically coupled robots; Adaptive training; Virtual Reality

1 Introduction

Work-related musculoskeletal disorders (WMSDs) represent a significant health concern in the construction industry, where physically demanding tasks are routine. Over 20% of nonfatal injuries in construction result from WMSDs, highlighting the heightened physical strain and risks associated with construction work [1]. Research shows that over 43% of construction workers report chronic musculoskeletal pain, which has been directly linked to reduced productivity, higher rates of absenteeism, and escalating healthcare costs [2]. Furthermore, the physically intensive nature of construction tasks exacerbates the risk of cumulative trauma disorders, impacting not only the physical well-being of workers but also the efficiency and cost-effectiveness of construction operations [3]. Addressing these challenges with innovative technological solutions has thus become a priority to improve both worker health and productivity in this high-risk industry.

Physically coupled robotic systems, such as exoskeletons and collaborative lifting aids, offer promising solutions to mitigate the high incidence of WMSDs within construction[4]. By directly enhancing human strength and endurance, these systems support workers in performing physically demanding tasks, such as heavy lifting, repetitive movements, and sustained awkward postures, without the same level of physical strain [5]. Unlike traditional assistive devices, physically coupled robots operate in close synchronization with human operators, offering seamless support that adapts dynamically to the user's movements. Exoskeletons, for instance, designed for leg support can reduce muscular fatigue up to 71.5% [6], a benefit that has shown potential in experimental studies to improve workers' endurance in repetitive tasks. However, fully realizing their benefits depends on highly coordinated human-robot interactions, where human movement is synchronized with robotic support to optimize safety and productivity [7].

Unlike robots in controlled industrial settings, working with physically coupled robots on construction sites requires workers to adapt control to unpredictable, site-specific conditions [8]. Successful operation of these robots relies on seamless coordination between humans and machines, where workers need the skills to adjust movement patterns, manage force distribution, and

respond dynamically to real-time construction challenges. This complex, close-coupling demands specialized training to ensure workers are equipped to handle the nuanced control and adaptability needed to maximize the potential of these robots.

Traditional training programs are generally characterized by their formulaic and rigid structure, focusing on standardized methods and broad, theory-based instruction. These programs often emphasize repetition and adherence to fixed protocols, which can be effective for routine tasks but fall short in addressing the complexities of dynamic and evolving work environments. A key limitation of such approaches is their lack of flexibility and adaptability, making them ill-suited for preparing workers to navigate the nuanced demands of human-robot interaction in physically coupled systems. While theoretical robotic operations may be addressed, these programs often lack the targeted specificity and adaptability essential for real-world scenarios that demand precise, synchronized human-robot collaboration required by physically coupled robots in close-coupling applications. This gap highlights an urgent need for research-driven training solutions that go beyond traditional approaches, offering adaptable training methods tailored to worker's skills and progress.

In response, this paper presents an adaptive VR-based Human Robot Interaction (HRI) training platform specifically designed for physically coupled robots in construction. Utilizing an Individualized Bayesian Knowledge Tracing (iBKT) framework, the platform continuously evaluates each worker's performance on essential skills, dynamically adjusting the training content and pace in real-time. This adaptive approach creates a personalized learning experience that enables workers to develop a deeper understanding of the precise techniques required for effective close coupling with physically coupled robots. By simulating construction scenarios within a virtual environment, this adaptive training approach allows workers to build practical HRI skills and refine their interaction techniques with physically coupled robots, ultimately facilitating a safer and more efficient human-robot partnership on construction sites. Through this adaptive VR-based training platform, the study aims to demonstrate the effectiveness of personalized, adaptive training in reducing safety risks, and advancing the integration of physically coupled robots in construction.

2 The Emergence of Physically Coupled Robots in Construction: Amplifying Human Capability

Physically coupled robots are specifically engineered to operate in direct physical contact with human operators, enhancing human strength, endurance, and precision in ways that expand the range of tasks workers can perform [9]. Unlike traditional industrial robots, which function autonomously or semi-autonomously, physically coupled robots are deeply integrated with the user, forming a symbiotic relationship that merges human intuition with robotic power to increase productivity and efficiency [10]. Exoskeletons are a prominent and widely adopted example of physically coupled robots. These wearable devices are designed to augment human physical abilities by providing mechanical support during strenuous tasks [11]. Working in close coordination with the human body, exoskeletons function as an extension of the user's own movements, enhancing strength, endurance, and precision. This makes them particularly valuable in environments involving repetitive, heavy, or physically demanding tasks.

The construction industry is especially positioned to benefit from this form of human-robot collaboration. Given the labor-intensive and physically demanding nature of many construction tasks, physically coupled robots can significantly reduce physical strain, minimize injury risk, and improve operational efficiency. Tasks that require heavy lifting, repetitive actions, or prolonged exertion can be performed more safely and effectively with exoskeleton assistance [12]. These devices amplify the user's strength, stabilize posture, and reduce fatigue, ultimately enhancing task performance and reducing physical stress on workers.

However, the integration of physically coupled robotics, such as exoskeletons, into construction introduces challenges [13]. Unlike conventional assistive tools, these systems require active human engagement, demanding continuous adjustment and synchronization of worker movements with robotic assistance. Improper usage, such as misaligned movement patterns or incorrect force distribution, can result in inefficient task execution, discomfort, or heightened physical strain, ultimately diminishing the intended benefits of these technologies. Consequently, equipping workers with the requisite motor coordination, adaptive control strategies, and cognitive decision-making skills is crucial for the safe and effective deployment of these systems. To achieve this, structured training programs are indispensable, ensuring that workers receive personalized, responsive instruction that enhances their ability to integrate these robotic systems seamlessly into dynamic construction workflows. Traditional training programs, typically formulaic and rigid, lack the specificity needed for these dynamic environments [14]. This limitation highlights the need for adaptive training solutions that personalize the learning experience, equipping workers with the skills necessary for effective interaction with wearable robotic systems in real-world construction scenarios.

3 The Need for Adaptive Training in Physically Coupled Robotics

Adaptive training is defined as a method in which the problem, stimulus, or task adjusts dynamically based on the trainee's performance [15]. Unlike conventional, static approaches, adaptive training dynamically customizes adaptive variables through a defined adaptive logic that continuously monitors performance metrics. This real-time adjustment creates a highly personalized learning experience, enhancing both efficiency and effectiveness. Performance metrics are indicators of a trainee's current skill level and progress, such as movement accuracy, reaction times, task completion rates, and overall accuracy. These metrics are gathered either through ongoing assessments that provide immediate feedback or post-session evaluations to measure overall progress. These metrics feed into the adaptive logic, which determines how and when to adjust adaptive variables, such as simulation difficulty or feedback type, to keep training optimally aligned with the trainee's evolving needs [16].

Adaptive training methodologies have found widespread application in industries including construction, adaptive training has been employed to improve worker safety, skill development, and operational efficiency. Studies have developed personalized safety training programs that dynamically adjust hazard recognition tasks based on workers' prior performance and learning progress, leading to enhanced hazard identification and risk mitigation capabilities [17]. In labor-intensive construction tasks, adaptive methods have been utilized to modify task difficulty in real time based on ergonomic assessments, resulting in improved lifting techniques and a reduction in musculoskeletal strain [18]. Additionally, cognitive training frameworks in construction have employed adaptive approaches to tailor problem-solving complexity to individual proficiency levels, thereby enhancing situational awareness and decision-making in dynamic work environments.

Previous research has demonstrated that adaptive training enhances learning efficiency and y, knowledge retention, and task-specific skill acquisition by dynamically adjusting instruction based on user performance metrics. Given the specific demands of exoskeleton training, such as the need for precise movement synchronization, optimized force distribution, and adaptive cognitive control, workers could greatly benefit from adaptive training approaches that tailor instruction to their biomechanics and real-time interaction with robotic assistance. However, research specifically focusing on adaptive training for physically coupled exoskeleton systems remains limited.

This study aims to develop an adaptive training solution to enhance human-robot collaboration in physically coupled robotics for construction. Given the need for training that adapts to individual learning progress, dynamically modifies task complexity, and provides real-time feedback on movement synchronization and force distribution, Virtual Reality (VR) serves as an effective platform [19]. Its immersive and interactive environment enables controlled, data-driven adjustments to training parameters within diverse construction scenarios that enhance adaptability.

To achieve this, the research focuses on developing a framework that continuously monitors, evaluates, and adjusts training variables based on real-time performance metrics, including movement synchronization, force distribution, and task-specific coordination. By integrating adaptive logic with VR, this study seeks to replicate the dynamic conditions of construction sites, addressing a critical gap in training methodologies and facilitating the broader adoption of wearable robotics in the industry.

4 Method: Development of Adaptive Training Platform for Enhancing Human Physically Coupled Robot Interaction in Construction

The VR-based adaptive training system for worker-physically coupled robot interaction is a data-driven approach to improve learning efficiency and optimize task mastery. The system relies on performance metrics to capture real-time motion data from user movements, which are processed by the iBKT-based adaptive logic to analyze the user's task mastery probabilities (Figure 1). When these probabilities exceed a predefined threshold, the system dynamically adjusts adaptive variables, such as task difficulty and task frequency, within the VR training environment. These adjustments are tailored to challenge the user appropriately while preventing skill stagnation, ensuring that users can master the required skills effectively and efficiently.

4.1 Performance Metrics

Key performance metrics were established to evaluate workers' capabilities in worker-physically coupled robot interaction, with a focus on movement accuracy as indicators of effective collaboration. Body angles were selected as the primary indicator of movement accuracy, measured through motion tracking sensors from Inertial Measurement Units (IMUs) and Lighthouse Tracking system (LTS) placed on key body segments, including the waist, knee and shoulder. These IMU sensors continuously monitor, enabling real-time analysis of worker posture and alignment within the robotic system. To mitigate drift errors commonly associated with IMUs, a sensor fusion approach

integrating LTS data was employed to provide periodic corrections, ensuring long-term accuracy. Additionally, drift compensation algorithms, including zero-velocity updates and complementary filtering, were applied to correct accumulated errors over time. To accurately assess movement accuracy, joint angles at critical body points are calculated using vector-based computations derived from IMU data. For each body segment $\vec{u}$, it's calculated by determining the difference between the three-dimensional coordinates of the segment's endpoints $\overrightarrow{P_1}$ and $\overrightarrow{P_2}$ as $\vec{u} = \overrightarrow{P_1} - \overrightarrow{P_2}$. With segment vectors established for adjacent body parts, the joint angle θ between two connected segments $\vec{u}$ and $\vec{v}$ is then computed using the cosine rule $\theta = arccos(\frac{\vec{u}\cdot\vec{v}}{|\vec{u}||\vec{v}|})$.

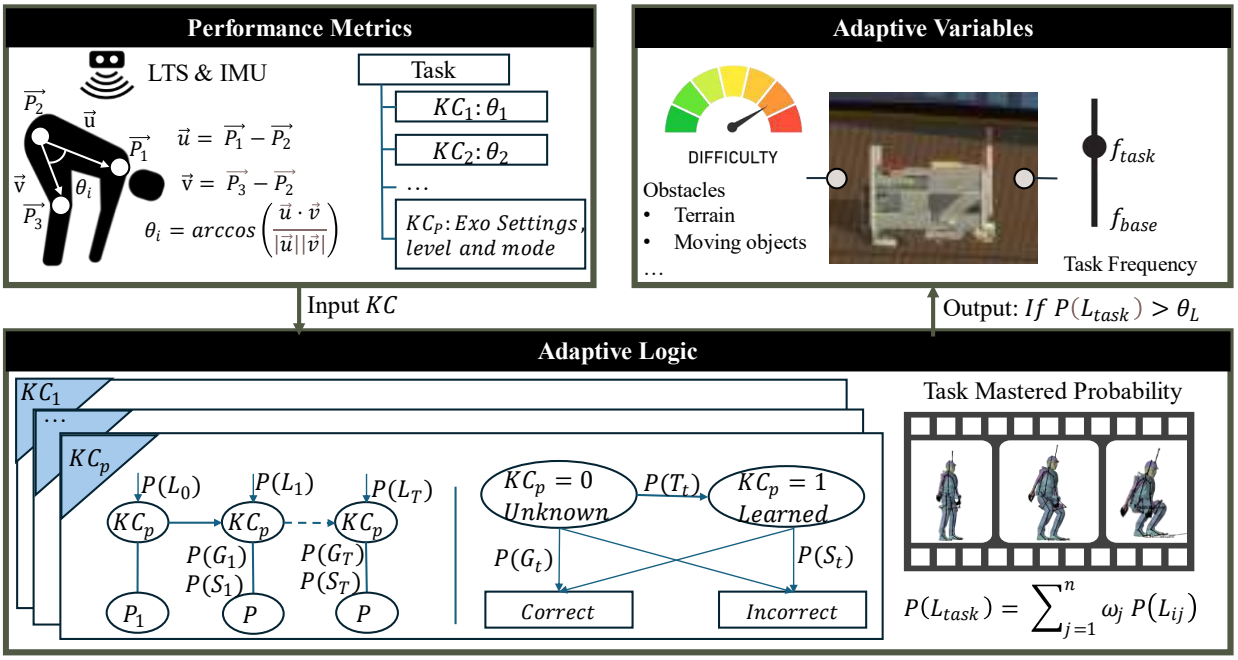


Figure 1. Overview of the developed VR-based adaptive training syste

4.2 Adaptive Logic

Effective adaptive training for physically coupled robotics requires a framework that continuously monitors user performance, dynamically adjusts training parameters, and personalizes learning trajectories. Bayesian Knowledge Tracing (BKT) is well-suited for this application as it models skill acquisition by estimating a learner's mastery probability and updating it based on observed performance. However, standard BKT assumes fixed learning and error probabilities, limiting its responsiveness to real-time changes in human-robot interaction. To address this, we propose an individualized BKT (iBKT) model, which dynamically updates learning probabilities based on real-time performance metrics, refining adaptation to individual learning curves and optimizing skill acquisition in exoskeleton training

The proposed iBKT model extends the traditional BKT framework by enabling real-time adaptation of key parameters based on observed learner performance, thus catering to individual characteristics. The training content is divided into knowledge components (KCs), where each KC represents key body angles extracted from the performance matrix, with a value of KC=1 indicating mastery and KC=0 indicating non-mastery. These components are selected according to ergonomic and task-specific requirements.

A preliminary test is administered in which participants complete a set of baseline tasks designed to evaluate their prior knowledge and motor coordination relevant to exoskeleton operation. This test is used to initialize four critical parameters within the iBKT model for each KC, including the initial probability that user i knows KC j $P(L_{0_{ij}})$; the likelihood of transitioning from a state of non-mastery to mastery of learning KC j after an interaction $P(T_{ij})$; the guess probability $P(G_{ij})$, which is the probability of a correct response to a KC j-related question despite lack of mastery; and slip probability, $P(S_{ij})$, which is the probability of an incorrect response despite mastery of the KC j. Due to the inherent uncertainty in the guessing and slipping behaviors, $P(G_{ij})$ and $P(S_{ij})$ are assumed to follow Beta distributions, allowing for flexible and continuous adjustment based on learner performance.

After initialization, these parameters are continually updated based on the learner's real-time performance to individualize the training process. The guess and slip probabilities, $P(G_{ij})$ and $P(S_{\text{ij}})$, are updated via Bayesian inference, refining these parameters based on learner performance over time. For $P(S_{\text{ij}})$, if the learner has mastered the knowledge but makes an incorrect response (indicating a slip), the posterior distribution is updated as

$$S|data \sim Beta(\alpha_S + f, \beta_S + m - f) \quad (1)$$

where m represents the total number of responses in the mastered state, and f is the number of incorrect responses. Similarly, $P(G_{\text{ij}})$ is updated when the learner is in an unmastered state and correctly guesses the answer, with the posterior distribution.

$$G|data \sim Beta(\alpha_G + g, \beta_s + n - g) \quad (2)$$

where n is the total number of responses in the unmastered state, and g represents correct responses within that state.

The learning probability is $P(T_{ij})$ dynamically adjusted based on recent task performance. If the learner's recent accuracy R meets or exceeds a predefined target accuracy R_{target}, $P(T_{ij})$ is incremented by a set step size ΔT. Conversely, if R falls below R_{target}, $P(T_{ij})$ is reduced by ΔT. $P(T_{ij})$ is updated as

$$P(T_{ij}) = \min(0, \max(1, P(T_{ij}) + \Delta T \cdot sign(R - R_{target}))) \quad (3)$$

Finally, the mastery probability $P(L_{0_{ij}})$ is updated at each time step t through the forward algorithm. Given a learner's response at time t, the probability of correctly answering a question related to KC j, denoted as $P\left(C_{t_{ij}}\right)$, is calculated by

$$P\left(C_{t_{ij}}\right) = P\left(L_{t_{ij}}\right) \times \left(1 - P(S_{ij})\right) + \left(1 - P\left(L_{t_{ij}}\right)\right) \times P(G_{ij}). \quad (4)$$

Following each response, $P\left(L_{t+1_{ij}}\right)$ is updated based on whether the answer was correct or incorrect. For a correct response, the updated mastery probability is $P\left(L_{t+1_{ij}} \middle| observation = C_{t_{ij}}\right) = \frac{P\left(L_{t_{ij}}\right) \times \left(1 - P(S_{ij})\right)}{P\left(C_{t_{ij}}\right)}$. Conversely, if the response is incorrect, the update is given by $P\left(L_{t+1_{ij}} \middle| obs = \neg C_{t_{ij}}\right) = \frac{P\left(L_{t_{ij}}\right) \times P(S_{ij})}{P\left(\neg C_{t_{ij}}\right)}$. The final mastery probability $P\left(L_{t+1_{ij}}\right)$ after time t is calculated as

$$P\left(L_{t+1_{ij}}\right) = P\left(L_{t+1_{ij}} \middle| obs\right) + P(T_{ij}) \times \left(1 - P\left(L_{t+1_{ij}} \middle| obs\right)\right) \quad (5)$$

For a specific human-robot interaction task, the overall task mastery probability $P(L_{task})$ is computed by aggregating the mastery probabilities across all relevant KCs required for the task as $P(L_{task}) = \sum_{j=1}^{n} \omega_j P\left(L_{ij}\right)$, where ω_j is the weight of KC j, satisfying the condition $\sum_{j=1}^{n} \omega_j = 1$.

4.3 Adaptive Variables

The training platform dynamically adjusts certain human-physically coupled robot interaction task difficulty and frequency based on the user's performance on certain task, optimizing the training experience to support mastery in human-robot interaction under realistic conditions. When the learner's mastery probability $P(L_{task})$ for a specific task reaches or exceeds a predefined threshold θ_T, indicating sufficient proficiency, the model increases task difficulty by introducing additional construction-related obstacles, such as uneven terrain or moving objects, to simulate the dynamic and unpredictable nature of a real-world construction environment. This adjustment helps further challenge the learner and prevent skill stagnation. To ensure task diversity and avoid overexposure to a single task, the model employs a rotation mechanism that varies task frequency f_{task} based on mastery levels. Task frequency f_{task} is adjusted according to the function.

$$f_{task} = f_{base} \cdot (1 - \alpha \cdot \frac{P(L_{task}) - \theta}{1 - \theta}) \quad (6)$$

where f_{base} is the base frequency of task rotation, and α is a scaling factor that controls how rapidly the task frequency decreases as $P(L_{task})$ approaches 1. This rotation mechanism ensures a balanced distribution of tasks across all relevant KCs, while placing greater emphasis on those areas that require further development, thus promoting well-rounded proficiency.

5 Case Study: Evaluating the Effectiveness of Adaptive HRI Training Platform

The adaptive training platform was developed in a virtual reality (VR) environment using Unity [20] and deployed on the VIVE PRO 2 [21] system. To enable real-time tracking of body movements, SteamVR motion trackers were attached to each participant, allowing the platform to dynamically adjust training variables based on real-time body position data.

To evaluate the effectiveness of the adaptive training

platform for exoskeleton operation, a case study was conducted with a diverse group of 30 participants with varying backgrounds and skill levels. Participants ranged from novice users with no prior exoskeleton experience to experienced operators. To control prior knowledge, participants were evenly divided into two groups based on their experience: the Adaptive Training Group (ATG), which received training through the iBKT-based adaptive platform, and the Traditional Training Group (TTG), which underwent standard, non-adaptive training. To ensure an accurate assessment of each participant's skill improvement, all participants first completed a pre-assessment to establish baseline task score in exoskeleton operation. The tasks include walking, bending, and squatting. During the process, each participant was instructed to perform the designated tasks while motion tracking sensors recorded joint angles and movement trajectories. The collected data was processed to calculate error rate (ER) metrics, which quantified deviations in posture and movement accuracy based on REBA scoring criteria [22], which assesses postural risks by assigning scores to different body regions. ER is calculated by quantifying deviations in key body joints, including shoulders, waist, and knees, and mapping these deviations to corresponding REBA risk levels. Following the pre-assessment, participants engaged in training according to their group assignment. Upon completing the training, participants were evaluated using the same ER metrics as in the pre-assessment.

6 Results

The effectiveness of the adaptive VR training platform was evaluated by comparing the mastery probability $P(L_{task})$ progression and error rates (ER) between the ATG and TTG across the training period.

6.1 Task Mastery Probability Progression

Figure 2 illustrates the progression of mastery probabilities over the training sessions for a randomly selected participant from the ATG. Three commonly encountered tasks in construction and main body joint mastery probabilities in each task are recorded. The upper panel of the figure (3 × 13 grid) depicts the evolution of the proficiencies of three KCs, $P(L_j)$, and the lower panel (3 × 13 grid) illustrates the progression of trained task proficiencies, $P(L_{Task_i})$, across 13 training sessions. From the figure, tasks with fewer correct attempts appear more frequently in the training sequence, ensuring the participant receives sufficient practice opportunities to reinforce weaker areas. The proficiencies in both panels reveal a clear upward trend, reflecting the gradual improvement in mastery levels for both knowledge components and tasks, driven by the adaptive nature of the training system.

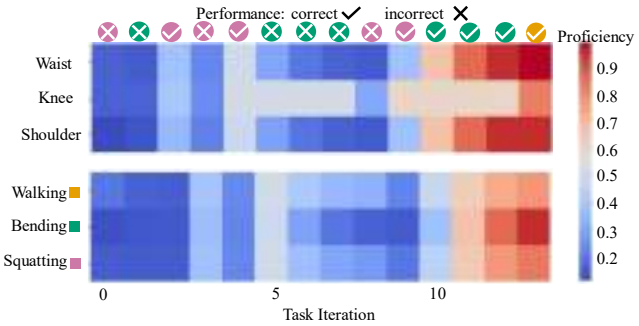


Figure 2. An example of the iBKT-based adaptive training process

6.2 Task Score Comparison

Figure 3 illustrates the pre- and post-training error rates (ER) for the targeted tasks—walking, bending, and squatting, broken down by body parts for the Adaptive Training Group (ATG) and the Traditional Training Group (TTG). For walking, the total ER for ATG decreased from 9 (waist: 7, knee: 2) to 3 (waist: 1, knee: 2), showing a reduction of 6 points. TTG also improved, with total ER decreasing from 8 (waist: 7, knee: 1) to 4 (waist: 3, knee: 1), a reduction of 4 points.

For bending, ATG's total ER decreased from 7 (waist: 2, shoulder: 5) to 4 (waist: 2, shoulder: 2), reflecting a 3-point improvement. TTG also showed a 2-point reduction, with total ER decreasing from 7 (waist: 3, shoulder: 4) to 5 (waist: 3, shoulder: 2). For squatting, ATG's total ER improved significantly, decreasing from 8 (waist: 4, knee: 3, shoulder: 1) to 3 (waist: 1, knee: 1, shoulder: 1), a reduction of 5 points. TTG's total ER decreased from 10 (waist: 4, knee: 4, shoulder: 2) to 5 (waist: 2, knee: 2, shoulder: 1), showing a reduction of 5 points. These results indicate that ATG exhibited greater reductions in postural errors across tasks, particularly in waist and shoulder alignment, with higher total score reductions compared to TTG. This suggests that the adaptive nature of the training platform contributed to more substantial improvements in movement accuracy and postural control. The observed enhancements in waist and shoulder positioning are particularly relevant, as these areas are critical for exoskeleton operation, influencing overall balance and load distribution. The ability of the system to dynamically modify task difficulty and provide real-time feedback likely played a key role in refining postural strategies and improving motor coordination.

Furthermore, paired t-tests indicate that ATG participants exhibited significantly greater improvements in knee ($p=0.0004$), shoulder ($p<0.0001$), and waist ($p=0.0001$) error rates across walking, bending, and squatting tasks, respectively. Independent t-tests further confirm that post-training ER values were significantly lower in ATG than TTG ($p<0.05$), demonstrating the

superior effectiveness of adaptive training in refining movement accuracy and postural stability in exoskeleton operation.

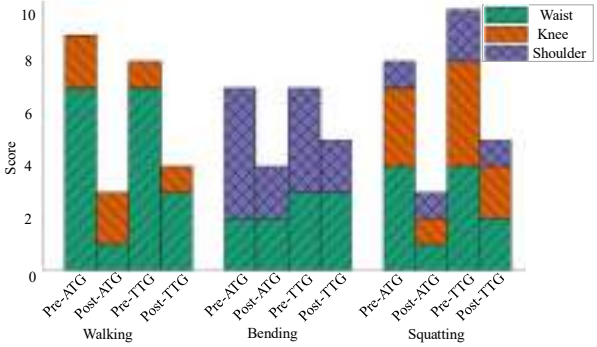


Figure 3. Pre- and Post-Training Task Scores for ATG and TTG

7 Conclusion

This research developed an adaptive training platform using the Individualized Bayesian Knowledge Tracing (iBKT) framework to enhance HRI skills for construction workers operating physically coupled robots, such as exoskeletons. The platform provided personalized, real-time adjustments to training content, effectively addressing individual performance gaps and ensuring efficient skill acquisition. Results demonstrated that the adaptive training led to steady improvement in task mastery and showed significant advantages compared to traditional training. This study underscores the importance of adaptive training in optimizing human-robot collaboration, particularly in dynamic and demanding construction environments.

However, several limitations remain. The current study was conducted in a simulated VR environment without real-world validation using physical exoskeletons, necessitating future studies to assess its practical applicability. Additionally, the system was designed for individual training, and its effectiveness in multi-user collaborative settings has yet to be explored. Long-term retention studies are also required to evaluate whether skills acquired through adaptive VR training persist over time or require.

By integrating real-time feedback and immersive VR simulations, the platform allowed workers to build essential skills safely and effectively. Future research could expand on this approach by exploring other adaptive algorithms and broader applications across different types of physically coupled robots. Ultimately, this adaptive training model represents a promising direction for facilitating the integration of advanced robotics in construction.

Acknowledgements

The work presented in this paper was supported financially by the National Science Foundation (Award No. DUE- 2402008 ‘Investigating the Impact of an Immersive VR-based Learning Environment for Learning Human-Robot Collaboration in Construction Robotics Education’ and Award No. DUE- 2410255 ‘NRI: Understanding Underlying Risks and Sociotechnical Challenges of Powered Wearable Exoskeleton to Construction Workers). Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] “CPWR Musculoskeletal Disorders (MSDs) in Construction.” Accessed: Dec. 01, 2024. [Online]. Available: https://www.cpwr.com/research/data-center/data-dashboards/musculoskeletal-disorders-in-construction/

[2] H. Zong, W. Yi, M. F. Antwi-Afari, and Y. Yu, “Fatigue in construction workers: A systematic review of causes, evaluation methods, and interventions,” *Saf Sci*, vol. 176, p. 106529, Aug. 2024, doi: 10.1016/J.SSCI.2024.106529.

[3] M. F. Antwi-Afari *et al.*, “A science mapping-based review of work-related musculoskeletal disorders among construction workers,” *J Safety Res*, vol. 85, pp. 114–128, Jun. 2023, doi: 10.1016/J.JSR.2023.01.011.

[4] “Automation and Robotics in the Architecture, Engineering, and Construction Industry,” *Automation and Robotics in the Architecture, Engineering, and Construction Industry*, 2022, doi: 10.1007/978-3-030-77163-8.

[5] C. Andrade and D. Nathan-Roberts, “Occupational Exoskeleton Adoption and Acceptance in Construction and Industrial Work: A Scoping Review,” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 66, no. 1, pp. 1325–1329, Sep. 2022, doi: 10.1177/1071181322661415.

[6] Z. Yan, B. Han, Z. Du, T. Huang, O. Bai, and A. Peng, “Development and testing of a wearable passive lower-limb support exoskeleton to support industrial workers,” *Biocybern Biomed*

Eng, vol. 41, no. 1, pp. 221–238, Jan. 2021, doi: 10.1016/J.BBE.2020.12.010.

[7] Y. Liu *et al.*, "Studying the Effects of Back-Support Exoskeletons on Workers' Cognitive Load during Material Handling Tasks," *Construction Research Congress 2024, CRC 2024*, vol. 1, pp. 659–669, 2024, doi: 10.1061/9780784485262.067.

[8] S. Crea *et al.*, "Occupational exoskeletons: A roadmap toward large-scale adoption. Methodology and challenges of bringing exoskeletons to workplaces," *Wearable Technologies*, vol. 2, p. e11, 2021, doi: 10.1017/WTC.2021.11.

[9] A. Ojha, Y. Gautam, H. Jebelli, and A. Akanmu, "Physiological impact of powered back-support exoskeletons in construction: Analyzing muscle fatigue, metabolic cost, ergonomic risks, and stability," *Autom Constr*, vol. 168, p. 105742, Dec. 2024, doi: 10.1016/J.AUTCON.2024.105742.

[10] S. Shayesteh *et al.*, "Evaluating the Impact of Powered Back-Support Exoskeletons and Virtual Reality Interventions on Gait Stability of Construction Workers at Heights," *Computing in Civil Engineering 2023: Resilience, Safety, and Sustainability - Selected Papers from the ASCE International Conference on Computing in Civil Engineering 2023*, pp. 392–399, 2024, doi: 10.1061/9780784485248.047.

[11] A. Ojha, H. Guo, H. Jebelli, A. Martin, and A. Akanmu, "Assessing the Impact of Active Back Support Exoskeletons on Muscular Activity during Construction Tasks: Insights from Physiological Sensing," *Computing in Civil Engineering 2023: Resilience, Safety, and Sustainability - Selected Papers from the ASCE International Conference on Computing in Civil Engineering 2023*, pp. 340–347, 2024, doi: 10.1061/9780784485248.041.

[12] S. J. Baltrusch, F. Krause, A. W. de Vries, and M. P. de Looze, "Arm-support exoskeleton reduces shoulder muscle activity in ceiling construction," *Ergonomics*, Nov. 2023, doi: 10.1080/00140139.2023.2280443.

[13] S. Shayesteh, A. Ojha, Y. Liu, and H. Jebelli, "Human-robot teaming in construction: Evaluative safety training through the integration of immersive technologies and wearable physiological sensing," *Saf Sci*, vol. 159, p. 106019, Mar. 2023, doi: 10.1016/J.SSCI.2022.106019.

[14] M. J. Burke, S. A. Sarpy, K. Smith-Crowe, S. Chan-Serafin, R. O. Salvador, and G. Islam, "Relative effectiveness of worker safety and health training methods," *Am J Public Health*, vol. 96, no. 2, pp. 315–324, Feb. 2006, doi: 10.2105/AJPH.2004.059840.

[15] C. R. Kelley, "What is Adaptive Training?1," *https://doi.org/10.1177/001872086901100602*, vol. 11, no. 6, pp. 547–556, Dec. 1969, doi: 10.1177/001872086901100602.

[16] M. Zahabi and A. M. Abdul Razak, "Adaptive virtual reality-based training: a systematic literature review and framework," *Virtual Real*, vol. 24, no. 4, pp. 725–752, Dec. 2020, doi: 10.1007/S10055-020-00434-W/FIGURES/5.

[17] W. Li, H. Huang, T. Solomon, B. Esmaeili, and L. F. Yu, "Synthesizing Personalized Construction Safety Training Scenarios for VR Training," *IEEE Trans Vis Comput Graph*, vol. 28, no. 5, pp. 1993–2002, May 2022, doi: 10.1109/TVCG.2022.3150510.

[18] Sudip Subedi, Nipesh Pradhananga, Armando Carrasquillo, and Francisco Lopez, "Virtual Reality-based Personalized Learning Environment for Repetitive Labor-intensive Construction Tasks." Accessed: Mar. 13, 2025. [Online]. Available: https://www.researchgate.net/

[19] S. Shayesteh, A. Ojha, and H. Jebelli, "Workers' Trust in Collaborative Construction Robots: EEG-Based Trust Recognition in an Immersive Environment," in *Automation and Robotics in the Architecture, Engineering, and Construction Industry*, Cham: Springer International Publishing, 2022, pp. 201–215. doi: 10.1007/978-3-030-77163-8_10.

[20] "Unity Real-Time Development Platform | 3D, 2D, VR & AR Engine." Accessed: Jan. 22, 2024. [Online]. Available: https://unity.com/

[21] "VIVE Pro 2 Full Kit - High-Resolution PC VR Gaming System." Accessed: Nov. 02, 2024. [Online]. Available: https://www.vive.com/us/

[22] S. Hignett and L. McAtamney, "Rapid Entire Body Assessment (REBA)," *Appl Ergon*, vol. 31, no. 2, pp. 201–205, Apr. 2000, doi: 10.1016/S0003-6870(99)00039-3.

A Digital Twin-based System for the Indoor Environmental Quality Monitoring in Off-Site Construction Facilities

Sena Assaf[1], Mohamed Assaf[1], Xinming Li[2], Ahmed Bouferguene[1], and Mohamed Al-Hussein[1]

[1]Department of Civil and Environmental Engineering, University of Alberta, Canada
[2]Department of Mechanical Engineering, University of Alberta, Canada
sassaf1@ualberta.ca, massaf2@ualberta.ca, xinming.li@ualberta.ca, ahmedb@ualberta.ca, malhussein@ualberta.ca

Abstract

Off-site construction (OSC) is a well-established paradigm for achieving productivity, safety, and sustainability in construction. The key factor in its success is the efficiency of the workforce transforming blueprints into actual buildings and building elements. This has driven efforts to improve working conditions at the production facility, paving the way towards human-centric construction manufacturing. Achieving a safe and healthy indoor environmental quality (IEQ) is among the most critical aspects of this endeavor. Given the dynamic nature of IEQ in OSC facilities, there is a need for real-time performance monitoring as a means of identifying any deviations from IEQ standards and regulations. As such, the objective of this study is to develop a digital twin-based system for real-time monitoring of the IEQ of OSC facilities with continuous management and alerting features. The proposed *O-IEQ Alert System* encompasses four phases: (1) IEQ metrics identification, (2) data acquisition, (3) digital modelling, and (4) development of the IEQ alerting component. The system is successfully applied in a case study of a wall-production manufacturing facility. A web-based alert system is developed that presents real-time monitoring of the IEQ indices, including Air Quality Index, thermal comfort Predicted Mean Vote Index, and the Sound Level Index for each workstation along with the time-series trends. In case any of the indices does not fall within the defined threshold, an alert is issued to management to take the necessary corrective and preventive measures. The ultimate aim is to provide a safe, healthy, and productive working environment in OSC facilities.

Keywords

Indoor Environmental Quality; Digital Twin; Alert System; Off-site Construction; Human-centricity

1 Introduction and Literature Review

The off-site construction (OSC) method has emerged as an alternative to conventional construction approaches in an effort to industrialize the construction industry. In OSC, building elements are fabricated in off-site production facility and then shipped to the construction site for installation and assembly [1]. It is well-established as a successful approach to achieving a faster, more productive, safer, and more sustainable construction as compared to conventional methods [1].

Typically, an individual spends about a third of their life at the workplace [2] emphasizing the need to create safe and healthy working environments, especially that employees tend to carry their workplace experiences into their personal lives [3]. In OSC, the production facility acts as the workers' workplace as they undertake different activities of the building fabrication process. The setup in a controlled factory environment has paved the way into human-centric construction manufacturing which requires placing the well-being of workers at the center of production [4]. Well-being covers the physical, emotional, social, financial, vocational, and environmental aspects [5]. Human centric research in OSC has been evolving in several directions, such as ensuring safety [6], assessing and improving ergonomics[7], enhancing productivity [8], etc.

Achieving a safe and healthy indoor environmental quality (IEQ) is among the most critical human-centric factors. IEQ is impacted by several dimensions: air quality, thermal comfort, acoustic comfort, and lighting quality. Studies showed that there is a direct and strong correlation between IEQ and productivity. Individuals tend to perform their tasks more efficiently when they feel comfortable within the environment [9]. It is a core element in human-centric construction manufacturing as workers in these facilities can be prone to uncomfortable sound levels, indoor air pollution, and uncomfortable thermal state. These can be due to several factors such as the heavy use of machinery, construction material cutting and processing, the use of specific personal protective equipment (PPEs), the installed heating and ventilation systems, available natural lighting, visual comfort, etc. However, IEQ studies in OSC are limited to the acoustic comfort through identifying the noise sources in an OSC facility [10] as well as the workers' exposure level in an effort to identify safe and high-risk working areas [11].

Monitoring the IEQ is the first step into achieving a safe and comfortable indoor environment that meets the required IEQ standards and regulations set by different organizations such as ASHRAE [12], the World Health Organization (WHO) [13], the Environmental Protection Agency (EPA) [14], etc. However, monitoring on its own is not sufficient, and must be accompanied by an information interaction system that can be leveraged by management to make timely decisions. This can be achieved by developing a real-time monitoring and alert system to detect deviations from acceptable ranges. Once a deviation is detected, management is alerted to take the necessary measures. Digital twins (DT) can be used in this case as an integrated system to monitor and control the indoor environment [15]. DT systems are composed of three main components including physical space, virtual space and the information connection and processing between the two layers [16]. They have been successfully implemented for IEQ applications in the built environment. For example, Arsiwala et al. [17] created a DT-based solution to monitor the indoor air quality of existing building assets to enable data-driven decision making for retrofitting strategies. This ultimately helps achieve net zero carbon emission targets. In another study, Shahinmoghadam et al. [18] proposed a DT model in order to monitor thermal comfort in buildings in a virtual reality environment. The model also allows modifications to modify certain parameters that impact thermal comfort to study what-if scenarios based on the real-time thermal comfort data feed.

A review of the relevant literature showed that although the OSC industry is shifting more towards human-centric approaches, the IEQ applications remain limited. Moreover, IEQ directly correlates to the health, safety, and productivity of the individuals in the workplace, and therefore must be addressed in the context of OSC. Monitoring IEQ enables the detection of deviations from the established IEQ standards and regulations. In this regard, DTs have been successfully incorporated into IEQ real-time monitoring. However, IEQ applications in OSC have been relatively limited. As such, the objective of the present study is to develop a DT-based system for real-time monitoring of the IEQ of OSC facilities with continuous management and alerting features. The contribution of the system is bifold, 1) it provides a real-time assessment of the IEQ and 2) it can assist management to take the necessary corrective and preventive measures to ensure a safe, healthy, and productive environment.

2 Methodology

This study adopts the Design Science Research (DSR) method for developing the proposed IEQ monitoring system. DSR is a structured approach to developing artifacts that address an identified problem, and to demonstrate the artifact's effectiveness [19]. The method consists of the following four steps: (1) identifying the problem, (2) defining the objective, (3) developing the artifact, and (4) demonstrating the artifact [20].

In the present study, the first step is a thorough review of the literature to obtain a solid understanding of the current human-centric research in OSC, IEQ in OSC workplaces, and the potential of DTs in the built environment. The literature review reveals the importance of IEQ as a core element in the workplace. While the controlled indoor environment of OSC provides an opportunity to apply IEQ standards and improvements, applications of these standards in OSC are limited. The problem—in the present case, the gap in the literature regarding IEQ applications in OSC—having been identified, the system's objective is articulated in the second step of "Defining the objective." In the present case, the objective is defined as the development of a structured approach for real-time IEQ monitoring and alerting in OSC capitalizing on DT technology. With real-time monitoring of IEQ parameters, any deviations between the actual conditions at the factory and the identified thresholds can be detected. In the case of a deviation, management will be alerted, and the necessary measures can be taken. Ultimately, such a system helps to ensure the health, safety, and productivity of workers at the production facility. The third step is the development of the proposed system, and this comprises four phases: (1) IEQ metric identification, (2) data acquisition, (3) digital modelling, and (4) development of the IEQ alerting component of the system. The system employs DT-based tools that have been validated in the built environment, tailoring them to IEQ real-time monitoring in OSC. Several tools are leveraged, including sensor technologies, laser scanning, 3D modelling, Unity game engine, and web-based functions. Details on the development of the system are presented in section 3. The fourth step is to present an illustrative application, a case study of an OSC facility prefabricated components for residential projects—to demonstrate the system's applicability.

3 Proposed *O-IEQ Alert System*

The aim underlying the development of the *O-IEQ Alert System* (i.e., Step 3) is to provide a DT-based system for real-time IEQ monitoring and alerting in OSC facilities to promote the health, safety, and productivity of workers. As noted above, its development is composed of four phases: (1) IEQ metric identification, (2) data acquisition, (3) digital modelling, and (4) development of IEQ alerting component of the system. The following sub-sections outline the developed system (see Figure 1). It should be noted that the proposed system can be

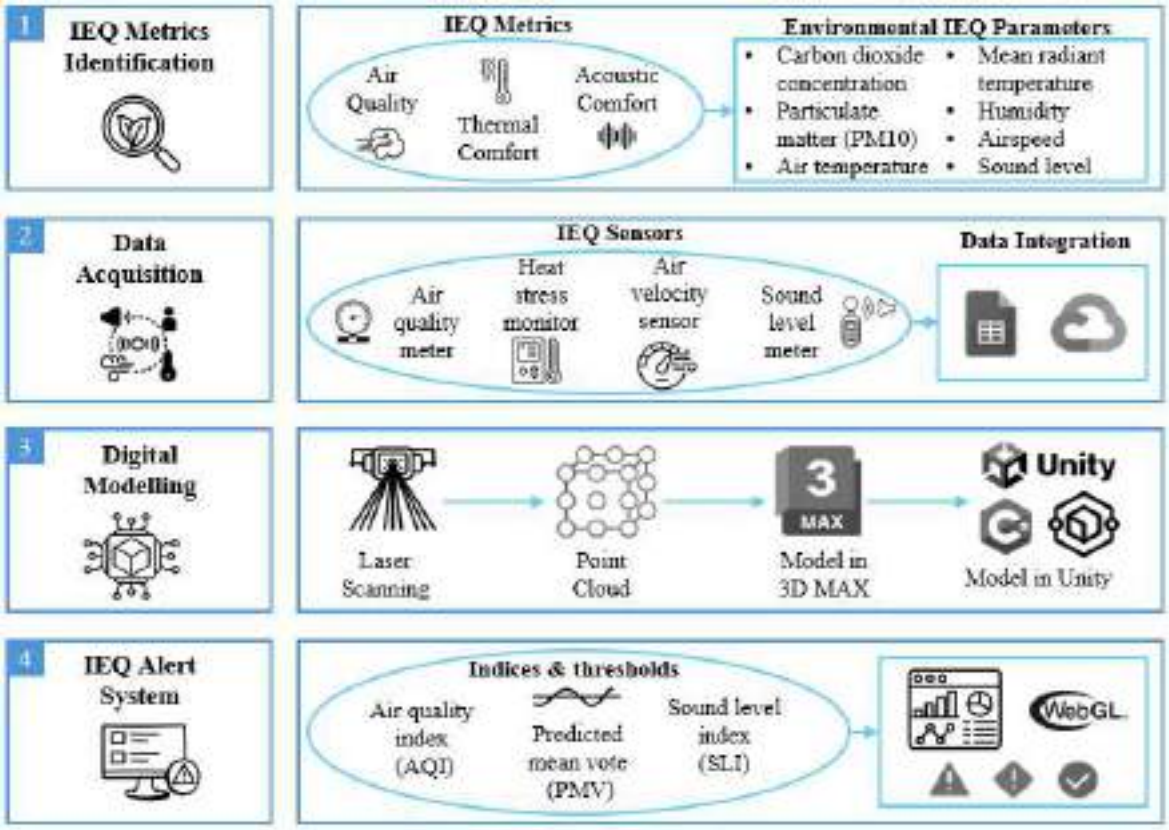


Figure 1. Proposed framework for the O-IEQ monitoring and alert system

customized as per the production operations as well as the IEQ needs and expectations of the production facility of interest.

3.1 IEQ Metric Identification

The proposed *O-IEQ ALERT System* encompasses three IEQ metrics: air quality, thermal comfort, and acoustic comfort. (Lighting quality may also be incorporated in future work.)

Indoor air quality is influenced by several parameters. In the present study, the system monitors levels of carbon dioxide (CO_2) and particulate matter (PM10). CO_2 is indicative of the freshness of indoor air, while PM10 measures the presence of coarse, harmful particles [21]

According to Fanger [22] a person's thermal comfort depends on six environmental and personal parameters: air temperature, mean radiant air temperature, relative humidity, air speed, metabolic rate, and clothing insulation. Air temperature refers to the temperature of the air surrounding the human body, while mean radiant air temperature refers to the heat radiated from a warm object. Air speed is the speed at which air circulates around humans. Relative humidity is the present state of humidity relative to the maximum humidity given at a certain temperature, and the metabolic rate refers to the physical effort required to perform tasks. Clothing insulation, finally, refers to the worker's ability to adapt their clothing to feel comfortable, while also considering the necessary personal protective equipment [22].

Acoustic comfort, meanwhile, has to do with the sound pressure levels perceived by the worker while performing tasks at the factory. Sound levels are influenced by the combined noise generated from different activities at the factory, whether manual or machine-based activities [11].

3.2 Data Acquisition

The parameters that affect the factory's IEQ having been identified, the next step is to develop a data acquisition system for collecting real-time data from each workstation. This is achieved by implementing various types of sensors at locations throughout the production facility. There are many types of sensors available for automated monitoring of IEQ parameters [23]. The selection of the sensors is typically case-specific and depends on sensor availability, range, accuracy, ease of use, ease of installation, cost, data transmission, and interoperability with other software.

Regarding the acquisition of IEQ data, an air quality meter can be used to measure the CO_2 and PM10 levels [24]. Personal parameters of thermal comfort are often assumed to be constant, whereas environmental factors—i.e., air temperature, mean radiant temperature, relative humidity, and air speed—are dynamic and must be monitored over time. Air temperature, mean radiant temperature, and relative humidity can be collected using a heat stress monitor, while air speed data can be collected using an air velocity meter [25]. To measure sound levels, a sound level meter can be employed [10]. Following the selection of commercially available sensors, sensing equipment is installed at the factory in accordance with the associated manufacturing guidelines.

The sensors, having been installed, data can be collected in real time at select intervals, such as one, five, or 10 minutes. The intervals are selected based on the criticality of the IEQ in the factory, as well as the variability of operations. Once the data on the different parameters is collected, it will undergo a process of data cleaning, processing, and structuring into the proper format so that it can be used as an input to the digital model. The specific steps depend on the selected sensors, the integration between the sensors, and the case study.

The cleaned and structured sensor data is then integrated with the digital model of the production facility (as discussed in the description of Phase 3 below) to ensure the link between the physical and virtual environments of the DT [26]. This integration is achieved by storing the sensor data in Google Sheets, with each row representing a single time stamp and each column representing a different IEQ parameter. These Google Sheets are hosted on a Google Cloud platform, which allows for separate sheets to be created for each workstation. An application programming interface (API) is then created in Google Cloud to integrate data from the sensors via the created Google Sheets. A URL link is then generated to present the data in JSON format. This allows for the data to be retrieved in the Unity modelling software. It should be noted that the APIs are restricted so that only employees of the case company and members of the research team can access the sensor data.

3.3 Digital Modelling

A 3D model that is reflective of actual production line operations in the facility and that represents the information flow of the IEQ indices and parameters in real time is critical, as it improves the user experience. In developing the 3D model, laser scanning is used to automatically measure and locate thousands of datapoints at the factory based on the selected positions of the laser scanners. This results in a spherical point cloud that maps to the factory. The point cloud, in turn, is imported to 3Ds Max for model development [26]. Finally, the model is imported to Unity software, which provides an interactive environment for the real-time monitoring and alert system.

3.4 IEQ Alert System

The purpose of the IEQ alert system is to provide a means of monitoring the quality of the indoor factory environment in real time. The system issues warnings if IEQ levels are not within the acceptable ranges set by the standards and regulations. Management can then take the necessary corrective or preventive measures to ensure a healthy, satisfactory, and productive environment for workers. For the alert system, three indices are selected corresponding with the three IEQ metrics of interest. The three indices are the Air Quality Index (AQI), Predicted Mean Vote index (PMV), and Sound Level Index (SLI). Each of these indices considers the influencing parameters identified in Phase 1 (see above). The indices having been identified, the next step is to determine the acceptable thresholds based on the relevant standards, guides, and regulations. A three-level warning system (good, fair, or poor) is introduced accordingly. Further details on the influencing parameters, index establishment, thresholds, and warning levels for each index are described as follows.

The AQI index is dependent on the CO_2 and PM10 concentration levels [21]. Since there is no universal index to measure the indoor air quality, the AQI was developed based on the thresholds of its dependent parameters. Table 1 shows the good, fair and poor ranges for the indoor CO_2 and PM10 concentration levels as per the ASHRAE 62.12 for CO2 [12]as well as the WHO [13] and the EPA [14] for PM10.

Table 1. CO_2 and PM10 levels

	Good	Fair	Poor
CO_2 concentration (ppm)	< 1,000	1,001–2,000	> 4,000
PM10 Level (ug/m³)	< 50	51–150	> 150

The next step identifies the 9 combinations of the CO_2 and PM10 levels to provide a qualitative indoor air quality index. If both CO_2 and PM10 are at acceptable levels, then the AQI is also acceptable with an index of 1. If either one is fair and the other is good, the AQI is moderate with an index of 2. If both are fair, the index will increase to 3. The index will be considered unhealthy at 4, if one is good and the other is poor. The index changes to 5 and becomes very unhealthy if either one becomes poor. Finally, the AQI index increases to 6 and becomes hazardous in case both parameters are in a poor state. Table 2 represents the developed AQI and its concern levels with a range from 1 to 6, with 1 being in good condition and 6 being in a hazardous state.

Table 2. AQI development

CO_2 Level	PM10 Level	AQI (index)	Level of Concern
good	good	1	Good
good	fair	2	Moderate
fair	good	2	Moderate
fair	fair	3	Moderate
good	poor	4	Unhealthy
poor	good	4	Unhealthy
fair	poor	5	Very unhealthy
poor	fair	5	Very unhealthy
poor	poor	6	Hazardous

Thermal comfort is dependent on 6 parameters, including personal and environmental parameters as per Fanger's model [22]. The metabolic rate and the clothing insulation are assumed to be constant values in the model. These can be selected based on the activity being performed by the worker and the clothing characterises respectively as per the ASHRAE Standard 55[27]. It should be noted that these values typically remain constant for a certain period and require an update upon major changes such as allocating a new worker to a certain task, entering a new season (winter and summer), or a change in the tasks of the production process, etc. As such, the manual selection of these values is feasible [28]. The environmental parameters include the air temperature, mean radiant air temperature, humidity, and air speed [22]. Based on these 6 parameters, thermal comfort can be quantified by calculating the Predicted Mean Vote (PMV) index as per the mathematical model developed by Fanger [22]. The PMV varies on a 7-class scale as follows: hot (−3), warm (−2), slightly warm (−1), neutral (0), slightly cool (−1), cool (−2), and cold (−4). Typically, a PMV between −0.5 and +0.5 reflects a comfortable indoor thermal environment.

The sound level index is developed based on the corresponding measured sound levels [11]. Table 3 presents the ranges of the noise levels, the level of concern, and the SLI. The SLI ranges from 1 to 6, with 1 being safe and 6 being extremely high risk.

Table 3. SLI development

Sound level dBA	Level of concern	Sound level index (SLI)
<66	Safe	1
66–71	Tolerable	2
71–76	Low risk	3
76–81	Moderate risk	4
81–85	High risk	5
85	Extremely high risk	6

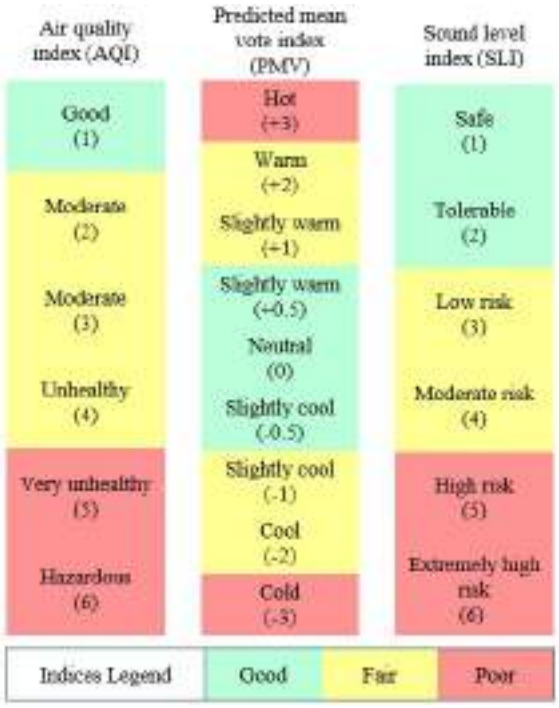


Figure 2. IEQ indices warning levels

Once the three IEQ indices are defined and established, the warning and alerting levels are introduced, as shown in Figure 2.

Once the indices, their thresholds, and warning levels are established, they are reflected in the Google Sheets on Google Cloud and thus are reflected in the APIs that feed data into the Unity model of the factory. A Canvas is created in Unity that shows two monitoring and alerting stages across two developed scenes. The first stage, Layout Scene, represents the layout of the production facility. This includes the sensors' location and status of the three IEQ indices for each workstation (good, fair, or poor). The second stage, Station Scene, shows the three indices' values, warning levels, and time series trends for each of the seven IEQ parameters for the station of interest. It should be noted that all indices, trends, and warning levels are updated in real time in Unity, given that the real time sensor data is fed via the API data integration.

The last step in the IEQ monitoring and alert system is to enable users (i.e., management at the production facility) to monitor the facility's IEQ remotely through a web-based decentralized network that can be accessed from any compatible browser [29]. Such solutions have been successfully applied in the planning of OSC projects [30] and are provided in Unity through the WebGL function as a JavaScript API [31].

4 Illustrative Application of the *O-IEQ Alert System*

The proposed *O-IEQ Alert System* is part of a larger study that aims to assess the quality of the indoor environment in OSC facilities to ensure a human-centric healthy, safe, and productive environment. A case study of a wall production line in an OSC facility is introduced to showcase an illustrative application of the proposed system. The application covers four stages of the process, including framing, sheathing, multi-function bridge, and the installation of windows and doors.

As discussed in the "Proposed O-IEQ Alert System" section, the three IEQ indices, including the AQI, PMV, and SLI, can be monitored by the *O-IEQ Alert System*, as well as the parameters of CO_2, PM10, air temperature, mean radiant air temperature, humidity, air speed, and sound level. Given that the data acquisition phase is under development, and the sensor installation and data collection steps are in progress, the authors used synthetic sample data to showcase the applicability of the system. This is a preliminary step that will help validate the system prior to real data acquisition and integration.

Although the authors relied on simulating data points to mimic the sensor data, the system is updating in real-time based on its acquired data. As such, the authors created a sample dataset with 200 data point stamps at 10-second intervals for each of the 7 parameters of interest. Then, the indices and the warning levels were accordingly calculated. The synthetic sensor data, along with the corresponding IEQ indices and warning levels, were uploaded to Google Sheets on Google Cloud, where the corresponding APIs were created. As such, the authors were able to replicate real-time data feed into the digital model for real time updating.

The digital model development of the production facility was achieved through laser scanning, whereby point cloud data was generated. The data was used to develop the model in 3Ds MAX, and the model was then transferred to Unity. The Unity model was uploaded to the web using the WebGL function to allow for online and remote access to the system. As such, the model can be accessed through a URL, as shown in Figure 3. Figure 3 also introduces the aim of the system and its main outcomes.



Figure 3. Web-based O-IEQ Alert System overview

Once the user runs the model, they will be taken to the Layout Scene, as shown in Figure 4. This scene presents the layout of the wall framing production line and the 4 stations understudy. Each of these stations is associated with a unique warning level, with coloured indices representing the IEQ index of interest. The user can select any of the 3 indices, and the warning levels will be updated accordingly.



Figure 4. Layout scene of O-IEQ Alert System

Once the user clicks on any of the stations, they will be taken to the Station Scene, as shown in Figure 5, providing more detailed insights regarding the IEQ. The user can view how the 3 indices are being updated and can select a certain index to view the time series of its corresponding parameters. Figure 5 shows the status of the framing station, where AQI is in a fair condition, and the PMV and SLI are in good condition. The user can select the other 2 indices to monitor the trends of the relevant parameters. It should be noted that the Layout Scene and the Station Scene are fed real time synthetic data using the aforementioned data integration workflow, and the indices, the warning levels, and the parameters trends are updated accordingly in real-time.

5 Conclusion and Future Work

In summary, this study proposed a digital twin-based system to monitor the IEQ of off-site construction facilities in real time with continuous management and alerting features. The developed *ALERT O-IEQ System* was developed in reference to the DSR method through identifying the need for the system, defining its objectives, developing and evaluating it. The system is composed of four phases including (1) IEQ metric identification, (2) data acquisition, (3) digital modelling, and (4) development of the IEQ alerting component of the system. An illustrative application was presented in a

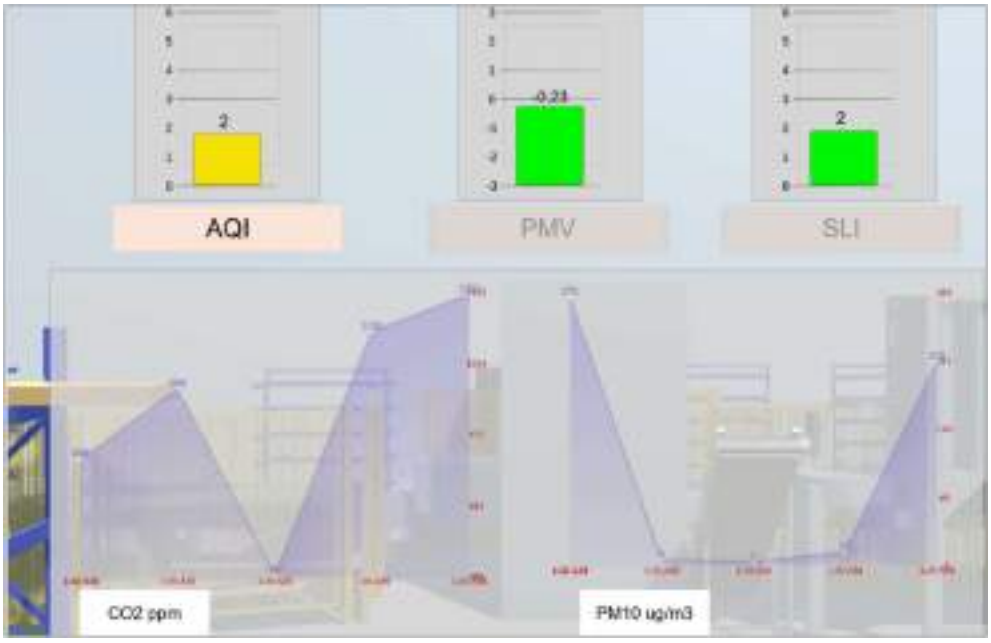


Figure 5. Framing station scene of the O-IEQ Alert System

wall-production manufacturing facility. A web-based Canvas in a 3D environment was developed to monitor the IEQ in real-time including the Air Quality Index, thermal comfort Predicted Mean Vote Index, and the Sound Level Index for each workstation. It also shows detailed insights into each workstation, including real-time time-series trends of the IEQ parameters (i.e. PM10 level, noise level, etc.). In case the IEQ is not within the acceptable thresholds, an alert is issued to the management in order to take the necessary measures. Certain limitations exist in the study that should be addressed. Firstly, the AQI index integrated in the developed system only considers two parameters, the CO2 and PM10 levels, while other parameters can impact the air quality. In the future, the AQI will be further developed to include the ground-level ozone, PM2.5, PM10, carbon monoxide, sulfur dioxide, and nitrogen dioxide as per EPA's U.S AQI to ensure more comprehensive and reliable air quality monitoring. Secondly, the system requires further validation to ensure its reliability in practice. As such, the system will be implemented on an actual case study of an OSC facility to test its effectiveness in the future.

References

[1] M. Kamali and K. Hewage, "Life cycle performance of modular buildings: A critical review," *Renewable and sustainable energy reviews*, vol. 62, pp. 1171–1183, 2016.

[2] P. Conrad, "Health and fitness at work: A participants' perspective," *Soc Sci Med*, vol. 26, no. 5, pp. 545–550, 1988.

[3] K. Danna and R. W. Griffin, "Health and well-being in the workplace: A review and synthesis of the literature," *J Manage*, vol. 25, no. 3, pp. 357–384, 1999.

[4] Y. Lu *et al.*, "Outlook on human-centric manufacturing towards Industry 5.0," *J Manuf Syst*, vol. 62, pp. 612–627, 2022.

[5] D. L. Stoewen, "Dimensions of wellness: Change your habits, change your life," *The Canadian veterinary journal*, vol. 58, no. 8, p. 861, 2017.

[6] M. Chatzimichailidou and Y. Ma, "Using BIM in the safety risk management of modular construction," *Saf Sci*, vol. 154, p., 2022, doi: 10.1016/j.ssci.2022.105852.

[7] C. Ritter, R. D. Barkokebas, X. Li, and M. Al-Hussein, "Integrated ergonomic and productivity analysis for process improvement of panelised floor manufacturing," in *18th International Conference on Modeling and Applied Simulation, MAS 2019*, 2019, pp. 174–180. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85074189873&partnerID=40&md5=c81226a63a5ddb4a9d128c00665e8288

[8] S. Assaf, F. Alsakka, M. Darwish, and M. Al-Hussein, "Simulation-Based Analysis of a Precast Factory Layout to Reduce Labor Travel

Time and Production Time," in *Construction Research Congress 2024, CRC 2024*, 2024, pp. 1087–1096. doi: 10.1061/9780784485262.111.

[9] N. S. Mahbob, S. N. Kamaruzzaman, N. Salleh, and R. Sulaiman, "A correlation studies of indoor environmental quality (IEQ) towards productive workplace," 2011.

[10] S. Dabirian, S. Han, and J. Lee, "Identification of modular construction activity noise levels by using k-means clustering," in *INTER-NOISE 2018 - 47th International Congress and Exposition on Noise Control Engineering: Impact of Noise Control Engineering*, 2018

[11] S. Dabirian, S. H. Han, and J. Lee, "Stochastic-based noise exposure assessment in modular and off-site construction," *J Clean Prod*, vol. 244, p., 2020, doi: 10.1016/j.jclepro.2019.118758.

[12] ASHRAE, "ASHRAE Position Document on Indoor Carbon Dioxide," 2022. [Online]. Available: www.ashrae.org

[13] W. H. Organization, *Air quality guidelines: global update 2005: particulate matter, ozone, nitrogen dioxide, and sulfur dioxide*. World Health Organization, 2006.

[14] R. Esworthy and J. E. McCarthy, "The National Ambient Air Quality Standards (NAAQS) for Particulate Matter (PM): EPA's 2006 Revisions and Associated Issues," Library of Congress, Congressional Research Service, 2013.

[15] Y. Zheng, S. Yang, and H. Cheng, "An application framework of digital twin and its case study," *J Ambient Intell Humaniz Comput*, vol. 10, pp. 1141–1153, 2019.

[16] E. Glaessgen and D. Stargel, "The digital twin paradigm for future NASA and US Air Force vehicles," in *53rd AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and materials conference 20th AIAA/ASME/AHS adaptive structures conference 14th AIAA*, 2012, p. 1818.

[17] A. Arsiwala, F. Elghaish, and M. Zoher, "Digital twin with Machine learning for predictive monitoring of CO2 equivalent from existing buildings," *Energy Build*, vol. 284, p. 112851, 2023.

[18] M. Shahinmoghadam, W. Natephra, and A. Motamedi, "BIM-and IoT-based virtual reality tool for real-time thermal comfort assessment in building enclosures," *Build Environ*, vol. 199, p. 107905, 2021.

[19] A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design science in information systems research," *MIS quarterly*, pp. 75–105, 2004.

[20] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A design science research methodology for information systems research," *Journal of management information systems*, vol. 24, no. 3, pp. 45–77, 2007.

[21] J. Saini, M. Dutta, G. Marques, J. Saini, M. Dutta, and G. Marques, "Indoor Air Quality and internet of things: the State of the Art," *Internet of Things for Indoor Air Quality Monitoring*, pp. 33–50, 2021.

[22] P. O. Fanger, "Assessment of man's thermal comfort in practice," *Occup Environ Med*, vol. 30, no. 4, pp. 313–324, 1973.

[23] J. Pantelic, N. Nazarian, C. Miller, F. Meggers, J. K. W. Lee, and D. Licina, "Transformational IoT sensing for air pollution and thermal exposures," *Front Built Environ*, vol. 8, p. 971523, 2022.

[24] S. Sleiman *et al.*, "Overview of occupant-centric KPIs for building performance and their value to various building stakeholders," *Energy Build*, p. 114704, 2024.

[25] C. Rînjea *et al.*, "Influence of the thermal environment on occupational health and safety in automotive industry: a case study," *Int J Environ Res Public Health*, vol. 19, no. 14, p. 8572, 2022.

[26] J. Shan and C. K. Toth, *Topographic laser ranging and scanning: principles and processing*. CRC press, 2018.

[27] A. Standard, "Thermal environmental conditions for human occupancy," *ANSI/ASHRAE, 55*, vol. 5, 1992.

[28] F. Lolli, A. M. Coruzzolo, and E. Balugani, "Indoor Environmental Quality (IEQ): A Comparison between TOPSIS-and PROMETHEE-Based Approaches for Indirect Eliciting of Category Weights," *Toxics*, vol. 11, no. 8, p. 701, 2023.

[29] M. Assaf, M. Al-Hussein, and X. Li, "Critical Review of Virtual Reality Applications in Offsite Construction Research," *J Constr Eng Manag*, vol. 150, no. 11, p. 03124007, 2024.

[30] M. Assaf, R. Lemouchi, M. Al-Hussein, and X. Li, "Virtual Reality-based Blockchain Application for Optimized Collaborative Decisions of Modular Construction," in *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, IAARC Publications, 2024, pp. 331–338.

[31] M. Chessa, C. Bassano, and F. Solari, "A WEBGL virtual reality exergame for assessing the cognitive capabilities of elderly people: a study about digital autonomy for web-based applications," in *International Conference on Pattern Recognition*, Springer, 2021, pp. 163–170.

Development of an assessment matrix for evaluating public participation procedures

Jonathan Matthei[1], Leon Streyl[1], Philip Balcar[1], Peter Bohnenkamp[1], Sven Mackenbach[1], Katharina Klemt-Albert[1]

[1] Institute for Construction Management, Digital Engineering and Robotics in Construction, RWTH Aachen University, Aachen, Germany
matthei@icom.rwth-aachen.de, leon.streyl@rwth-aachen.de, philip.balcar@rwth-aachen.de, p.bohnenkamp@icom.rwth-aachen.de, mackenbach@icom.rwth-aachen.de, klemt-albert@icom.rwth-aachen.de

Abstract

Public participation is crucial for designing livable cities and accelerating planning procedures. A key challenge is the complexity of planning documents, which are often difficult for non-experts to understand. Digital participation and techniques like 3D visualizations and Building Information Modeling (BIM) offer promising solutions to enhance the comprehensibility and acceptance of construction projects.

This study aims to develop an assessment matrix for participation procedures that enables the quality and effectiveness of such processes to be comprehensively evaluated. A systematic literature analysis is first used to identify success factors and challenges of public participation as well as specific success factors and challenges for digital forms of participation (e-participation). These success factors are divided into the categories of resources, communication and process-related aspects, while challenges are categorized into stakeholder-specific and overarching aspects.

Based on these findings, an assessment matrix is created to evaluate public participation in the context of planning and approval procedures. The assessment matrix includes parameters such as background information, the participation process, planning documents, and responses and their respective impacts. This matrix will be applied in the BIM4People research project, funded by the Ministry of Digital and Transport (BMDV), to identify improvements in current participation processes and explore how BIM models can optimize them.

Keywords

Building Information Modeling (BIM), e-participation, evaluation matrix for public participation, success factors and challenges of public and e-participation

1 Introduction

"*A good city is like a good party — people stay longer than really necessary because they are enjoying themselves*" - Jan Gehl [1]

Public participation plays a central role in the design of livable cities and infrastructures. Nevertheless, such participation processes regularly lead to conflicts, particularly in the case of complex infrastructure projects. Protests, resistance and legal disputes on the part of citizens are often the consequences, which can have a negative impact on the progress and success of such processes [2].

One of the major factors contributing to these conflicts are planning documents, which are often complex and difficult to understand, especially for people who are not familiar with technical drawings. The traditional presentation of planning documents is often in the form of textual elaborations and two-dimensional drawings that are difficult for the non-expert to access [3]. This barrier can make participation in planning processes considerably more difficult, as those affected often find it difficult to grasp and understand the full dimensions of the planned projects.

In this context, new possibilities for participation and visualization of construction projects, such as 3D visualizations and Building Information Modelling (BIM) models, offer promising approaches for improving public participation [4]. These technologies promise to provide more vivid and intuitive representations, making planning documents more comprehensible and potentially increasing transparency and public acceptance. To systematically evaluate the effectiveness of participation processes, various approaches and guidelines have been developed in the past [5–7]. However, these primarily focus on traditional, non-digital formats or specific types of participation. A comprehensive framework that incorporates both traditional and digital participation formats—while

enabling a comparative analysis of text-based, two-dimensional, and three-dimensional planning documents—has not yet been established.

With this background, an assessment matrix for planning and approval procedures is developed that makes it possible to systematically evaluate the quality of such public participation processes. It provides a structured method for evaluating various aspects and parameters that play an important role in such processes. The matrix is used to systematically identify weaknesses and potential for improvement in order to make the entire process both more efficient and more citizen friendly.

In order to develop this assessment matrix, a comprehensive overview of the success factors and challenges of public participation and digital participation opportunities, so-called e-participation measures [8], is first required. Analyzing these factors makes it possible to determine the relevant parameters that are important for the evaluation of the planning documents. Once the success factors and challenges have been analyzed, the specific assessment parameters can be derived and integrated into the assessment matrix.

The detailed methodological procedure for developing the assessment matrix is described in the following chapter. The identified success factors and challenges are then highlighted, and the developed assessment matrix is presented. Finally, the key results are discussed and summarized. This section also contains an outlook that shows the context in which the assessment matrix is to be used as an example.

2 Methodology

To develop the assessment matrix, a systematic literature analysis was carried out based on the approach of Webster and Watson (2002) [9] and Vom Brocke et al. (2009) [10].

According to Webster and Watson, a literature search essentially involves searching scientific databases using keywords. The authors recommend using a backward and forward search to find relevant literature. A forward search looks for literature that cites the literature already found. In a backward search, on the other hand, the literature found is searched for cited literature in order to include it after checking its suitability [9].

Vom Brocke et al. (2009) take up these approaches and at the same time note that only a few detailed methods and standardized procedures exist for literature research. They emphasize the difficulty of searching for literature at a time when more and more sources and literature are being published every year. In their explanations, the authors emphasize the importance of a careful literature review. Strict documentation and presentation of the research processes is required in order to make the procedure comprehensible and reproducible [10]. Therefore, the systematic literature analysis of this study is based on three phases.

In the **first phase** of this study, search criteria and search parameters are defined. The databases SpringerLink, ScienceDirect, and Scopus were selected as the foundation for the literature search. Scopus and ScienceDirect, as predominantly English-language databases, were used exclusively for English-language literature. In contrast, SpringerLink was specifically chosen to cover German-language sources. Google Scholar was included as it is considered to cover German-language literature more efficiently than comparable databases [11]. Moreover, its inclusion ensured that relevant insights from non-academic sources, such as municipal master plans and governmental reports, were also captured, as these can provide valuable findings for the research context. However, Google Scholar does not provide literature directly but refers to other databases [12].

The literature search is based on predefined keywords to systematically identify relevant sources. Keywords related to public participation, including ‘public participation’ and ‘citizen participation,’ as well as their German equivalents, were selected. Additionally, terms associated with e-participation, such as ‘e-participation,’ its German translation, and ‘digital participation,’ were included. This targeted selection ensures a comprehensive review of relevant literature in both English and German.

To refine the search scope, only open-access sources classified under the subject area *Engineering* were considered. In Scopus, the presence of the respective keyword was required, while in ScienceDirect, the term *participation* had to appear in the title. Additionally, the search was limited to the last 15 years (2008–2023) to ensure relevance, as scientific findings are continuously updated. Given the rapid innovation cycles in e-participation technologies, a more recent timeframe is particularly relevant. [13]. Furthermore, by 2008, over three-quarters of German households had broadband access, establishing the foundation for digital participation.[14]. This approach resulted in 979 identified sources.

In the **second phase**, duplicates were first removed. The remaining sources then underwent an abstract screening to assess their relevance. After this process, 169 sources met the criteria and were selected for detailed analysis.

In the **third phase**, the success factors and challenges for both public participation and e-participation were analyzed. Based on the respective success factors and challenges, 23 evaluation parameters were then derived.

The entire procedure, including each phase from initial literature search to the development of the assessment matrix, is summarized in the following figure.

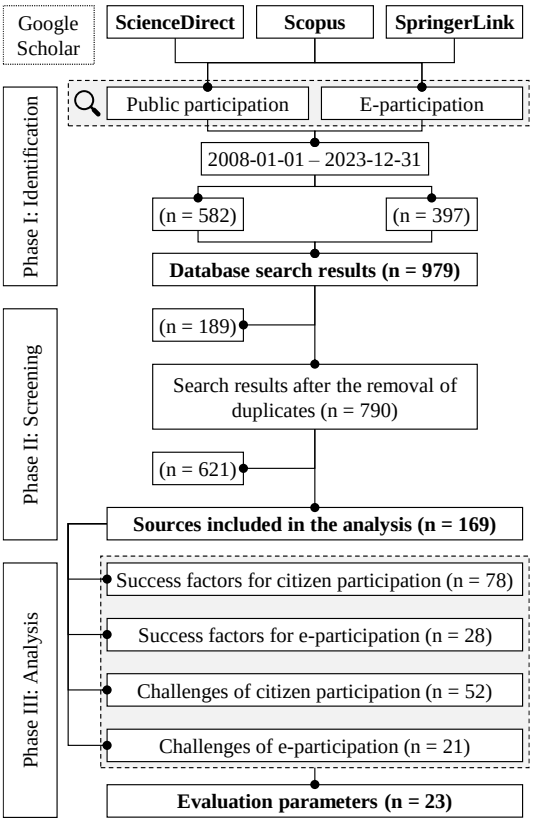


Figure 1. Methodology Findings

The first step is to define the terms 'success factor' and 'challenge' for a shared understanding in evaluating public participation processes. Success can be perceived subjectively by different actors, with citizens viewing success as the assertion of their interests, while organizers may see functional participation as successful [15]. In 1988, Lutz Hildebrandt defined the 'success factor' as a key variable linking performance indicators and planning units [16].

The following definition can be derived from this regarding the success of public and e-participation:

Success factors in public participation and e-participation measures are those variables that help to make the participation process as pleasant and simple as possible for all those involved, as well as creating a feeling of satisfaction and promoting and strengthening user behavior.

According to the Oxford English Dictionary, challenges are defined as demanding situations or tasks designed to test one's abilities or character. Lazarus and Folkmann, on the other hand, write that the evaluation of challenges focuses on the winning or growth potential of an encounter (and is characterized by pleasurable emotions such as eagerness, excitement and enthusiasm) [17]. The focus/stimulus of challenges is therefore to achieve the desired goal by completing the demanding task. However, there is always the risk of failing at the task or that the desired goal is not achieved. Challenges in the context of public and e-participation measures can therefore be defined as follows:

Challenges in public participation and e-participation measures are all matters that could lead to more successful participation but involve a lot of effort and are not guaranteed to have a positive impact.

2.1 Success factors for public participation and e-participation

The study identified 78 success factors for public participation and 28 specifically for e-participation measures. Following an inductive categorization method as outlined by Mayring [18], these success factors can be grouped into three main categories: "Resources," "Communication," and "Process-related Aspects." Each main category further divides into several subcategories (see Figure 2 and 3).

	Success factors for public participation		
	Resources	Communication	Process-related aspects
	1	2	3
a	Personnel n = 6	Real participation n = 8	Traceability in the process n = 4
b	Organisational n = 17	At eye level n = 10	Preparation and follow-up n = 10
c	Range n = 7	Create clarity n = 9	Quality and continuity in the process n = 7

Figure 2. Success factors for public participation

	Success factors for e-participation		
	Resources	Communication	Process-related aspects
	1	2	3
a	Personnel n = 3	Avoiding frustration n = 3	Preparation and follow-up n = 2
b	Organisational n = 2	Exchange medium n = 12	Quality and continuity in the process n = 5
c	Range n = 1		

Figure 3. Success factors for e-participation

In the area of public participation, **resources** are crucial to the success of participation processes. Resources are understood as 'certain means that are necessary to achieve a goal' [19]. Human resources, such as the stakeholders involved, play a key role. Social competence from authorities is a crucial success factor, as it facilitates communication and interaction with citizens, improving the process's effectiveness [20]. Another example is the involvement of external, neutral moderators. These help to promote the acceptance of procedures and minimize barriers to participation, as they act as impartial mediators [21]. The availability of technical support plays a decisive role in the area of e-participation measures. Technical support ensures that all participants receive the necessary assistance with digital matters [22]. This makes a significant contribution to ensuring that sufficient support is provided when giving feedback, which is an important assessment parameter in the evaluation of participation processes.

Organizational success factors include the general willingness of the authority to carry out participation [23]. This willingness is reflected in the openness of the authorities to actively involve citizens in decision-making processes. Other factors are the existence of a harmonious relationship between the costs and benefits of participation [20] and the general accessibility of information [24]. In the context of e-participation measures, it is also necessary to ensure barrier-free online access [23].

Outreach is another aspect of successful participation processes. The success factors in this sub-category include the use of various media to disseminate information [20] and widespread publicity for the implementation of a participation process [25]. Another success factor for e-participation measures is the awareness of and attention generated by digital participation opportunities [26].

The **communication** category refers to the direct exchange between the participants and the participating institutions. Both verbal and non-verbal exchange is crucial here.

In this context, various success factors can be identified that ensure genuine participation. Key success factors include avoiding the impression of a fictitious participation [27] and a recognizable prospect of success through the investment [24]. This knowledge motivates active participation and increases the willingness to get involved in the process. A concrete evaluation parameter that captures this aspect of communication is the consideration of feedback in further planning. This shows the participants that their contributions are taken seriously and are actually incorporated into the decision-making processes.

Furthermore, success factors that build trust are crucial, including openness to communication, willingness to discuss, and avoiding specialized terminology [20].

Clear communication is also conducive to extensive participation and a process that is perceived as fair. Unclear statements and feedback can lead to false expectations, which should be avoided [28]. A key evaluation parameter that ensures this clarity is the thematic response to feedback from the public. If the authority's feedback specifically addresses the points raised, misunderstandings can be avoided and the satisfaction of those involved can be increased.

In the area of e-participation measures, there are additional factors to consider that help to avoid frustration when using digital technologies. One decisive factor is a user-friendly interface that makes it easier for users to find their way around digital platforms [29]. In addition, it is important to provide visual and auditory content to cater for different learning and communication styles and to facilitate access for all user groups [26].

The category of **process-related aspects** includes factors that influence 'traceability in the process', 'preparation and follow-up' and 'quality and continuity in the process'.

Transparent procedures and the provision of background information are necessary for traceability in the process [21]. Transparent procedures enable those involved to clearly understand the progress and decisions within the process. This evaluation parameter increases the confidence of those involved in the process and ensures that they feel they are taken seriously. To achieve this goal, the planning documents themselves must also be relevant, comprehensible and transparent.

Preparation and follow-up include early activation and mobilization of the public [15] and the start of participation before irreversible decisions are made [24]. This enables a well-founded opinion to be formed and promotes the quality of the contributions. In this context, the procedure and the key points of participation must be known [30].

In order to ensure high quality and continuity in the process, it is of central importance that comprehensive information on participation is provided [21]. In addition, proactively taking up feedback during the planning phase is a key success factor. This ensures that the feedback from those involved is taken into account and integrated into further planning, which increases the effectiveness and acceptance of the entire participation process.

E-participation also benefits from limiting anonymity for accountability and conducting objective fact-checks to maintain accurate and current information [31].

2.2 Challenges of public participation and e-participation

Using Mayring's inductive categorization method [18], the study identified 52 challenges in public

participation and 21 specific to e-participation. These are categorized as "Stakeholder-specific" and "Overarching challenges" (see Figure 4).

	Challenges of public participation and *e-participation	
	Stakeholder-specific challenges	Overarching challenges
	1	2
a	On the authority side n = 12 *n = 7	Representation n = 7 *n = 1
b	Interaction with the citizen side n = 13 * n = 7	Framework conditions n = 12 *n = 4
c	Citizen side n = 8 *n = 2	

Figure 4. Challenges of public participation and e-participation

Stakeholder-specific challenges in public participation are categorized into three sub-categories: public authorities, citizens, and their interactions.

Public authority challenges mainly involve internal structures and processes, including resource shortages that hinder effective participation implementation in terms of staff and finances [32]. Additional challenges arise in the area of e-participation measures. E-participation introduces technical and organizational challenges, requiring authorities to develop and maintain digital platforms. [33]. Managing large volumes of digital data also demands effective systems for collection, analysis, and interpretation. [26].

There are numerous challenges on the citizens' side, with a central problem being insufficient information about participation opportunities and the underlying procedures [20]. In addition, there is often a lack of trust in the authorities based on previous experience of insufficient consideration of citizens' opinions [29]. In this context, it is essential to ensure the credibility of the planning documents. Consistent and complete planning documents are crucial for this. In addition, the fulfilment of technical requirements for e-participation measures poses a challenge. Many citizens do not have the necessary digital skills or access to suitable technology to participate in online participation procedures [26].

The interaction between authorities and citizens also harbors specific challenges. A frequent problem is dealing with different interest groups [34] and the scope of information provision [29]. These difficulties can lead to time and financial burdens on the one hand and make it difficult to satisfy individual concerns on the other. With regard to e-participation measures, there are additional challenges, particularly in terms of ensuring the credibility of these measures. The credibility of e-participation processes is often questioned as there are concerns regarding the manipulability of online surveys and polls [35].

Overarching challenges include representation and framework conditions, which are crucial for the effectiveness and legitimacy of participation processes.

A common issue is the dominance of well-organized, vocal groups, which can overshadow diverse perspectives [21]. Additionally, proponents of a project are often underrepresented, with participation processes frequently attracting more opponents, leading to an imbalanced discourse [21].

The framework conditions of public participation processes are also of crucial importance. A frequent challenge is the lack of clarity about the influence of citizens' opinions on the actual decisions [28]. In addition, the high complexity of construction projects poses a further challenge [29]. Construction projects often involve complex technical, legal and economic aspects that can be difficult for laypeople to understand. Consequently, more vibrant, clear and understandable documentation and feedback are crucial and consequently represent concrete evaluation parameters.

E-participation faces challenges like software dependency [26], where functionality and usability are crucial, and resistance to change from both citizens and authorities unfamiliar with digital tools [36].

3 Assessment matrix

The assessment matrix includes aspects, parameters, descriptions, and a scale from 1 to 6. Aspects are broad categories, with parameters serving as specific criteria. The scale evaluates the degree of fulfillment, from 1 (not fulfilled) to 6 (fully fulfilled), allowing for a detailed assessment of strengths and weaknesses.

The assessment matrix includes four aspects: (i) background information, (ii) participation process, (iii) planning documents, and (iv) response and impact. These were derived from the identified success factors and challenges. Since evaluation parameters often influence multiple factors, a direct one-to-one mapping was not possible. The matrix focuses on evaluating public participation and e-participation from the citizens' perspective to help identify strengths and weaknesses for the initiators.

The 'background information' aspect ensures citizens understand the process, deadlines, and project details. The 'participation process' evaluates citizen involvement and support. The 'planning documents' aspect assesses the quality and review challenges of the provided information. The 'response and impact' aspect examines how authorities respond to citizen input and its impact.

Aspect	Assessment parameter	Description	Ranking					
			1	2	3	4	5	6
Background-information	Familiarity with the procedure	Assesses whether the course of the procedure is known.						
	Awareness of dates and deadlines for participation	Evaluates whether the relevant deadlines and dates for participation have been clearly communicated.						
	Sufficient information on the project	Measures whether sufficient information has been provided about the project.						
	Feeling of sufficient participation present	Assesses whether the feeling of having been given sufficient opportunities to participate.						
Participation process	Sufficient information for providing feedback	Checks whether the provision of information on participation was adequate.						
	Sufficient support for providing feedback	Measures the extent to which support measures were available for the submission of feedback.						
Planning documents	Accessibility of the planning documents	Evaluates the accessibility of the planning documents.						
	Completeness of the planning documents	Checks whether the planning documents contain all the necessary information.						
	Comprehensibility of the planning documents	Checks the comprehensibility of the planning documents.						
	Attractive visual presentation of the planning documents	Evaluates the clear presentation of the planned impacts in the planning documents.						
	Clarity of the planning documents	Evaluates the clarity of the planning documents.						
	Relevance of the planning documents	Evaluated whether all relevant information was easy to find in the planning documents						
	Consistency of the planning documents	Measures the consistency within the planning documents provided.						
	Credibility of the planning documents	Assesses the credibility of the planning documents.						
	Comprehensibility of the planning documents	Checks whether the presentation of the impacts in the planning documents is comprehensible.						
	Transparency of the planning documents	Evaluates the transparent presentation of decisions in the planning documents.						
	Appropriate timeframe for reviewing the planning documents	Measures whether the time available for reviewing the planning documents was sufficient.						
	Financially unproblematic inspection of the planning documents	Evaluates whether access to the planning documents was associated with financial barriers.						
Response and impact	Comprehensibility of the feedback	Measures the comprehensibility of the feedback received.						
	Thematic response to the feedback	Evaluates whether the feedback has been answered thematically.						
	Satisfaction with the feedback received	Checks whether the feedback could satisfy the individual concerns.						
	Discussion of the feedback	Checks to what extent the submitted feedback has been discussed.						
	Feeling of integration of feedback into the planning process	Evaluates whether there is a feeling that the feedback has been taken into account in further planning.						

1 = Not achieved 2 = Barely achieved 3 = Partially achieved 4 = Well achieved 5 = Almost fully achieved 6 = Fully achieved

Table 1. Assessment matrix for evaluating public participation procedures in planning and approval procedures

4 Discussion and Conclusion

Public participation is a crucial factor in accelerating planning and approval processes, significantly contributing to the success of complex infrastructure projects. Despite its importance, dissatisfaction and complaints remain common. Digital participation formats offer new opportunities to improve the comprehensibility and acceptance of these processes. However, existing evaluation approaches often focus on traditional, non-digital participation methods or specific participation types, lacking a comprehensive framework that equally considers both conventional and digital formats. A holistic evaluation approach is essential to systematically assess the transition to digital participation and analyze its potential and challenges in a structured manner.

This study aimed to develop an assessment matrix that systematically evaluates the quality and effectiveness of public participation processes in the context of planning and approval procedure, identifying weaknesses and potential areas for improvement, particularly regarding digital visualization techniques. A systematic literature review was conducted to identify relevant success factors and challenges, which were categorized into four dimensions: 'resources,' 'communication,' 'process-related aspects,' and 'stakeholder-specific and overarching challenges.' The resulting assessment parameters were incorporated into the matrix, which consists of four core aspects: background information, participation process, planning documents, and response and impact.

The developed assessment matrix serves as a structured tool for objectively evaluating public participation processes. It enables a systematic comparison of different planning document formats—whether text-based, 2D, or 3D visualizations—contributing to a deeper understanding of their effectiveness in participation processes. Since the matrix is designed for use in both analog and digital participation formats, it offers a flexible means of identifying and addressing shortcomings. If certain aspects—such as comprehensibility or accessibility—receive low scores, particularly in relation to the format of planning documents, alternative representations (e.g., replacing text-based or 2D plans with 3D visualizations) can be tested. The matrix then facilitates a systematic evaluation of whether these adjustments lead to improvements in specific assessment parameters, ensuring that participation strategies can be iteratively refined based on empirical findings.

This structured evaluation fosters dialogue between citizens and decision-makers, ensuring that participation processes not only meet formal requirements but also align with the needs and expectations of stakeholders. By applying the matrix, targeted recommendations for improvement can be developed to enhance the quality and acceptance of public participation processes. Additionally, incorporating free-text input allows for capturing qualitative feedback that extends beyond predefined assessment parameters, further refining the evaluation and optimization of participation strategies.

Acknowledgement

This research paper was written as part of the project BIM4People [19FS2057A], which is funded by the Federal Ministry for Digital and Transport (BMDV) in Germany.

References

[1] Gehl, J. (2010) *Cities for People*. Washington DC: Island Press.

[2] Albrecht, R. et al. (2013) *Optionen moderner Bürgerbeteiligung bei Infrastrukturprojekten – Ableitungen für eine verbesserte Beteiligung auf Basis von Erfahrungen und Einstellungen von Bürgern, Kommunen und Unternehmen*.

[3] Brettschneider, F.; Spieker; Arne (2017) *Bauprojekte Visulaisieren – Leitfaden für die Bürgerbeteiligung* in: Schriftenreihe der Baden-Württemberg, H. 86.

[4] Matthei, J.; Rausch, S.; Klemt-Albert, K. *BIM basierte Planfeststellung für Infrastruktur-Großprojekte*. https://doi.org/10.1002/bate.202400016

[5] Arbter, K. (2011) *Standards der Öffentlichkeitsbeteiligung – Praxisleitfaden*.

[6] Ritzi, C.; Kaßner, J. (2019) *Evaluationsleitfaden für Beteiligungsverfahren. vhw-Schriftenreihe 11. vhw*.

[7] Nadeem, O.; Fischer, T. B. (2011) *An evaluation framework for effective public participation in EIA in Pakistan* in: Environmental Impact Assessment Review 31, Nr. 1, pages 36–47. https://doi.org/10.1016/j.eiar.2010.01.003

[8] Le Blanc, D. (2020) *UN Department of Economic and Social Affairs (DESA) Working Papers*. United Nations Publications.

[9] Webster, J.; Watson, R. T. (2002) *Analyzing the past to prepare the future: Writing a literature review*.

[10] vom Brocke, J. et al. (2009) *Reconstructing the giant: On the importance of rigour in documenting the literature search process*.

[11] Breuer, W. (2009) *Google Scholar as a Means for Quantitative Evaluation of German Research Output in Business Administration – Some Preliminary Results*.

[12] Delgado Lopez-Cozar et al. (2019) *Google Scholar as a Data Source for Research Assessment* in: Glänzel, W. et al. [Hrsg.] *Springer Handbook of Science and Technology Indicators*. Cham: Springer International Publishing, pages 95–128.

[13] Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag (2017) *Chancen und Risiken mobiler und digitaler Kommunikation in der Arbeitswelt*.

[14] Statistisches Bundesamt, Wirtschaft und Statistik (2011) *Internetnutzung in privaten Haushalten in Deutschland – Ergebnisse der Erhebung 2010*, pages 709–718.

[15] Uhlendahl, T. C. (2011) *Öffentlichkeitsbeteiligung in der Umweltplanung?* in: Standort 35, H. 1, pages 22–28. https://doi.org/10.1007/s00548-011-0152-6

[16] Hildebrandt, l. (1988) *Store Image and the Prediction of Performance in Retailing*, pages 91–100.

[17] Lazarus, R. S.; Folkman, S. (1984) *Stress, Appraisal, and Coping*. New York: Springer Publishing Company.

[18] Mayring, P. (2022) *Qualitative Inhaltsanalyse: Grundlagen und Techniken* 13. überarbeitete Auflage.

[19] Bundeszentrale für politische Bildung (2024) *Ressouren* [online]. https://www.hanisauland.de/wissen/lexikon/grosses-lexikon/r/ressourcen.htmlm Accessed: 5/10/2024.

[20] Bundesministerium für Verkehr und digitale Infrastruktur (2014) *Handbuch für eine gute Bürgerbeteiligung*.

[21] Bertelsmann Stiftung (2013) *Mehr Transparenz und Bürgerbeteiligung – Prozessschritte und Empfehlungen am Beispiel von Fernstraßen, Industrieanlagen und Kraftwerken*.

[22] Schoßböck, J.; Rinnerbauer, B.; Parycek, P. (2018) *Digitale Bürgerbeteiligung und Elektronische Demokratie* in: Leitner, M. [Hrsg.] *Digitale Bürgerbeteiligung*. Wiesbaden: Springer Fachmedien Wiesbaden, pages 11–40.

[23] Märker, O.; Wehner, J. (2008) *E-Partizipation* in: Standort 32, H. 3, pages 84–89. https://doi.org/10.1007/s00548-008-0088-7

[24] Klages, H.; Vetter, A. (2013) *Bürgerbeteiligung auf kommunaler Ebene*. Nomos.

[25] Regener, M. (2010) *Umweltbelange in der Öffentlichkeitsbeteiligung – eine Bestandsaufnahme* https://doi.org/10.1007/s00548-010-0143-z

[26] Radtke, J.; Saßmannshausen, S. M. (2020) *Auf dem Weg zur responsiven Demokratie? Online-Öffentlichkeitsbeteiligung in der Stadtentwicklung als aktiver Link zwischen Kommunalpolitik und Bevölkerung* in: Zeitschrift für Politikwissenschaft 30, H. 2, pages 329–358. https://doi.org/10.1007/s41358-020-00233-4

[27] Hofmann, E. (2012) *Die Modernisierung des Planungsrechts: das Energierecht als neues Paradigma der Öffentlichkeitsbeteiligung in einer Planungskaskade?* in: JuristenZeitung, pages 701–711.

[28] Bauer, C. (2015) *Stiftung von Legitimation oder Partizipationsverflechtungsfalle* in: Zeitschrift für Public Policy, Recht und Management, pages 273–293.

[29] Rademacher, L.; Lintemeier, K.; Kretschmer, H. (2020) *Öffentlichkeitsbeteiligung bei Infrastrukturprojekten als Herausforderung für Politik und Verwaltung* in: Kocks, K.; Knorre, S.; Kocks, J. N. [Hrsg.] *Öffentliche Verwaltung – Verwaltung in der Öffentlichkeit*. Wiesbaden: Springer Fachmedien Wiesbaden, pages 165–184.

[30] Kallenbach, B.; Barth, R.; Brohmann, B. (2008) *Anforderungen an die Gestaltung von Öffentlichkeitsbeteiligung – Zur Planung von Endlagerstandorten für hochradioaktive Abfälle* in: Technikfolgenabschätzung – Theorie und Praxis, pages 72–78.

[31] Himmel, W. *Bürgerbeteiligung gelingt auch digital* in: *Luppold, Himmel et al. (Hg.) 2021 – Berührende Online-Veranstaltungen. https://doi.org/10.1007/978-3-658-33918-0_3*, pages 27–45.

[32] Landua, D. et al. (2013) *Auf dem Weg, nicht am Ziel – Aktuelle Formen der Bürgerbeteiliung - Ergebnisse einer Kommunalbefragung*.

[33] Krämer, T. (2015) *Online-Bürgerbeteiligung bei der Entwicklung der Aplpenregion* in: Doleski, O. D.; Lorenz, K. [Hrsg.] *Energie der Alpen*. Wiesbaden: Springer Fachmedien Wiesbaden, pages 35–44.

[34] Senatsverwaltung für Stadtentwicklung und Umwelt Berlin (2012) *Handbuch zur Partizipation*.

[35] Leitner, M. [Hrsg.] (2018) *Digitale Bürgerbeteiligung*. Wiesbaden: Springer Fachmedien Wiesbaden.

[36] Doppler, S.; Kraut, M.; Steffen, A. (2021) *Herausforderung Customer Experience bei digitalen Veranstaltungen – Erkenntnisse aus der Corona-Krise* in: Luppold, S.; Himmel, W.; Frank, H.-J. [Hrsg.] *Berührende Online-Veranstaltungen: So gelingen digitale Events mit emotionaler Wirkung*. Wiesbaden, Heidelberg: Springer Gabler, pages 277–305.

Enabling human-robot collaboration in unstructured construction environment through sharing control: Classification, Challenges, and Future Trends

Guohao Wang[1,*], Changrui Zhu[2], Honghu Chu[1,3], Abdul-Majeed Mahamadu[1] , Vijay M. Pawar[1]

[1]Bartlett School of Sustainable Construction, University College London, United Kingdom
[2]Computer Science, University College London, United Kingdom
[3]College of Civil Engineering, University of Hunan, China

*ucbvanb@ucl.ac.uk

Abstract -

Robots hold great potential for enhancing efficiency, quality, and safety in construction, yet the unstructured and dynamic nature of construction sites limits the feasibility of full automation. Human-robot collaboration (HRC) offers a solution by integrating human expertise with robotic precision to improve adaptability. However, there remains a significant gap in understanding the systematic integration of human and robotic intelligence, particularly in dynamic control sharing. This study addresses this gap through a systematic review of shared control in HRC within unstructured environments, reflecting on its application to construction. A two-layer classification framework is proposed, encompassing "human-in-the-loop" (HITL) and "human-on-the-loop" (HOTL) systems, further divided into five subcategories. Key challenges and future research directions are also identified, providing actionable insights for designing and optimizing HRC systems in complex construction scenarios.

Keywords -

Robotic construction; Human-robot collaboration; Control sharing; Unstructured construction environment

1 Introduction

The construction industry, which contributes approximately 13% of global GDP, continues to face persistent challenges related to safety and labor shortages. In the UK, construction sector accounted for 51 worker fatalities in 2023, representing 37% of all workplace deaths [1], alongside a shortfall of over 150,000 skilled workers jeopardizing housing plans [2]. Robotics presents a promising avenue for mitigating workforce deficits and undertaking hazardous tasks. However, the inherently unstructured and dynamic nature of construction sites—characterized by irregular components and unpredictable environmental conditions—demands a high degree of adaptability. This complexity underscores the difficulty of achieving fully autonomous robotic operations in construction, highlighting the need for intelligent systems capable of navigating uncertainties and adjusting dynamically [3].

Humans' agility and expertise are crucial for managing uncertainties in unstructured environments, making human-robot collaboration (HRC) a viable approach for deploying construction robots. Although originally developed for industrial settings, HRC has expanded into fields such as surgery and military operations, while its application in construction remains in an early stage. Recent reviews have examined HRC in construction, focusing on robot autonomy [3], worker skill transfer methods [4], multidimensional taxonomy frameworks [5], and applications scenarios [6], laying the foundation for advancements in this field. Nevertheless, it remains unclear how the integration of human operators' intelligence and robotic autonomy is achieved in unstructured construction environments. In this study, we focus primarily on human-robot collaboration systems under teleoperation scenarios. Within this context, as the robot is the primary entity executing tasks, human-robot collaboration can also be referred to as shared control of the robot. A limited understanding of shared control constrains HRC system design, restricts autonomy, heightens risks, reduces efficiency, and undermines user trust—ultimately impeding adoption in complex construction settings.

To fill the aforementioned knowledge gaps, this study investigates shared control mechanisms of human-robot collaboration systems in an unstructured construction environment. A systematic literature review methodology is employed, involving comprehensive retrieval and rigorous screening to ensure the scientific rigor and reliability. The findings aim to provide theoretical support for the design and optimization of HRC systems in such complex settings. The research questions are formulated as following: (1) What shared control methods have been developed in existing human-robot collaboration systems in unstructured environment? And how can they be systematically classified? (2) What are the key challenges in implementing shared control systems for construction

robots operating in unstructured environments? (3) What are the emerging research trends for the advancement of shared control systems in construction robotics?

2 Methodology

This study conducted a systematic literature review following the PRISMA guideline [7]. As instructed in PRISMA flow diagram, there are two steps in PRISMA i.e. studies identification and screening each of which will be introduced in details in this section.

2.1 Identification

In PRISMA, relevant records need to be systematically identified first and then initially filtered out of duplicates and ineligible records using automated tools. To ensure the comprehensiveness of the literature review, we carefully selected search terms as shown in the Figure 1. Specifically, we incorporated a broad range of terms capable of reflecting various human-robot collaboration paradigms that combine human expertise with robot autonomy. It is important to note that these terms do not have standardized definitions, and in some cases they are cross-used. In addition, we selected "robot" and "uav" to cover common types of robots. For built environments, the search strategy included terms describing unstructured environments similar to buildings, ensuring coverage of a wide range of dynamic and changing built environments, including "building", "architecture", "infrastructure ", "unstructured environments", "changing environments", "dynamic environments", and "clutter ".

Through this systematic search, 656 records were initially retrieved. Automated filters were then applied to refine the dataset, restricting it to English publications from 2010 to September 2023 within the engineering and computer science fields to ensure a comprehensive and up-to-date literature review. Finally, 521 records remained after removing duplicates.

2.2 Screening

In this subsection, the screening criteria and rigorously assessment process will be introduced. This study focused on the remotely collaboration between human and real robots within unstructured construction environment. Thus, studies in which only systems with partial shared control and physical robots were included, excluding review papers, fully human-controlled or fully autonomous systems, soft robots, and miniature robots. Additionally, selected studies focused on applications in unstructured, dynamic, and cluttered environments typical of construction engineering, ensuring relevance to complex construction scenarios.

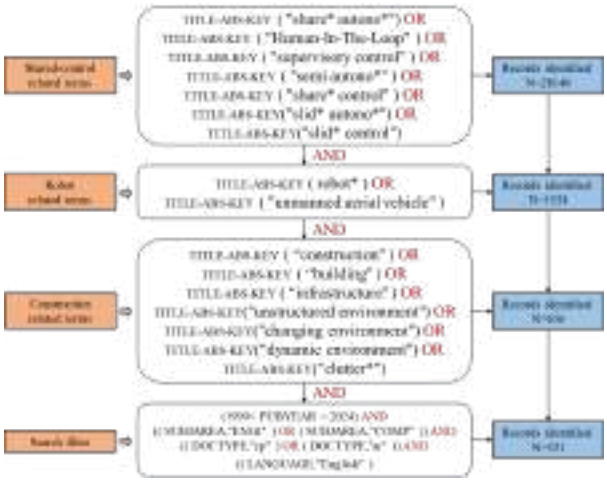


Figure 1. Combination of searching terms

To ensure rigor in the screening process and reduce subjective bias, a literature assessment team consisting of three PhD students was formed. The assessment process consists of three steps: (1) independent assessment of titles, abstracts, and keywords to identify relevant or uncertain studies; (2) independent full-text review of selected articles; and (3) consensus among raters, requiring at least two agreements for inclusion. Before final consensus, each rater followed the above screening criteria without any interference from others. By following the above criteria and process, a total of 65 relevant articles were finally included for detailed review.

3 Bibliometric Analysis

This section analyzes publications on human-robot collaboration based on shared control in unstructured environments, highlighting trends, robot types, and key venues.

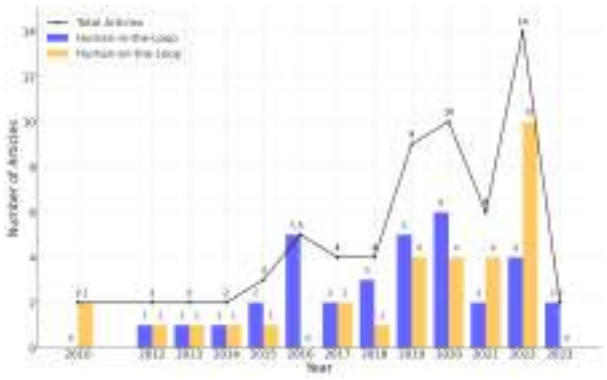


Figure 2. Numbers of publications of each shared control category by year

As shown in Figure 2, the annual publication count of

shared control systems has steadily increased, with significant growth since 2016 and a peak in 2022 (14 articles). The trends of "Human-in-the-Loop" (HITL) and "Human-on-the-Loop" (HOTL) indicate a shift towards supervisory roles, as HITL research peaked in 2020 while HOTL rapidly expanded in 2022, driven by advancements in AI and the demand for balancing automation and human oversight.



Figure 3. Distribution of robot types

Figure 3 shows that robot arms dominate the field (48.4%), with 90% being single-arm systems, reflecting their maturity in industry. Mobile robots (29.7%), led by UAVs (42%) and UGVs (58%), highlight their versatility, while mobile manipulators (21.9%) showcase rising interest in dynamic environments. Humanoids (4.3%) and excavators (5.7%) remain niche areas.

4 Classification of shared control based human-robot collaboration systems

Based on retrieved studies, a two-layer classification framework of human-robot collaboration systems is proposed in this section based on control sharing mechanisms. Table 1 shows a summary of the classification framework.

4.1 Human-in-the-loop system (HITLs)

Human-in-the-loop (HITL) systems involve the active participation of human operators within the robot's control loop, where human input is essential to ensure task execution. Depending on the mechanism by which human and robotic control are integrated, HITL systems can be categorized into three subtypes as shown in Figure 4: semi-autonomous human-in-the-loop systems, robot-guided human-in-the-loop systems, and human-instructed human-in-the-loop systems.

4.1.1 Semi-autonomous human-in-the-loop system (SA-HITLs)

SA-HITL refers to a shared control system capable of fusing motion commands from both human operators and autonomous controllers, which are then executed by the robot. In SA-HITL, the operator provides low-level motion commands in a human-in-the-loop manner, while the autonomous controller simultaneously generates motion commands to assist the operator in performing tasks such as obstacle avoidance in unstructured environments. To fuse the commands from both the operator and the autonomous controller, existing studies primarily adopt two methods: (1) degrees of freedom (DOF) allocation and (2) command fusion. In the DOF allocation method, the operator and the autonomous controller independently control different components or DOF of the robotic system. For example, [8] developed a shared control system for motor graders, where human commands control steering and autonomous commands control blade motion. Such studies typically assign certain robot DOFs, often those challenging to automate, to human control, enhancing the system's adaptability in uncertain environments.

In another type of SA-HITL, operator and robot commands act on the same DOFs and are fused, typically using linear, probabilistic or constraint optimization methods, to generate the final execution command. Adjusting fusion strategies, such as the weights in linear fusion, allows the system to better align with human intentions or objectives of autonomous controller. For example, a linear command fusion strategy was proposed [9] to utilize robot-generated obstacle avoidance commands to mitigate collision risks caused by reduced situational awareness during human teleoperation of robots in cluttered and unknown environments.

4.1.2 Robot-guided human-in-the-loop system (RG-HITLs)

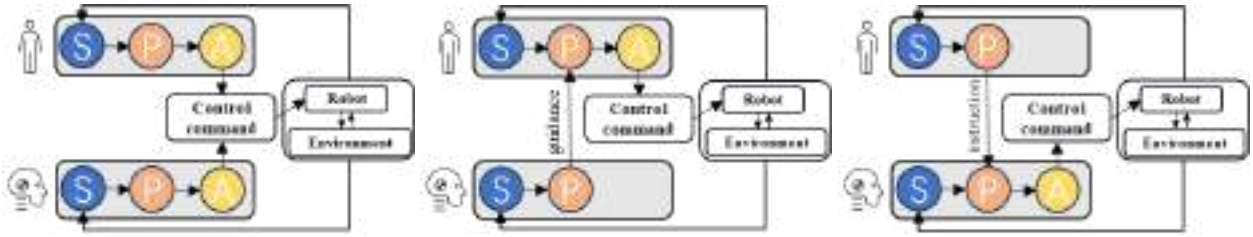
RG-HITL systems enable shared control by guiding, prompting, or constraining the operator's inputs, without the autonomous controller directly generating commands for the robot's actuators. This approach allows operators to effectively control the robot while leveraging the autonomous system's planning and situational awareness. RG-HITL systems can be classified into sensation-based and virtual fixture-based systems.

Sensation-based systems enhance operators' situational awareness by conveying environmental information to them through multimodal feedback, such as visual [10], auditory [11], or haptic cues [12]. For instance, a UAV haptic interface guided operators in navigating obstacle-laden environments, where tactile feedback not only reduced collision risks but also improved performance in scenarios with limited visibility [13].

Table 1. Classification of shared control based human-robot collaboration systems

Class	Subclass	Human Role	Robot Role	Techniques	Potential HRC applications in construction
HITL	SA-HITL [8, 9]	Provide control commands	Generates control commands, fuses operator commands, and executes.	DOF allocation [8]; Commands fusion [9]	Tasks with redundant robot's DOFs such as grading and rolling
	RG-HITL [10, 11, 12] [13, 14, 15]	Provide control commands with robot's guidance	Provide guidance to operator and executes operator's commands.	Visual[10], auditory [11], and haptic cues[12]; Virtual fixtures [14, 15]	Dynamic tasks requiring accurate control such as Welding or drilling
	HI-HITL [16, 17, 18] [19, 20]	Provide high-level planning instructions	Convert operator's high-level instructions into low-level planning, generate executable control commands, and execute them.	Geometric [16], template-based [19], and demonstration-based motion planning [17]	Tasks requiring high-complexity planning such as grasping structural demolition components
HOTL	HI-HOTL [21, 22, 23]	Supervise robot performance and intervene when necessary	Execute tasks autonomously and await human intervention when needed.	Gesture, haptic[21], or voice-based "take over" mechanisms [22]	Tasks with conditional robot autonomy such as drone-based construction site safety monitoring
	RI-HOTL [24, 25, 26]	Intervene when triggered by robot based on external information	Execute tasks autonomously, monitor system/environment state, and adjust autonomy.	Task failure risk prediction [25, 26], system gain prediction [24]	Tasks with high robot autonomy: drone-based 3D reconstruction of the exterior of buildings

Note: Due to page limitations, identified works for each shared control category cannot be fully indexed. The full list is provided in the appendix A.



(a) SA-HITL (b) RG-HITL (c) HI-HITL

Figure 4. Schematic diagram of three HITL shared control systems

The second type of RG-HITLs is virtual fixture-based, which employs attractors or repulsors to guide or constrain operator inputs. Attractor fixtures create a converging force field, directing the operator's control toward a desired trajectory or target, enhancing precision and simplifying tasks. For example, Adjigble et al. [14] applied attractor-based fixtures in robotic telemanipulation for industrial assembly, enabling stable and accurate peg-in-hole operations even in unstructured environments. Conversely, repulsor fixtures generate a diverging force field to prevent the operator from moving the robot into restricted or unsafe areas. For instance, Gottardi et al. [15] utilized repulsor fixtures in a teleoperated UAV for search-and-rescue missions, helping operators navigate around collapsed structures and debris, thereby reducing collision risks and enhancing safety in disaster response scenarios.

4.1.3 Human-instructed human-in-the-loop system (HI-HITLs)

Human-instructed human-in-the-Loop (HI-HITL) systems are characterized by a distinct interaction model where the operator provides high-level instructions rather than low-level motion commands. These high-level instructions are integrated by the robot's autonomous planner to generate detailed, low-level control commands. By eliminating the need for real-time low-level robot control, with HI-HITLs operators can focus on higher-level planning.

Components on construction sites tend to be irregular in shape and material compared to parts in manufacturing. To achieve robust autonomous gripping requires tedious computation, yet this is a subconscious action for humans. The HI-HITL system can effectively utilize human dexterity and tacit knowledge in grasping irregular objects to compensate for deficiencies in robotic reasoning. For example, existing studies have proposed that operators could instruct robots through (1) setting a target grasping pose [19], (2) selecting from a set of grasping candidates generated by the robot [20], or (3) defining geometric constraints for grasping [16, 17]. In addition to grasping assistance, another type of application for HI-HITL is in guiding mobile robots through complex environments. For example,

Jia et al. [18] used a graphical interface enabling operators to modify the robot's cost map by marking non-traversable obstacles that are difficult for the robot to perceive, improving path planning in cluttered environments.

4.2 Human-on-the-loop system (HOTLs)

Unlike the HITL system, the HOTL system places more emphasis on robot autonomy, i.e., the robot has the ability to perform tasks autonomously under ideal circumstances, but retains the authority of the human operator to intervene. Thus HOTL systems have the capability of control authority switching, which is also referred to as sliding control or supervisory autonomy in some contexts. Depending on how control authority switching is triggered, HOTL can be categorized into human-initiated HOTLs and robot-initiated HOTLs. These systems reduce operator workload, ensure reliability and safety, and balance efficiency with adaptability in complex environments by limiting intervention to critical situations.

4.2.1 Human-initiated human-on-the-loop system (HI-HOTLs)

HI-HOTL systems enable operators to "take over" control as needed, often through emergency stop buttons or GUI-based triggers in teleoperation scenarios. Other that these, haptic devices also play a role, allowing intuitive takeovers by detecting operator actions, such as moving the device, as demonstrated by Brosque et al. [21] in collaborative welding tasks. Additionally, systems can transition to operator control based on confidence in predicting the operator's intent, ensuring alignment with human goals.

While the reliability of the HI-HOTL system is ensured by human supervision, it relies on the operator's ability to understand the robot's state, environment and condition accurately and in a timely manner. Traditional approaches typically require operators to continuously monitor sensor data, plans, and environmental information, which can be unintuitive and cognitively demanding. To address this, Song et al. [22] introduced a system using visual-language models to present the robot's perception and motion planning in natural language, enhancing intuitiveness. Additionally, research aims to enable robots to learn from operator interventions, improving autonomy and reducing the need for future takeovers [23].

4.2.2 Robot-initiated human-on-the-loop system (RI-HOTLs)

RI-HOTL systems enable robots to operate autonomously while involving human operators only when necessary. By monitoring their state and environment, robots identify situations requiring intervention, relieving operators from constant monitoring. This further reduces operators' cognitive load and allows them to manage other tasks, improving efficiency and multitasking capabilities.

In existing RI-HOTL systems, the decision to involve human assistance is fundamentally rooted in predictions concerning the environment or task execution. This decision-making process primarily revolves around two key aspects:

1. **Estimating System Gains from Human Assistance:** For instance, hierarchical reinforcement learning has been utilized to predict the potential gains of human intervention, such as improving rescue efficiency and safety in cluttered environments. Here, the robot learns to decide when to involve the human operator to maximize task performance and ensure safety [24].

2. **Assessing Task Failure Risks:** Another approach involves evaluating the risks associated with task failures. For example, in [25], a 3D environmental uncertainities estimation model-based, risk identification and management framework was presented in which the failure probabilities of various type of mobile manipulator's task such as collisions and missed grasping were estimated. Except for utilizing task information, environmental information such as environmental accessibility [26] can also be used to regulate the autonomy of the system.

5 Challenges of deploying shared-control system in unstructured construction environment

In this section, we identify three key challenges associated with the deployment of shared control HRC in an unstructured construction environment.

5.1 Low latency bidirectional communication systems

A stable and low-latency bidirectional communication system is essential for shared control systems. Communication delays can cause command mismatches and feedback distortion, undermining operator control and robot performance [27]. This is particularly critical for SA-HITL, RG-HITL, and HI-HOTL frameworks. In SA-HITL and RG-HITL, operators provide real-time low-level control commands, and latency disrupts synchronization between operator inputs and robot actions. For instance, hydraulic excavators experience efficiency drops to 40–60% of manned operation levels due to communication delays [28]. Similarly, in HI-HOTL systems, where operators monitor and intervene, delays impair situational awareness, as seen in crane operations, which face increased task difficulty and extended completion times [29].

5.2 Intuitive human-robot interaction interface

An intuitive human-robot interaction interface is essential for shared control in construction, as it determines whether a construction worker, without robotics expertise, can accurately interpret the robot's state and remote environment, make appropriate decisions, and efficiently issue control commands with minimal errors. While most existing research on shared control focuses on the fusion mechanisms of human and robot commands, the impact of interaction interface design has received significantly less attention.

HRI interfaces can be categorized into feedback and control interfaces. Feedback interfaces are used to communicate information about the robot, the mission and the environment to the remote operator, and are therefore critical in SA-HITL, RG-HITL and HI-HOTL systems that require real-time operator feedback or supervision. Existing studies have shown that the addition of sensory dimensions such as vision, auditory and haptic can help improve situational awareness under teleoperation [30]. However, other studies have found that providing too much information to operators increases their cognitive load and, in turn, decreases their productivity. For example, relying only on visual modal information when teleoperating an excavator can overload the operator, cause fatigue and poor decisions to operators, and even damage underground utility facilities such as pipes and wires [31]. Therefore, what and how information is presented to the remote operator for different construction tasks and environments still needs a lot of human subject validation. Unlike feedback interfaces, the core challenge of control interfaces for human-robot collaboration in the field of construction engineering lies in the mismatch between limited robotic degrees of freedom and highly flexible human movement. Construction workers often require extensive training to overcome mismatches with field robots during teleoperation, thereby raising the barrier to widespread adoption.

5.3 Building trust between human and robots

Trust between humans and robots reflects the confidence and willingness of humans to rely on or cooperate with robots to accomplish tasks. Without some level of trust, it is difficult to apply any shared control method to a real construction site. Typically, the more autonomous a robot is, the more it needs human trust. For example, the HOTL system has the ability to operate fully autonomously compared to HITL, and therefore also requires more trust from the human supervisor.

To foster trust between humans and construction robots, key questions include: (1) Trust in Specific Autonomy Levels: Which shared control framework is most trusted for a given task and environment? (2) Transparency and Explainability: How can robotic systems improve transparency to enhance operator trust? (3) Impact of Training: Does continuous training increase operator trust and acceptance of higher autonomy?

6 Future trends

In this section, two future research trends identified from related works are introduced which helps to tackle with the above challenges as shown in Figure 5.

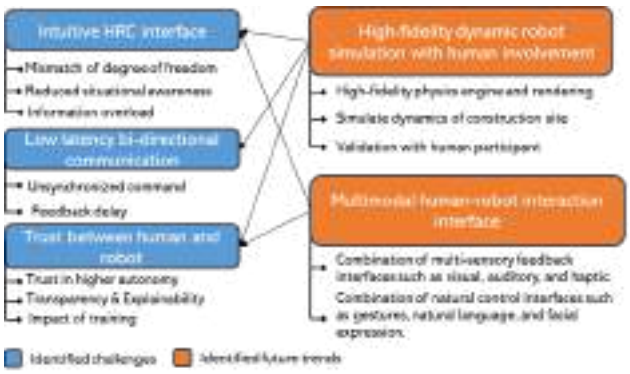


Figure 5. Mapping between identified challenges and future trends

6.1 High-fidelity dynamic robot simulation with human involvement

Compared to traditional static, low-fidelity robot simulations, shared control systems in unstructured construction environments require high-fidelity simulations to reflect the dynamics of the site and the operator's perception of the environment. High-fidelity means that the simulated environment accurately replicates the physical conditions affecting the robot and operator. For example, simulating a drone in a tunnel must account for lighting and airflow, while a robotic crane needs to model the dynamics of a pendulum in a wind or rush-stop situation to address payload swing. Similarly, simulations of inspection robots navigating incomplete structures should include uneven surfaces, temporary scaffolding, and obstacles such as materials or moving people to ensure effective robot adaptation. Future research should prioritize high-precision physics engines and advanced graphics rendering for realistic simulations. Digital twin technology can be used to capture the dynamics of the construction site and integrate real-time data to effectively simulate the ever-changing construction environment.

6.2 Multimodal human-robot interaction interface

Current research on shared control is primarily based on GUI to convey signals to operators in a visual format.

In recent years, auditory and haptic modalities have been shown to enhance remote operator situational awareness [30]. Thus, future research could focus on the integration of multiple sensory modalities, such as vision, auditory, and haptic, to provide remote operators with an immersive experience. For example, by leveraging vision-language models, robots can explain their understanding of the environment and their own actions to operators in natural language [22], while operators can also issue commands to the robots using natural language.

7 Conclusion

This paper reviews shared-control human-robot collaboration systems, focusing on their application in unstructured construction environments. A two-layer classification framework was proposed that categorizes these systems into "Human-in-the-Loop" (HITL) and "Human-on-the-Loop" (HOTL), highlighting varying autonomy levels and human involvement. Bibliometric analysis reveals significant growth, with a shift toward supervisory HOTL roles enabled by advances in autonomy. However, challenges persist, including limited focus on usability, and insufficient exploration of real-world deployment, particularly regarding trust-building, training, and adaptability. Future research should prioritize standardized frameworks, user-centered design, high-fidelity simulations, and multi-modal interaction interface to bridge the gap between theory and practice, accelerating adoption and improving safety, efficiency, and collaboration in construction robotics.

References

[1] Health and Safety Executive. Fatal injuries in great britain 2023/24, 2024. Accessed: 2024-12-05.

[2] Financial Times. Labour's homebuilding plans at risk from skills shortage, industry says. 2024.

[3] Ci-Jyun Liang, Xi Wang, Vineet R. Kamat, and Carol C. Menassa. Human–Robot Collaboration in Construction: Classification and Research Trends. *Journal of Construction Engineering and Management*, 147(10):03121006, October 2021.

[4] Haitao Wu, Heng Li, Xin Fang, and Xiaochun Luo. A survey on teaching workplace skills to construction robots. *Expert Systems with Applications*, 205: 117658, November 2022.

[5] Patrick B. Rodrigues, Rashmi Singh, Mert Oytun, Pooya Adami, Peter J. Woods, Burcin Becerik-Gerber, Lucio Soibelman, Yasemin Copur-Gencturk, and Gale M. Lucas. A multidimensional taxonomy for human-robot interaction in construction. *Automation in Construction*, 150:104845, June 2023.

[6] Ming Zhang, Rui Xu, Haitao Wu, Jia Pan, and Xiaowei Luo. Human–robot collaboration for on-site construction. *Automation in Construction*, 150: 104812, June 2023.

[7] Matthew J. etc. Page. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372:n71, March 2021.

[8] Peng Qi and Liang Gong. 3D Guidance Navigation and Human-Robot Shared Control for Motor Graders in Field Operation. In *2022 8th International Conference on Mechatronics and Robotics Engineering (ICMRE)*, pages 79–83, February 2022.

[9] P. Pappas, M. Chiou, G.-T. Epsimos, G. Nikolaou, and R. Stolkin. VFH+ based shared control for remotely operated mobile robots. In *IEEE Int. Symp. Saf., Secur., Rescue Robot., SSRR*, pages 366–373, 2020.

[10] N.V. Navkar, Z. Deng, D.J. Shah, K.E. Bekris, and N.V. Tsekos. Visual and force-feedback guidance for robot-assisted interventions in the beating heart with real-time MRI. In *Proc IEEE Int Conf Rob Autom*, pages 689–694, 2012.

[11] M. Larsson, H. Yoshida, and T. Igarashi. Human-in-the-loop fabrication of 3D surfaces with natural tree branches. In Spencer S.N., editor, *Proc.: SCF - ACM Symp. Comput. Fabr.* Association for Computing Machinery, Inc, 2019.

[12] F. Abi-Farraj, C. Pacchierotti, O. Arenz, G. Neumann, and P.R. Giordano. A Haptic Shared-Control Architecture for Guided Multi-Target Robotic Grasping. *IEEE Transactions on Haptics*, 13(2):270–285, 2020.

[13] Tomasz Piessens, Marinus M. van Paassen, and Max Mulder. UAV Haptic Interface for Dynamic Obstacle Avoidance. In *AIAA Scitech 2020 Forum*. American Institute of Aeronautics and Astronautics.

[14] M. Adjigble, N. Marturi, V. Ortenzi, and R. Stolkin. An assisted telemanipulation approach: Combining autonomous grasp planning with haptic cues. In *IEEE Int Conf Intell Rob Syst*, pages 3164–3171, 2019.

[15] Alberto Gottardi, Stefano Tortora, Elisa Tosello, and Emanuele Menegatti. Shared Control in Robot Teleoperation With Improved Potential Fields. *IEEE Transactions on Human-Machine Systems*, 52(3): 410–422, June 2022.

[16] M. Gridseth, O. Ramirez, C.P. Quintero, and M. Jagersand. ViTa: Visual task specification interface for manipulation with uncalibrated visual servoing. In *Proc IEEE Int Conf Rob Autom*, volume 2016-June, pages 3434–3440, 2016.

[17] Rafael Papallas, Anthony G. Cohn, and Mehmet R. Dogar. Online Replanning With Human-in-the-Loop for Non-Prehensile Manipulation in Clutter — A Trajectory Optimization Based Approach. *IEEE Robotics and Automation Letters*, 5(4):5377–5384, October 2020.

[18] Raymond Jia, Nathanael Koh, Nicholas Leone, Mohit Singh, Zixuan Wu, and Patricio Vela. Teleoperation of Semi-autonomous Robots Through Uncertain Environments. In *2022 Opportunity Research Scholars Symposium (ORSS)*, pages 31–37, April 2022.

[19] A. Romay, S. Kohlbrecher, D.C. Conner, A. Stumpf, and O. Von Stryk. Template-based manipulation in unstructured environments for supervised semi-autonomous humanoid robots. In *IEEE-RAS Int. Conf. Humanoid Rob.*, volume 2015-February, pages 979–986, 2015.

[20] C. Pohl, K. Hitzler, R. Grimm, A. Zea, U.D. Hanebeck, and T. Asfour. Affordance-based grasping and manipulation in real world applications. In *IEEE Int Conf Intell Rob Syst*, pages 9569–9576. Institute of Electrical and Electronics Engineers Inc., 2020.

[21] C. Brosque, E.G. Herrero, Y. Chen, R. Joshi, O. Khatib, and M. Fischer. Collaborative Welding and Joint Sealing Robots With Haptic Feedback. In *Proc. Int. Symp. Autom. Robot. Constr.*, volume 2021-November, pages 1–8, 2021.

[22] Y. Song, P. Sun, P. Fang, L. Yang, Y. Xiao, and Y. Zhang. Human-in-the-loop Robotic Grasping using BERT Scene Representation. In *Proc. Main Conf. Int. Conf. Comput. Linguist., COLING*, volume 29, pages 2992–3006. Association for Computational Linguistics (ACL), 2022.

[23] W. Si, Y. Guan, and N. Wang. Adaptive Compliant Skill Learning for Contact-Rich Manipulation With Human in the Loop. *IEEE Robotics and Automation Letters*, 7(3):5834–5841, 2022. Publisher: Institute of Electrical and Electronics Engineers Inc.

[24] A. Hong, O. Igharoro, Y. Liu, F. Niroui, G. Nejat, and B. Benhabib. Investigating Human-Robot Teams for Learning-Based Semi-autonomous Control in Urban Search and Rescue Environments. *Journal of Intelligent and Robotic Systems: Theory and Applications*, 94(3-4):669–686, 2019. Publisher: Springer Netherlands.

[25] S. Al-Hussaini, N. Dhanaraj, J.M. Gregory, R.J. Joseph, S. Thakar, B.C. Shah, J.A. Marvel, and S.K. Gupta. Seeking Human Help to Manage Plan Failure Risks in Semi-Autonomous Mobile Manipulation. *Journal of Computing and Information Science in Engineering*, 22(5), 2022. Publisher: American Society of Mechanical Engineers (ASME).

[26] H. Belaidi, A. Hentout, and H. Bentarzi. Human–Robot Shared Control for Path Generation and Execution. *International Journal of Social Robotics*, 11(4):609–620, 2019.

[27] Miran Seo, Samraat Gupta, and Youngjib Ham. Exploratory study on time-delayed excavator teleoperation in virtual lunar construction simulation: Task performance and operator behavior. *Automation in Construction*, 168:105871, 2024.

[28] Yuzuki Okawa, Masaru Ito, Ryota Sekizuka, Seiji Saiki, Yoichiro Yamazaki, and Yuichi Kurita. An assistive interface of a teleoperation system of an excavator by overlapping the predicted position of the arm. pages 74–78. International Association for Automation and Robotics in Construction (IAARC), October 2020.

[29] Joshua Vaughan, Kelvin Chen Chih Peng, William Singhose, and Warren Seering. Influence of remote-operation time delay on crane operator performance. *IFAC Proceedings Volumes*, 45(14):85–90, 2012.

[30] Pengxiang Xia, Fang Xu, Zhuoyuan Song, Shuai Li, and Jing Du. Sensory augmentation for subsea robot teleoperation. *Computers in Industry*, 145:103836, 2023.

[31] Yogeeta Desai, Dorian Davis, Steven Jiang, and Alexandria Ward. The effect of auditory cues on haptic-controlled excavator operator performance. *IISE Annual Conference.Proceedings*, pages 825–832, 2014.

A List of Identified Works

The full list of identified works can be found here.

Effects of Visual Fidelity on Task Performance and Mental Workload in IVEs

Jin-Bin Im[1], Rong-Lu[1], Hong, Jae-Ho Jang[1], Enlian Zhang[1], Seung-Hye Shin[1], Moonboo Joo[1], Ki-Won Lee[1], Kang-Moo Lee[1], Kyung-Ho Lee[1], Seo-Young Park[1] and Ju-Hyung Kim[1]

[1]Department of Architectural Engineering, Hanyang University, Republic of Korea
jbim0202@hanyang.ac.kr, abraham0814@hanyang.ac.kr, 0404mm@naver.com, eunryeon@hanyang.ac.kr, sshye@hanyang.ac.kr, joomb@hanyang.ac.kr, ondamco@gmail.com, es8282@naver.com, mbl00@naver.com, fosys8651@hanmail.net, kcr97jhk@hanyang.ac.kr

Abstract –

Visual fidelity affects the user experience and performance in virtual environments. While past studies have explored its effects, most have relied on qualitative measures, and few quantitative studies have specified visual quality standards. This study examined the influence of visual fidelity on task performance and mental workload in immersive virtual environments (IVEs). Two environmental models were tested in virtual reality (VR) and mixed reality (MR): a high-visual quality model featuring detailed materials, realistic lighting, and outdoor scenery, and a low-fidelity research model with basic colors, dimensional lighting, and no external views. Sixty participants completed cognitive tasks (digit span, spatial working memory updating, go/no-go), while the EEG measured mental workload. The results showed that high-fidelity increased the cognitive load in complex tasks, whereas MR environments supported better performance. These findings emphasize the need to balance realism and cognitive demands to optimize virtual experiences.

Keywords –

Immersive Virtual Environments; Visual Fidelity; Task Performance; Mental Workload

1 Introduction

Immersive Virtual Environments (IVEs) are increasingly being used in research and industry for training, simulation, and entertainment. Advancements in head-mounted display (HMD) hardware and software have enhanced visual fidelity by incorporating detailed textures, realistic lighting, and function-focused representations [1]. Visual fidelity, defined as the level of realism and detail in rendered virtual environments (VEs), is widely assumed to enhance the user's sense of presence (SoP), the subjective perception of "being there" [2].

However, previous studies have primarily focused on qualitative methods, such as surveys and interviews [3], leaving the impact of visual fidelity on objective measures of task performance and mental workload less explored. Task performance is a critical factor in indoor behavior and influences business and educational outcomes [4,5]. Mental workload, which represents cognitive effort during work or study, is influenced by complexity and environmental outcomes [6,7]. While high-fidelity elements enhance realism and immersion, they may also increase the cognitive load by demanding greater processing resources, particularly in scenarios requiring complex decision-making or problem-solving within Virtual (VR) and Mixed Reality (MR) [8,9].

This study aimed to investigate the influence of high and low visual fidelity on task performance and mental workload in IVEs. By comparing VR and MR settings, we provide insights into leveraging visual presentation to balance performance and cognitive demands.

2 Literature Reviews

Visual fidelity is often correlated with user engagement, presence, and satisfaction with a VEs [10]. Previous studies have suggested that realistic lighting and detailed textures significantly enhance SoP [11,12]. However, high-fidelity elements also require greater computational resource and may inadvertently increase the cognitive load [13]. Experimental investigations on IVEs have shown that variations in environmental realism can significantly affect user experience [14,15]. Beyond a certain threshold, further enhancements in graphical detail may yield diminishing returns on engagement and fail to substantially improve task performance [16].

Milgram and Kishino [17] described the Reality-Virtuality Continuum, bridging the gap between fully virtual and real environments. Whereas VR places the user in a completely simulated space, whereas MR overlays digital objects in the real-world. This dual

processing may lead to different cognitive demands than those in the VE. Supporting this notion, Hong et al. [18] demonstrated that users' subjective evaluations of an urban soundscape did not significantly differ between real and IVEs, suggesting a certain level of equivalence in the perceived experience.

Physical environmental factors, such as ceiling height, nature elements, and window size, are well known to influence work performance [19–22]. The translation of these insights into IVEs, particularly those designed with varying degrees of visual fidelity, is an area of ongoing research. High visual fidelity can enhance the user's SoP and realism in virtual experiences. However, they may also increase the cognitive load, especially during complex decision-making or problem-solving tasks. For instance, a study on visuospatial problem-solving in MR environments found that although high-fidelity visuals improved task accuracy, they also led to longer completion times, indicating increased cognitive effort [23]. Similarly, research on the cognitive load in 3D Ves suggests that higher visual complexity can elevate the cognitive load, potentially affecting user performance [9].

This gap in understanding, explains how visual fidelity in immersive environments influences task performance, mental workload. By systematically examining user performance and mental workload across varying levels of visual realism, we aimed to clarify whether increased visual fidelity truly benefits VE users. While previous research has highlighted the potential of high-fidelity environments to enhance SoP and immersion, their impact on practical outcomes, such as cognitive efficiency and task performance, remains underexplored.

3 Methods

3.1 Experimental Settings

3.1.1 Physical Space

An L-shaped classroom was selected as the experimental setting for this study, because its layout is particularly relevant for examining the cognitive load and occupant engagement while reducing distractions [24]. The physical space used was 15.5 m^2 L-shaped educational facility in Seoul, Republic of Korea. Thes room was furnished with basic furniture, including four tables and chairs, a television, two pendent lights, bar tables and chairs, and air conditioners. In addition, the wall interior was painted using a green diagonal design and punched plate finishing. The features of this room, along with general images, are shown in Figure 1.

3.1.2 Virtual Spaces

Two distinct VEs were created in this study. First, the virtual setting was modelled in Unity 3D using a 3D model developed from the original CAD floor plan in Revit software. Second, the 3D modelling file was exported as an fbx file and imported into the Unity 3D environment. Finally, detailed virtual settings were designed in the Unity 3D engine such as texture, lighting, and outdoor scenery. This design process was performed based on different visual fidelity levels, with high and low quality implemented in each VR and MR devices.

3.1.3 Visual Quality Levels in VEs

Each virtual environment was presented at two visual quality levels, following the framework outlined by Graham et al. [25]. The authors classified visual fidelity into two levels of detail (LOD 1 and LOD 2) in VR environments, using identical base modelling files. In LOD 1, textures were applied to 3D models using a generic texture library available on software platforms



Figure 1. Floor plan and cross-section photos of the physical space

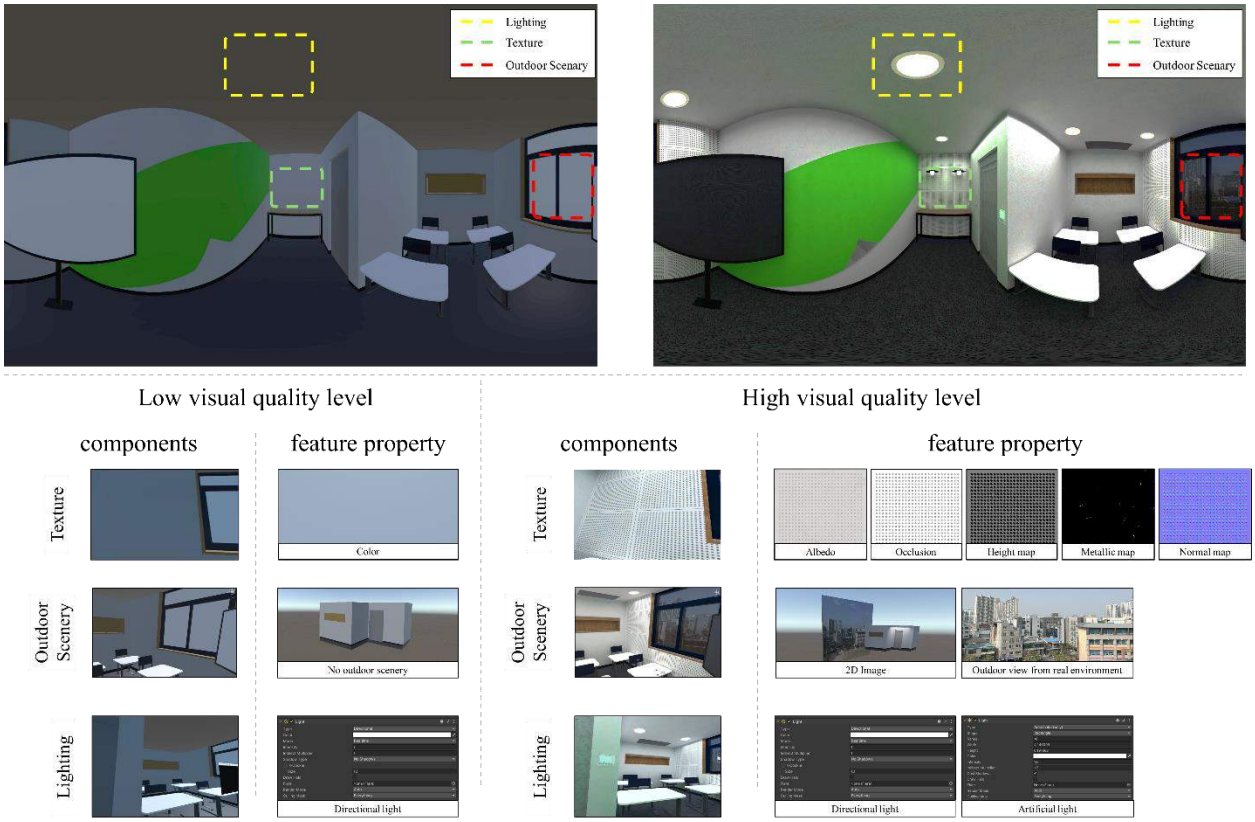


Figure 2. Description of LQ and HQ visual fidelity features in VEs

such as Autodesk Revit, Rhinoceros, or Unity 3D. Conversely, LOD 2 incorporates highly detailed texture features, leveraging advanced techniques such as detail maps (including normal, height, and adobe maps) or data derived from laser scanning and photogrammetry.

In this study, the High-Quality (HQ) visual fidelity level includes realistic artificial indoor lighting that enhanced the immersive quality of the environment. In addition, the outdoor scenery visible through windows was modelled using real photographs, further contributing to the realism of the virtual environment. However, the Low-Quality (LQ) visual fidelity level relies solely on directional lighting provided by Unity 3D, which simplifies the lighting setup. Unlike the HQ, the LQ does not include outdoor scenery, which results in a less immersive virtual experience.

3.2 Subjects

A one-way analysis of variance (ANOVA) was conducted. Based on the central limit theorem and assumption of normality in the distribution for statistical analysis, a sample size of 30 participants was used [26]. The specific sample size was calculated using G*power 3.1 software. We used an ANOVA: fixed effects, omnibus, one-way function with an effect size of 0.60 as the middle scale, an alpha error probability of 0.05, and a power level of 0.80 with three groups.

As the experimental groups were conducted using VR, MR, and real-world settings, a total of 50 participants was calculated, and 60 subjects were recruited, considering a dropout rate of 20%. All the study procedures were approved by the Institutional Review Board and Ethics Committee of Hanyang University (HYUIRB-202406-014). In total, 60 subjects participated in this study, none of whom had a history of mental or physical health problems. Among the 60 collected data points, three were excluded because of poor data quality, two from the HQ group and one from the LQ group.

The study sample consisted of 60 participants aged 20 – 32 years, reflecting the target demographics of university students. Among the participants, 42 were male, and 18 were female. The average age for male participants is 25.43 years, and the average age for female participants is 24.06 years. The overall average age of the sample is 25.02 years. The decision to recruit participants primarily in their twenties aligns with the study's focus on classroom environments, as university students represent the primary users

3.3 Measurements

3.3.1 Task Performance

Cognitive tests are widely used to assess domains such as executive function, attention, memory, intelligence, language ability, and reasoning [27]. These cognitive domains are regarded as key indicators of concentration and predictors of future job performance [28]. While specific cognitive requirements may vary depending on job demands, executive function, attention, and memory are fundamental to general work performance [29].

To evaluate productivity across different spatial environments, three cognitive tests were conducted, the 2-back test, spatial working memory updating, and go/no-go test. In an MR environment, where holographic information is overlaid onto the real world, the use of a traditional keyboard can be challenging for users. To address this limitation, cognitive tests that rely on simple button interactions, such as pressing a space bar or number keypads, were adopted.

The 2-back test presents sequential visual stimuli to the users, requiring them to memorize and respond correct time [30]. In the 2-back test, participants must identify when the current stimulus matches the one presented two steps earlier, necessitating continuous updating of information.

Spatial working memory updating was utilized to assess memory ability. The original test uses a 2x2 matrix displayed on the screen, in which a color-filled square briefly appeared in one spot before disappearing [31]. In the updated version of the test, a 3x3 blank square matrix was used to further assess participants' memory abilities.

The go/no-go test is an intuitive measure of participants' attention ability [32]. In this test, a green or checkerboard pattern sphere appeared on the screen to evaluate the response time and accuracy. This test effectively gauges participants' ability to focus, process stimuli, and exercise inhibitory control.

To quantify learning performance, accuracy and response time were recorded for all cognitive tests. The learning performance was calculated using Equation (1) and employed as the primary indicator [3].

3.3.2 Mental Workload

The experiments imposed a cognitive load on participants, making it necessary to quantify and measure this load. The cognitive load experienced by participants during the task increased as they completed the given assignments, with workload and stress being the primary factors involved in the process. To assess this, we used the EPOC X+ device to collect raw EEG data. Brainwave data were extracted from the frontal (AF3, F7, F3, FC5, FC6, F4, F8, and AF4), parietal (T7, P7, P8, and T8), and occipital lobes (O1 and O2 through 14 electrodes, capturing frequencies within the range of 0.16 to 43 Hz.

Raw data often contain noise, such as eye blinks, head movements, and muscle activity, which necessitates a noise removal process. A finite-impulse response (FIR) filter was applied to eliminate these artifacts, and data outside the 4-40 Hz range were discarded. Next, an independent component analysis (ICA) was performed to remove impractical features from the filtered EEG data. Finally, the collected EEG data were classified into theta, alpha, beta, and gamma wave categories using fast Fourier transform (FFT). This method, which is widely validated in neuroscience, allows for the differentiation of brainwave activity. In simple term, theta waves are typically active during states of relaxation [34], alpha waves during calmness [35], beta waves during thinking or when experiencing tension [36], and gamma waves during heightened focus and concentration [37].

To conduct an in-depth analysis of the cognitive load during tasks, the mental workload was calculated and used as indicators. The methods for measuring each indicator follow Equations (2) [38].

3.4 Research Procedure

Each experiment was conducted in the same classroom with a modelled environment to prevent misunderstandings regarding the spatial layout. Participants underwent a preparation phase that consisted of an explanation of
the experiment, a Short Portable Mental Status Questionnaires (SPMSQ), and a short training of VR and MR to familiarize themselves with the navigation and interaction features in VE. The SPMSQ was used to check whether the participants had cognitive impairment [39]. Including the preparation phase, all VE experiences using the HMD device were limited to two minutes to prevent motion sickness, cyber-sickness, and dizziness. To measure brainwaves, the participants wore an EEG device on the MR HMD. After the preparation phase, the subjects performed three cognitive tests in each experimental environment, and the accuracy rate, response time, and brain wave data were collected. Cognitive tests were conducted in random order to prevent ordering effects [28].

$$Learning\ Performance = \left[(Accuracy, \%)^{0.5} \times \frac{1}{(Response\ Time,\ ms)^{0.5}}\right]^2 \times 100 \quad (1)$$

$$Mental\ workload = \frac{(Relative\ Beta)_{Frontal}}{(Relative\ Theta)_{Frontal} + (Relative\ Alpha)_{Frontal}} \quad (2)$$

4 Results

4.1 Task Performance Result

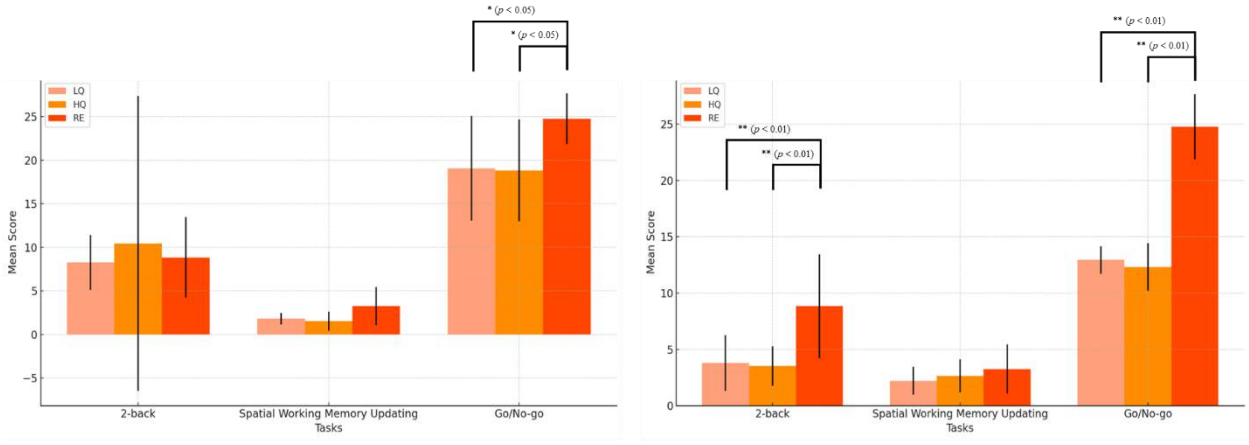
For the 2-back task, no significant differences in performance were observed among LQ, HQ, and RE in VR ($p = 0.350$). However, in MR, a significant difference was observed ($p < 0.001$), with RE outperforming HQ and LQ ($p < 0.01$). The performance advantage of MR may stem from the added realism of the real environment, which likely reduces the cognitive load.

In the Spatial Working Memory Updating Task (SWM), performance remained consistent across LQ, HQ, and RE in both VR ($p = 0.077$) and MR ($p = 0.356$). This indicates that visual fidelity does not have a strong impact on working memory tasks, regardless of the medium.

For the go/no-go task, RE significantly outperformed HQ and LQ across both VR ($p < 0.05$) and MR ($p < 0.001$). In both environments, RE outperformed HQ and LQ, with a post-hoc analysis showing that RE was significantly better than both HQ and LQ. This highlights the advantage of intuitive familiarity of the real environment in tasks requiring rapid decision-making and response inhibition.

Table 1. Statistical results of task performance across visual fidelity in VR and MR environments

Scenario	Task	Method	Type	Statistical Results					Post hoc
				Sum of squares	Df	Mean sqaure	F	p	
VR	2B	Kruskal Wallis	LQ HQ RE	2.101	2	-	-	0.350	-
	SWM	ANOVA-Welch	LQ HQ RE	20.7747	2	10.3874	4.8596	0.014* (0.077)	-
	GN	ANOVA	LQ HQ RE	271.1824	2	135.5912	5.1799	0.011*	HQ=LQ<RE
MR	2B	ANOVA	LQ HQ RE	204.3144	2	102.1572	9.4686	0.000*	HQ=LQ<RE
	SWM	ANOVA	LQ HQ RE	6.2397	2	3.1199	1.0706	0.356	-
	GN	ANOVA-Welch	LQ HQ RE	1121.5860	2	560.7930	113.9649	0.000*	HQ=LQ<RE

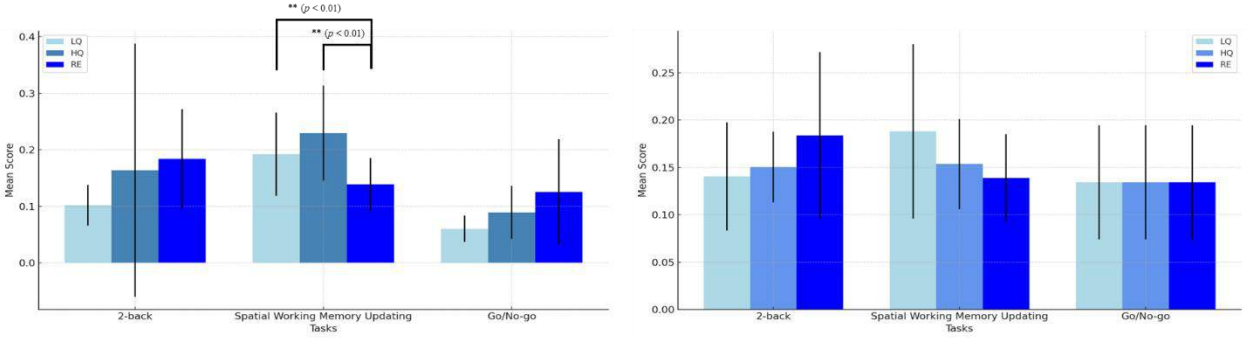


(a) VR Environment (b) MR Environment

Figure 3. Comparison of task performance across visual fidelity in VR and MR environments

Table 2. Statistical results of mental workload across visual fidelity in VR and MR environments

Scenario	Task	Method	Type	Statistical Results					Post hoc
				Sum of squares	Df	Mean sqaure	F	p	
VR	2B	Kruskal Wallis	LQ HQ RE	5.592	2	-	-	0.061	-
	SWM	ANOVA-Welch	LQ HQ RE	0.026	2	0.0128	3.3464	0.048* (0.036)	RE<LQ=HQ
	GN	ANOVA-Welch	LQ HQ RE	0.049	2	0.0247	3.1715	0.055 (0.027)	-
MR	2B	ANOVA-Welch	LQ HQ RE	0.012	2	0.006	1.3644	0.271	-
	SWM	ANOVA	LQ HQ RE	0.007	2	0.0033	0.6591	0.525	-
	GN	ANOVA	LQ HQ RE	0.002	2	0.0008	0.2246	0.800	-



(a) VR Environment (b) MR Environment

Figure 4. Comparison of mental workload across visual fidelity in VR and MR environments

4.2 Mental Workload Result

In terms of mental workload, no significant differences were observed for the 2-back task across the environments in VR (p = 0.061) and MR (p = 0.271). This indicated that the cognitive load associated with the 2-back task was not significantly affected by the environment or visual fidelity levels.

For SWM, VR showed a significant difference in mental workload (p = 0.048), with post hoc analysis indicating a higher cognitive load in HQ than in RE. In contrast, no significant differences were observed in MR (p = 0.525), suggesting that visual fidelity did not affect the mental workload for this task in MR

The go/no-go task exhibited no significant differences in the mental workload across environments or media, as indicated by the results for both VR (p = 0.055) and MR (p = 0.800). This suggests that the visual fidelity levels did not significantly influence the cognitive load in this task.

5 Conclusion

This study examined the nuanced relationships among visual fidelity, task performance, and mental workload in VR and MR environments. For the 2-back task, the significant difference observed in MR but not in VR underscores the importance of considering specific medium when designing IVEs. In MR, the presence of real-world elements or cues may enhance task performance by providing familiar and intuitive spatial references that can reduce the cognitive demands. However, SWM showed consistent performance across all environments in both VR and MR, suggesting that

visual fidelity had minimal effect on tasks involving working memory. By contrast, excessive visual detail in the HQ can burden cognitive resources without contributing to performance improvements. For the go/no-go task, the real environment consistently outperformed the VEs in both the VR and MR tasks, emphasizing the role of familiarity and intuitive interactions in optimizing task efficiency. The mental workload did not vary significantly for this task, indicating that the cognitive demands of response inhibition were stable regardless of environmental realism.

While this study was conducted in a L-shaped classroom setting, the cognitive process examined, such as executive function, memory, attention, are fundamental to human performance and can have implications beyond this specific environment. In real-world construction settings, particularly in open-plan sites and industrial environments, workers often navigate complex and unpredictable surroundings that require continuous cognitive adaptation. Given these considerations, our findings may provide valuable insights for applications in virtual construction environments, such as hazard perception training, safety simulation, and the understanding of complex procedural sequences. By integrating these cognitive insights into virtual construction platforms, training programs can replicate real-world conditions precisely, allowing for improved spatial awareness.

To enhance the generalizability of this study, future research should explore the way in which cognitive mechanisms influence decision-making and situational awareness in active construction environments. Expanding this research into dynamic, real-world settings could further validate its applicability in construction safety management.

Acknowledgement

This research was funded by the National Research Foundation of Korea (NRF) grant funded by the Korean government (Ministry of Science and ICT) (grant number RS202400356697).

References

[1] W.R. Sherman and A.B. Craig, Understanding Virtual Reality Interface, Application, and Design. Morgan Kaufmann, 2018.

[2] M. Slater and A. Steed, A virtual presence counter, Presence: Teleoperators and Virtual Environments. 9(5):413–434, 2000.

[3] L. Bergefurt, M. Weijs-Perrée, R. Appel-Meulenbroek, and T. Arentze, The physical office workplace as a resource for mental health – A systematic scoping review, Building and Environment. 207, 108505, 2022.

[4] H.W. Brink, M.G.L.C. Loomans, M.P. Mobach, and H.S.M. Kort, Classrooms' indoor environmental conditions affecting the academic achievement of students and teachers in higher education: A systematic literature review, Indoor Air. 31(2):405–425, 2021.

[5] H.W. Brink, S.C.M. Lechner, M.G.L.C. Loomans, M.P. Mobach, and H.S.M. Kort, Understanding how indoor environmental classroom conditions influence academic performance in higher education, Facilities. 42(3/4):185–200, 2023.

[6] J.M. Juliano, N. Schweighofer, and S.L. Liew, Increased cognitive load in immersive virtual reality during visuomotor adaptation is associated with decreased long-term retention and context transfer, Journal of NeuroEngineering and Rehabilitation. 19(1):1–14, 2022.

[7] H. Cui, R. Cui, D. Gong, W. Ma, and D. Yao, An EEG study of the brain workload under virtual reality and real environment, International Journal of Psychophysiology. 168, S163, 2021.

[8] F. Putze, C. Herff, C. Tremmel, T. Schultz, and D.J. Krusienski, Decoding Mental Workload in Virtual Environments: A fNIRS Study using an Immersive n-back Task, Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS. pages 3103–3106, Berlin, Germany, 2019.

[9] M. Rettinger, X. Jiang, J. Yang, and G. Rigoll, Visual Complexity in VR: Implications for Cognitive Load, Proceedings - 2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops, pages 783–784, Orlando, USA, 2024.

[10] H. Iriye and H.H. Ehrsson, Perceptual illusion of body-ownership within an immersive realistic environment enhances memory accuracy and re-experiencing, iScience. 25(1):103584, 2022.

[11] I. Yu, J. Mortensen, P. Khanna, B. Spanlang, and M. Slater, Visual realism enhances realistic response in an immersive virtual environment - Part 2, IEEE Computer Graphics and Applications. 32(6):36–45, 2012.

[12] W. Natephra, A. Motamedi, T. Fukuda, and N. Yabuki, Integrating building information modeling and virtual reality development engines for building indoor lighting design, Visualization in Engineering. 5(1), 2017.

[13] A. Mostafavi, M. Vujovic, T.B. Xu, and M. Hensel, Impacts of Illuminance and Correlated Color Temperature on Cognitive Performance: A VR-Lighting Study, arXiv, 2024.

[14] J.R.J. Neo, A.S. Won, and M.M.C. Shepley,

Designing Immersive Virtual Environments for Human Behavior Research, Frontiers in Virtual Reality. 2, 603750, 2021.

[15] K. Graham., C. Pybus, N. Arellano, J. Doherty, L. Chow, S. Fai, and T. Grunt, Defining geometry levels for optimizing BIM for VR: Insights from traditional architectural media, Technology| Architecture+ Design, 3(2):234-244. 2019

[16] A. Gronowski, D.C. Arness, J. Ng, et al., The impact of virtual and augmented reality on presence, user experience and performance of Information Visualisation, Virtual Reality. 28(3):1–10, 2024.

[17] P. Milgram, A Taxonomy of Mixed Reality Visual Displays. 1994.

[18] J.Y. Hong, B. Lam, Z.T. Ong, K. Ooi, W. S. Gan, J. Kang, S. Yeong, I. Lee, and S. T. Tan, A mixed-reality approach to soundscape assessment of outdoor urban environments augmented with natural sounds, Building and Environment. 194 107688, 2021.

[19] J. Kim, S.H. Cha, C. Koo, and S. keung Tang, The effects of indoor plants and artificial windows in an underground environment, Building and Environment. 138:53–62, 2018.

[20] S.H. Cha, C. Koo, T.W. Kim, and T. Hong, Spatial perception of ceiling height and type variation in immersive virtual environments, Building and Environment. 163, 2019.

[21] D. Jung, D.I. Kim, and N. Kim, Bringing nature into hospital architecture: Machine learning-based EEG analysis of the biophilia effect in virtual reality, Journal of Environmental Psychology. 89, 2023.

[22] J.B. Im, M. Joo, K.T. Lee, and J.H. Kim, A mixed reality approach to identify cognitive performance and mental states in preferred vs. non-preferred individual and collaborative work environments, Building and Environment. 274:112806, 2025.

[23] D. Garkov, T. Piselli, E. Di Giacomo, K. Klein, G. Liotta, F. Montecchiani, and F. Schreiber, Collaborative Problem Solving in Mixed Reality: A Study on Visual Graph Analysis, arXiv, 2024.

[24] P.C. Lippman, The L-Shaped Classroom: A Pattern for Promoting Learning., DesignShare (NJ1). 2004.

[25] K. Graham, L. Chow, and S. Fai, From BIM to VR: Defining a level of detail to guide virtual reality narratives, Journal of Information Technology in Construction. 24:553–568, 2019.

[26] L.M. Sullivan, J. Weinberg, and J.F. Keaney, Common statistical pitfalls in basic science research, Journal of the American Heart Association. 5(10), 2016.

[27] A. Diamond, Executive functions, Annual Review of Psychology. 64: 135–168, 2013.

[28] K.T. Lee and J.H. Kim, Relationship between psycho-physiological indicators and task performance under various indoor space designs for telecommuting environment by introducing mixed-reality, Scientific Reports. 14(1): 2024.

[29] D.-Y. Hwang, S.-H. Ryu, K.-H. Kwon, C.-R. Choi, and S.-A. Kim, Correlation Between Cognitive Impairment Screening Test (CIST), Korean-Mini Mental State Examination, 2nd Edition (K-MMSE~2) and Clinical Dementia Rating (CDR) of Patients with Stroke, Therapeutic Science for Rehabilitation. 11(2):53–62, 2022.

[30] D. V. Meegan, R. Purc-Stephenson, M.J.M. Honsberger, and M. Topan, Task analysis complements neuroimaging: an example from working memory research, NeuroImage. 21(3):1026–1036, 2004.

[31] T.A. Salthouse, R.L. Babcock, and R.I. Shaw, Effects of Adult Age on Structural and Operational Capacities in Working Memory, 1991.

[32] R. Ratcliff, A Theory of Memory Retrieval, Psychological Review, 82(2), 1978.

[33] H. Kim, T. Hong, J. Kim, and S. Yeom, A psychophysiological effect of indoor thermal condition on college students' learning performance through EEG measurement, Building and Environment. 184:107223, 2020.

[34] W. W. Ismail, M. Hanif, S. B. Mohamed, N. Hamzah, and Z. I. Rizman, Human emotion detection via brain waves study by using electroencephalogram (EEG). International Journal on Advanced Science, Engineering and Information Technology, 6(6):1005-1011, 2016

[35] I.A. Rahman and A.H. Memon, The Way Forward in Sustainable Construction: Issues and Challenges, Article in International Journal of Advances in Applied Sciences. 2013.

[36] N. Kumar, K. Khaund, and S.M. Hazarika, Bispectral Analysis of EEG for Emotion Recognition, Procedia Computer Science. 84:31–35, 2016.

[37] S. Nishifuji, M. Sato, D. Maino, and S. Tanaka, Effect of acoustic stimuli and mental task on alpha, beta and gamma rhythms in brain wave, In Proceedings of SICE Annual Conference. pages 1548–1554, Taipei, Taiwan, 2010.

[38] N.R. Bailey, M.W. Scerbo, F.G. Freeman, P.J. Mikulka, and L.A. Scott, Comparison of a brain-based adaptive system and a manual adaptable system for invoking automation, Human Factors. 48(4), 693-709, 2006.

[39] G. Wolber, M. Romaniuk, E. Eastman, and C. Robinson, Validity of the Short Portable Mental Status Questionnaire with elderly psychiatric patients, Journal of Consulting and Clinical Psychology. 52(4):712–713, 1984.

Design considerations of ergonomic human-robot collaboration (EHRC) for modular construction manufacturing

Yonglin Fu[1] and Weisheng Lu[2]

[1, 2]Department of Real Estate and Construction, The University of Hong Kong, Pokfulam Road, Hong Kong, China
u3009531@connect.hku.hk, wilsonlu@hku.hk

Abstract –

Manufacturing is a critical stage in modular construction, as it directly affects cost, schedule, and quality. Human-robot collaboration (HRC) has the potential to enhance modular construction manufacturing (MCM) productivity while preserving essential flexibility. Despite these benefits, previous studies on designing HRC workplaces for MCM have primarily focused on efficiency, often overlooking ergonomics and human factors. This oversight can negatively affect workers' occupational safety and health (OSH) and overall well-being. Therefore, this research aims to promote the concept of ergonomic HRC (EHRC) and inform the EHRC design for MCM. Three cases are further presented to demonstrate the design of EHRC under different variables and considerations. This study encourages researchers and practitioners to place human well-being at the center of the production system.

Keywords –

Human-robot collaboration (HRC); modular construction; ergonomics; human-centric HRC; workplace design

1 Introduction

Modular construction is a process where building modules are manufactured offsite and then transported to the construction sites for the final assembly [1]. Various terms are used in the literature to describe this manufacturing-oriented construction approach, including offsite construction [2], prefabricated construction [3], or industrialized construction [4]. To clarify, this study uses the term "modules" to encompass all prefabricated components, including 1D elements, 2D panelized systems, and 3D volumetric units [5]. Modular construction has been considered a viable alternative to traditional construction methods due to its improved productivity, quality, safety, and sustainability [6]. With its controlled manufacturing environment, robots can play a significant role in modular construction manufacturing (MCM) processes by alleviating workers' fatigue, improving quality, and boosting productivity [7, 8].

Despite the increasing use of automation in producing standardized building modules, significant challenges persist in MCM. First, MCM must accommodate diverse design styles to meet varying demands and adapt to economic fluctuations, resulting in considerable variability in manufacturing workflows [9]. This customization presents a challenge for standard automation, as conventional robotic systems lack the flexibility needed to handle frequent production changes. Second, fully automating MCM processes requires substantial upfront investments, such as the cost of equipment and factory reconfiguration [10]. This can be a barrier for small and medium-sized manufacturers with limited tolerance for financial risks [11]. Third, many MCM tasks require human involvement for the dexterous use of construction tools and the improvisational skills to adapt to material variations [12], while robots struggle to cope with these complexities.

To address these challenges, combining human expertise with robotic systems in MCM processes presents a promising solution through efficient human-robot collaboration (HRC) [13]. HRC capitalizes on the complementary strengths of humans and robots: humans excel at flexibility and adaptability for creative and ill-defined tasks, while robots offer unparalleled precision and consistency. Consequently, HRC offers significant potential for achieving the trade-off between MCM productivity and flexibility.

HRC workplace designs require careful consideration of various factors. Existing studies on HRC for MCM often prioritize productivity, efficiency, and flexibility by developing innovative workflows. For example, Tavares, Costa [14] presented a collaborative welding system where both a robot and a human share part of the physical welding task. Lee and Han [15] designed a novel robot system for welding, with an interactive interface for human fine-tuning. However, these studies paid limited attention to ergonomics and human factors, which may adversely affect workers' occupational safety and health

(OSH) and well-being [16]. With the advance of Industry 5.0, there is an increasing focus on human-centric HRC, where human well-being is prioritized in production processes [17]. Safety and ergonomics are central to this human-centric paradigm, yet current studies often fail to integrate ergonomic factors into HRC designs for MCM [18]. Besides, little research has been conducted in exploring HRC design considerations for MCM.

To bridge the gap, this paper aims to explore different HRC schemes in MCM and discuss critical considerations in designing ergonomic HRC (EHRC) workplaces. This study adopts an exploratory research approach, combining a literature review with case study analysis. First, we synthesize existing research to define EHRC and identify its design considerations in MCM. Second, we illustrate these considerations through three case studies derived from prior work, representing diverse MCM tasks, human roles, and robot types. By shifting the focus from productivity to human well-being, this study contributes to the development of human-centric and ergonomically optimized HRC applications in MCM.

2 Preliminary concepts for EHRC

HRC is gaining prominence to achieve the trade-off between flexibility and efficiency in production [19]. Its main goal is to reduce the physical and repetitive workload on human workers by enabling robots to assist in labor-intensive tasks. In the manufacturing industry, HRC typically involves humans and robots collaborating in a shared workplace towards a common goal. While similarities between MCM and traditional product manufacturing have been indicated [20], the make-to-order and elaborate nature of MCM presents unique challenges for implementing industrial HRC solutions effectively. Therefore, this paper adopts a broad definition of "collaboration" emphasizing joint efforts between individuals, especially in intellectual or creative pursuits [21].

To address worker well-being in such systems, ergonomics offers a critical foundation. The International Ergonomics Association (IEA) defines ergonomics as, "*the scientific discipline concerned with the understanding of interactions among humans and other elements of a system, and the profession that applies theory, principles, data, and methods to design in order to optimize human well-being and overall system performance.*" [22]. Although the terms "ergonomics" and "human factors" are sometimes distinguished - ergonomics being linked to physical aspects and human factors to psychological and cognitive aspects - they are often used interchangeably [22-24]. For clarity, this study adopts the IEA definition of ergonomics, encompassing both the physical and cognitive/psychological dimensions of human interaction. Furthermore, this study narrows its focus to the individual worker level, aligning with the scope of micro-ergonomics, which addresses both physical and cognitive factors [25] that directly impact worker well-being. Building on this, the study defines ergonomic human-robot collaboration (EHRC) as "using ergonomics principles to design a workplace where humans and robots collaborate safely, efficiently, and in support of worker well-being".

The EHRC further draws on the industrial human needs pyramid [17], which consists of five levels: (1) safety, the foundational level, which ensures physical protection and legal rights; (2) health, the next level, which focuses on creating a healthy work environment to support physical, cognitive, and psychological well-being; (3) belonging, which meets social needs of workers and to promote connections and trust between human-machine team; (4) esteem, which enhances workers' confidence, self-belief, and respect from peers; and (5) self-actualization, the highest level, which enables personal growth, fulfillment, and sense of purpose through meaningful work. This pyramid highlights that human health is founded on physical safety while facilitating higher levels of well-being. Mapping to the HRC domain, safe HRC primarily fulfills the safety level by mitigating physical hazards, while EHRC extends this to the health level by addressing physical, cognitive, and psychological factors to promote long-term well-being. Figure 1 illustrates this relationship, highlighting how EHRC builds upon safe HRC within the industrial human needs pyramid to enhance workplace design.



Figure 1. Role of EHRC in the industrial human needs pyramid (adapted from [17])

3 Design considerations of EHRC for MCM

3.1 HRC variables in MCM

Different HRC variables involve different requirements. Workers collaborating with robots on MCM tasks face specific ergonomic challenges that differ from the standardized and lightweight tasks in conventional manufacturing. This paper summarizes three critical HRC variables in MCM, i.e., MCM tasks, human roles, and robot types.

MCM tasks can be generally summarized into three types, i.e., small-size component production, 1D and 2D prefabricated component production, and 3D volumetric module production:

- Small-size component production: This involves manufacturing smaller components (e.g., windows and plumbing elements) that are later assembled on larger modules or buildings. In MCM, tasks to produce small components are less repetitive since size and specification needs vary in different projects. Intermittent and varied tasks require workers to make frequent postural or tool adjustments, increasing physical strain and cognitive burden.
- 1D and 2D prefabricated component production: This involves manufacturing structural elements such as steel beams (1D) and timber panels (2D) that bear the loads and stresses of the building. These components can be produced in assembly lines or fixed-position setups. Assembly-line workers face shoulder and back strain from repetitive lifting (e.g., aligning beams for robotic welding), while fixed-position tasks require prolonged leaning or twisting, increasing musculoskeletal risks.
- 3D volumetric module production: This refers to the manufacture of 3D volumetric housing prefabricated units. These volumetric modules can be prefabricated, pre-finished, and even pre-furnished, representing different levels of completion in modular construction. The large size and weight of these modules often confine workers and robots to fixed-position workstations. When working on the interior of a module, workers must collaborate with robots in tight spaces, which increases safety hazards while potentially further restricting workers' movement and forcing them to maintain awkward postures. As for exterior work such as cladding or window sealing, workers have more room to maneuver but face the challenges of working from heights or extending their arms.

Human roles can be categorized into four types: supervisor, teammate, operator, and bystander [13]:

- Supervisor: Supervisors are responsible for monitoring the robot and instructing the robot to perform the task. This role requires sustained attention to one or more robots performing diverse tasks in a cluttered MCM environment. As a result, this role requires constant vigilance, with increased cognitive load and mental fatigue.
- Teammate: Teammates are responsible for completing physical activities beyond the capabilities of the robot to achieve a common goal. As teammates are required to share physical tasks, they cannot avoid the postural risks associated with lifting, pushing, or adjusting materials. Collaborating in close proximity to the robot also increases the mental workload of the worker.
- Operator: Operators take direct control of the actions of the robot during the execution of the task. The control can be either in close proximity (e.g., guiding a gantry robot installing a window) or remotely (e.g., guiding a robot with VR controllers). Proximity control requires repetitive body postures, while telecontrol increases eyestrain and cognitive load.
- Bystander: Bystanders work in the same space as robots without direct interaction, but should be concerned with issues related to safety and situation awareness. Bystanders must be distracted from the robot's movements while focusing on the MCM task at hand, which causes additional mental stress and cognitive burden.

Robots commonly adopted in MCM can be classified into three categories, i.e., stationary, mobile, and wearable robots [26]:

- Stationary robot: Stationary robots are fixed in the workplace, have a limited working range, and cannot move freely. The most common stationary robot is robot arms that are bolted in the workstation. Due to their large payloads and high power, safety protocols prioritize isolation over proximity HRC. This physical isolation shifts the ergonomic challenge to remote monitoring.
- Mobile robot: Mobile robots provide enough flexibility to create an adaptive workplace. Automated guided vehicles (AGVs) are commonly used to transport materials around the factories. Robot arms (e.g., collaborative robots) mounted on AGVs or gantries are also considered mobile as they can reconfigure their location freely in the factory. Since these robots offer flexibility to work around workers, they are either with lower payload and power, or are less automated and require workers' direct control to ensure safety. Collaboration with these robots can be physically

demanding and cognitively taxing.

- Wearable robot: Wearable robots, commonly known as exoskeletons, are designed to boost human physical strength and endurance. These robots not only enable individuals to handle heavier loads but also help improve body posture during tasks, reducing the risk of strain and injury. Due to the complexity of controls and fewer applications, they can be an added cognitive burden for workers.

3.2 Design considerations for EHRC

Designing EHRC workplaces in MCM requires careful attention to safety strategies, physical ergonomic design, and cognitive ergonomic design.

3.2.1 Safety strategies

Safety strategies in HRC include a blender of separation protection, passive protection, and proactive protection. These strategies should be carefully selected in the EHRC design.

- Separation protection: Separation protection ensures safety by keeping humans and robots physically separated. This protection method is typically achieved through safeguards such as guards (e.g., fences, barriers, and light curtains). When collaborating with heavy industrial robots, maintaining a strict separation between humans and robots is essential.
- Passive protection: Passive protection ensures safety with robots' built-in design features without requiring active monitoring and prediction of human state. Collaborative robots typically have passive protection mechanisms such as limited payload and end-effector state detection, since they are designed to allow closed collaboration in a shared workplace.
- Proactive protection: Proactive protection perceives and predicts the human state and adapts robot behavior in real time to avoid safety risks. This method integrates advanced sensors such as depth cameras and wearable devices to monitor humans and adopt actions before an unsafe event occurs.

3.2.2 Ergonomic design considerations

In addition to keeping workers away from safety hazards, EHRC prioritizes worker health and well-being through ergonomic design, consisting of both physical and cognitive aspects [17]. Physical ergonomics examines human postures and behaviors that could cause musculoskeletal injuries [27], assessed through tools such as rapid upper limb assessment (RULA) [28] and rapid entire body assessment (REBA) [29], which evaluate posture risks, and biomechanics analysis (e.g., OpenSim), which measures muscle tension and joint loads. Cognitive ergonomics focuses on reducing the mental stress and psychological challenges faced by workers sharing the workplace with robots [30], evaluated using self-report scales such as NASA task load index (NASA-TLX) [31], supplemented by physiological data (e.g., electrodermal activity, heart activity, electroencephalography, electromyogram, eye activity) from wearable sensors. These physical and cognitive ergonomic principles are integrated into three design considerations tailored to MCM:

- Ergonomic task design: This balances task demands to reduce physical and cognitive strain. By optimizing task sequences, repetitive actions are delegated to robots, while workers engage in varied movements to lower musculoskeletal injury risks in assembly-line or fixed-position setups. Dynamic task adaptation allows robots to sense human states (e.g., physical fatigue and cognitive overload) in order to react by taking over demanding tasks or adjusting the speed of work, ensuring the well-being of workers in a custom-driven workflow. Real-time task feedback can help workers make informed decisions with a high level of trust and low cognitive workload.
- Ergonomic workstation design: This optimizes the physical workspace to minimize strain and environmental stress. Adjustable workstations allow workers to maintain natural postures minimizing the need for excessive bending or overextension during tasks. Additionally, mitigating adverse environmental factors such as noise and vibration generated by robots can preserve worker well-being and minimize ergonomic stress.
- Ergonomic interaction design: This enhances the usability and cognitive comfort of human-robot interaction. Intuitive interfaces allow workers to interact seamlessly with the robot, improving usability while reducing frustration and cognitive effort. Predictable robot motion can enhance worker trust and situational awareness and reduce stress when interacting with robots. Personalized interface settings cater to a diverse workforce in MCM, avoiding information overload for experienced operators while ensuring sufficient guidance for novices.

Although EHRC is a human-centric paradigm prioritizing human well-being, the system's efficiency cannot be ignored. However, the efficiency is not deeply examined in the paper, but its impact is presented in the discussion.

4 Case study

Three case studies under MCM scenarios are presented based on different EHRC requirements.

4.1 Case 1: Window installation on 3D modules

The first case is from our previous work [18], where a material-handling robot is applied to install a window component in an opening of a building module, as shown in Figure 2(a). Since 3D volumetric module production typically involves fixed-position workplace layouts that require workers and robots to work around the module, robot mobility is critical, which can be achieved with gantry cranes specialized for factory environments. The case has developed a 5-degree-of-freedom (DOF) gantry robot with 3 translational and 2 rotational servo-controlled axes. It is used to handle heavy window glasses; thus, the end-effector consists of a gripper with four vacuum cups. The worker is required to physically operate the robot to secure the window and haul the robot to its final location. Specifically, the worker first operates the two-handed handle with sensors, by moving it downward until it makes contact with the window glass. Once in position, the worker activates the gripping mechanism to secure the glass, enabling easy manipulation. Subsequently, the forward rotation button is engaged to orient the window to an upright position. The worker then guides the robot, along with the secured glass, to align it with the designated opening in the building module. After carefully adjusting the window to fit the opening, the worker releases the glass by pressing the release button. Finally, the robot is guided back to its initial position, and the gripper's rotation is reset, thereby completing one task cycle.

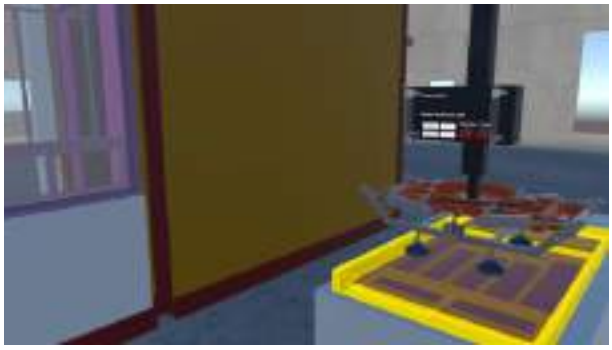


Figure 2. VR-based iterative EHRC design in case 1 [18]

The design of this collaborative robot allows humans to operate in close proximity, and the method of ensuring safety relies on its built-in mechanism design. The end-effector releases the part only after all the weight has been removed. If any problem occurs, the emergency stop button can be pressed.

The design of the HRC workstation meets ergonomics requirements from both physical and cognitive aspects. To improve physical ergonomics, the height of the worktable and the width between the two handles on the collaborative robot have been carefully designed, and both RULA and REBA tools show low risk. In terms of cognitive ergonomics, the cognitive workload and trust in the robot can be kept at an acceptable level due to the simplicity of the operation process and the intuitive human-robot interface. It is worth noting that the ergonomic assessment of this case was conducted in a virtual reality (VR) environment [Figure 2(b)], which offers a cost-efficient and iterative platform for designers to evaluate and refine ergonomic parameters with minimal resource expenditure.

4.2 Case 2: 2D timber panel production

The case 2 [11] adopts the TIM platform [32], which incorporates two KUKA KR-500 industrial robots and a swiveling table. Figure 3(a) shows the configuration of case 2. The industrial robots utilized have a scope of 2830 mm, including the end effector, a payload capacity of 500 kg, excluding the end effector, and maintain a pose repeatability accuracy of +- 0.08 mm. The robots are equipped with different end effectors each, namely a mechanical gripper, pneumatic gripper, and 3D-Camera for the PINP robot, and a nail gun, spindle and glue gun for the GNM robot, thus defining their respective skills. In this case study, the PINP robot uses the mechanical gripper for picking up the slat, placing and securing it, whereas the GNM robot employs the nail gun to fasten it. The swiveling table facilitates an extended reach for the robots by repositioning the workpieces.

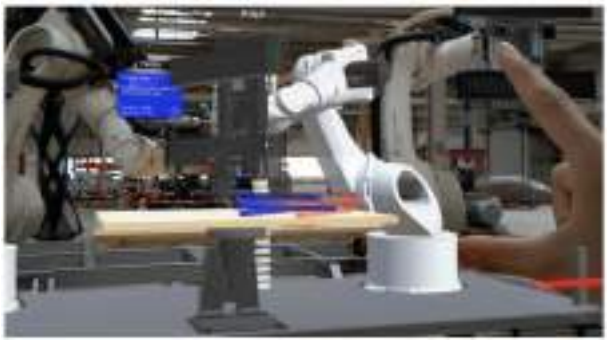


Figure 3. AR-based EHRC in case 2 [11]

This case adopts industrial robots to handle large-scale materials, which are too dangerous for physical collaboration. Therefore, separation protection is used to ensure human safety. By getting away from robots, human supervisors can avoid hazards from potential

crashes. When humans need to enter the range of the high-payload industrial robots to conduct tasks such as material supply, confirming that the robots are inactive is crucial to prevent accidents.

As supervisors, humans are required high-level cognitive capacities. This case provides an intuitive and smooth augmented reality (AR) interface for humans to improve their cognitive ergonomics, as shown in Figure 3(b). Although improper use of HRC systems could affect physical ergonomics, the motion introduced by AR (e.g., shoulder uplift motion associated with gesture recognition used in human-robot interaction) is neglectable from a biomechanical loading perspective.

4.3 Case 3: Window frame processing

The last case is from [33], which adopts a collaborative robot, UR5e robot arm, for holding the workpiece (i.e., a window frame), and a human conducts processing tasks on the workpiece. Specifically, the human is required to hold a power drill to punch four holes in each corner of the window frame, while the robot is responsible for adapting the workpiece's position and orientation to provide an ergonomic work condition. The configuration of case 3 is shown in Figure 4.



Figure 4. Proactive EHRC in case 3 [33]

This EHRC workstation ensures safety through proactive protection. On the one hand, the built-in safety mechanisms of the UR5e are designed in compliance with EN ISO 13849-1 and EN ISO 10218-1. On the other hand, the RealSense camera is used to perceive the safety condition of humans so as to prevent hazards proactively.

The robot can improve physical ergonomics by monitoring the 3D skeleton of the human, evaluating posture-related risks based on RULA tools, and adapting its motion accordingly to reduce the risks. In addition, the human wears AR glasses, which provide task-related information and hands-on instruction. The worker awareness and trust in robots can also improve through robot digital twins.

4.4 Discussion

Table 1 summarizes the three EHRC cases under study. These MCM cases illustrate differences in MCM tasks, human roles and robot types. Additionally, they involve different design considerations, offering insights for EHRC design in MCM:

- For 3D volumetric module production (Case 1), the fixed-position workplace layout confines workers and robots to work around the module, exposing workers to both physical and cognitive ergonomic risks. The case assigns heavy lifting to the robot, while the worker is only responsible for precise positioning, thus greatly reducing the burden on muscles. To design an ergonomic workstation, this case carefully designs a mobile gantry robot with adjustable gripper height, reducing awkward postures during window installation. The intuitive interface supports ergonomic interaction by simplifying operation, minimizing cognitive effort in this task. By adopting an ergonomic design, this case remains ergonomic risks to an acceptable level both physically (i.e., RULA and REBA) and cognitively (i.e., cognitive workload and trust).
- In 2D timber panel production (Case 2), heavy-duty industrial robots require separation protection, shifting human roles to supervisors focused on monitoring. The case assigns almost all of the physical work to the robot, with the worker needing only supervision and guidance. Ergonomic workstation design plays a minimal role here due to physical isolation. The AR interface enhances ergonomic interaction by providing seamless and intuitive task oversight, improving comprehension and reducing mental fatigue. This case does not directly demonstrate ergonomic design improvements for workers from physical and cognitive perspectives, but the intuitiveness, task support, and skill enhancement provided by AR technology are highly aligned with cognitive ergonomics principles.
- For small-size component tasks like window frame processing (Case 3), this case leaves the dexterity task to the worker, while the robot is responsible for helping to adjust the position of the workpiece, thus allowing the worker to reduce postural risks. The workstation senses human states (e.g., improper posture) in real time through sensors and adjusts the robot's pose to provide ergonomic working conditions that ensure worker health. An AR-driven interface enhances ergonomic interactions by providing ergonomic warnings, task guidance, and robot trajectory prediction, thereby improving situational awareness and cognitive capacity. The ergonomic design of this case keeps the RULA risk scores of the workers' arms and wrists at a low-risk level and improves their cognitive understanding of the HRC task and their trust in the robot.

Table 1. Summary of EHRC cases under study

Variables	Case 1	Case 2	Case 3
Task	Window installation on 3D modules	2D timber panel production	Window frame processing
Robot type	Mobile robot (Gantry robot)	Stationary robot (Industrial robots)	Mobile robot (collaborative robots)
Human role	Operator	Supervisor	Teammate
Safety strategy	Passive	Separation	Proactive
Ergonomic task design	Aligning robots for heavy lifting	Aligning robots for primary work	Aligning robots for workpiece adjustment
Ergonomic workstation design	Adjustable design	-	Adjustable design
Ergonomic interaction design	Intuitive controls	Intuitive AR, predictable robot motion	Intuitive AR, predictable robot motion

5 Conclusion

This paper establishes a rigorous definition of EHRC, emphasizing the design of workplaces where humans and robots work together safely, comfortably, and efficiently. It further delineates design considerations under MCM scenarios, considering HRC variables and specific challenges in MCM. Through the analysis of three case studies, this paper elucidates practical insights into the development of EHRC systems tailored to MCM. Theoretically, this study advances this field by identifying current HRC ergonomic challenges in MCM and proposing tailored design solutions that prioritize human-centric systems. It also provides a roadmap for practitioners to enhance worker safety and comfort, while mitigating expenditures on workers' compensation claims.

This paper serves as an initial blueprint for EHRC. Three future directions are proposed:

- *Multidisciplinary integration in EHRC design*. The complexity of EHRC requires a multidisciplinary approach, bridging ergonomics with robotics, medicine, psychology, human resources, and end-user perspectives. To achieve this, interoperability between these diverse fields is crucial [34]. A collaborative team of experts from these disciplines must work together to ensure that their insights and solutions are not only comprehensive but also seamlessly integrated.
- *Customized interaction interfaces.* Customized human-robot interaction interfaces are essential for achieving effective EHRC. Future efforts should focus on developing diverse interfaces for users to choose from, complemented by empirical user studies. Such studies are critical to assessing the impact of varying interface designs on different workers, optimizing information load, presentation modalities, and input mechanisms for enhanced usability and performance.
- *Dynamic adaptation in real-time systems*. Static and offline methodologies fail to account for the variability in human actions and states during real-time operations. Future systems must incorporate dynamic task allocation within human-robot teams [35] and adaptive robotic responses contingent on real-time human states. This requires perceiving and responding to human conditions, such as improper postures or elevated cognitive loads, ensuring safe and ergonomic collaboration.

References

[1] Bertram, N., et al. *Modular construction: from projects to products*. McKinsey & Company: Capital Projects & Infrastructure. On-line: https://www.mckinsey.com/capabilities/operations/our-insights/modular-construction-from-projects-to-products, Accessed: December 10, 2024

[2] Lou, J., et al., *Nomenclature for offsite construction.* Building Research & Information, 2022. 50(8): p. 894-908.

[3] Li, Z., G.Q. Shen, and X. Xue, *Critical review of the research on the management of prefabricated construction.* Habitat international, 2014. 43: p. 240-249.

[4] Qi, B., et al., *A systematic review of emerging technologies in industrialized construction.* Journal of Building Engineering, 2021. 39: p. 102265.

[5] Lim, Y.-W., et al., *Planning and coordination of modular construction.* Automation in Construction, 2022. 141: p. 104455.

[6] Razkenari, M., et al., *Perceptions of offsite construction in the United States: An investigation of current practices.* Journal of Building Engineering, 2020. 29: p. 101138.

[7] Parascho, S., *Construction robotics: from automation to collaboration.* Annual Review of Control, Robotics, and Autonomous Systems, 2023. 6(1): p. 183-204.

[8] Bock, T., *Construction robotics.* Autonomous Robots, 2007. 22(3): p. 201-209.

[9] Goulding, J.S., et al., *New offsite production and business models in construction: priorities for the future research agenda*. Architectural Engineering

and Design Management, 2015. 11(3): p. 163-184.
[10] Yang, Y. and W. Pan, *Automated guided vehicles in modular integrated construction: potentials and future directions.* Construction Innovation, 2021. 21(1): p. 85-104.
[11] Amtsberg, F., et al. *iHRC: an AR-based interface for intuitive, interactive and coordinated task sharing between humans and robots in building construction*. in *Proceedings of the 38th International Symposium on Automation and Robotics in Construction (ISARC)*. 2021. International Association for Automation and Robotics in Construction (IAARC).
[12] Yu, H., et al., *Mutual physical state-aware object handover in full-contact collaborative human-robot construction work.* Automation in Construction, 2023. 150: p. 104829.
[13] Fu, Y., J. Chen, and W. Lu, *Human-robot collaboration for modular construction manufacturing: Review of academic research.* Automation in Construction, 2024. 158: p. 105196.
[14] Tavares, P., et al., *Collaborative welding system using BIM for robotic reprogramming and spatial augmented reality.* Automation in Construction, 2019. 106: p. 102825.
[15] Lee, D. and K. Han, *Vision-based construction robot for real-time automated welding with human-robot interaction.* Automation in Construction, 2024. 168: p. 105782.
[16] Chen, J., et al., *Augmented reality-enabled human-robot collaboration to balance construction waste sorting efficiency and occupational safety and health.* Journal of Environmental Management, 2023. 348: p. 119341.
[17] Lu, Y., et al., *Outlook on human-centric manufacturing towards Industry 5.0.* Journal of Manufacturing Systems, 2022. 62: p. 612-627.
[18] Fu, Y., W. Lu, and J. Chen, *A virtual reality-based ergonomic assessment approach for human-robot collaboration workstation design in modular construction manufacturing.* Advanced Engineering Informatics, 2025. 64: p. 103054.
[19] Mukherjee, D., et al., *A survey of robot learning strategies for human-robot collaboration in industrial settings.* Robotics and Computer-Integrated Manufacturing, 2022. 73: p. 102231.
[20] Lu, W., Z. Yang, and L. Kong, *Identification of learning effects in modular construction manufacturing.* Automation in Construction, 2023. 154: p. 105010.
[21] Green, S.A., et al., *Human-robot collaboration: a literature review and augmented reality approach in design.* International Journal of Advanced Robotic Systems, 2008. 5(1): p. 1.
[22] IEA. *What Is Ergonomics (HFE)?* On-line: https://iea.cc/about/what-is-ergonomics/, Accessed: Novermber 21 2024
[23] Salvendy, G. and W. Karwowski, *Handbook of human factors and ergonomics*. 2021: John Wiley & Sons.
[24] Wilson, J.R., *Fundamentals of systems ergonomics/human factors.* Applied Ergonomics, 2014. 45(1): p. 5-13.
[25] Lorenzini, M., et al., *Ergonomic human-robot collaboration in industry: A review.* Frontiers in Robotics and AI, 2023. 9.
[26] Rodrigues, P.B., et al., *A multidimensional taxonomy for human-robot interaction in construction.* Automation in Construction, 2023. 150: p. 104845.
[27] Benos, L., A. Bechar, and D. Bochtis, *Safety and ergonomics in human-robot interactive agricultural operations.* Biosystems Engineering, 2020. 200: p. 55-72.
[28] McAtamney, L. and E. Nigel Corlett, *RULA: a survey method for the investigation of work-related upper limb disorders.* Applied Ergonomics, 1993. 24(2): p. 91-99.
[29] Hignett, S. and L. McAtamney, *Rapid Entire Body Assessment (REBA).* Applied Ergonomics, 2000. 31(2): p. 201-205.
[30] Gualtieri, L., E. Rauch, and R. Vidoni, *Emerging research fields in safety and ergonomics in industrial collaborative robotics: A systematic literature review.* Robotics and Computer-Integrated Manufacturing, 2021. 67: p. 101998.
[31] Hart, S.G. and L.E. Staveland, *Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research*, in *Advances in Psychology*, P.A. Hancock and N. Meshkati, Editors. 1988, North-Holland. p. 139-183.
[32] Wagner, H.J., et al., *Flexible and transportable robotic timber construction platform–TIM.* Automation in Construction, 2020. 120: p. 103400.
[33] Fu, Y. and W. Lu, *A mutual-cognitive human-robot collaboration framework to improve ergonomics in industrialized construction*, in *Computing in Civil Engineering 2024 (Accepted)*. 2025.
[34] Chen, J., et al., *Building "RoboAvatar": Industry Foundation Classes–Based Digital Representation of Robots in the Built Environment.* Journal of Computing in Civil Engineering, 2024. 38(4): p. 04024013.
[35] Lu, W., et al., *Digital twin-enabled human-robot collaborative teaming towards sustainable and healthy built environments.* Journal of Cleaner Production, 2023. 412: p. 137412.

Unsupervised Multimodal Learning for Fault Detection of Fan Coils Units Using Building Automation System Data

Chin Choy Chai, and J. J. McArthur*

Department of Architectural Science, Toronto Metropolitan University, Canada.
chaicc@torontomu.ca; jennifer.mcarthur@torontomu.ca**corresponding author*

Abstract

An unsupervised multimodal learning scheme based on autoencoder (AE) is proposed and developed for fault detection of building heating, ventilation, and air conditioning (HVAC) systems, applicable across multiple types of fan coil units (FCUs), i.e., cooling, heating, cooling & heating FCUs. Building automation system data of these three types of FCUs are used in developing, training and evaluating the proposed models and algorithms. The main contributions of this work include: (i) a methodology for developing a generalized AE model through multimodal and transfer learning; (ii) generalizing the AE model across three specific types of FCUs through cross-testing and transfer learning, which enables fault detection at the device-level (a specific FCU type), and system-level (across different FCU types).

Keywords–Unsupervised learning; multimodal learning; autoencoder; fan coil units; fault detection, building HVAC

1 Introduction

According to Intergovernmental Panel on Climate Change (IPCC) report [1], buildings account for 37% of global carbon emissions and 34% of energy demand worldwide. In particular, a building's heating, ventilation, and air conditioning (HVAC) system accounts for half of its total energy use, making them a primary focus for energy optimization [2]. Due to existing design flaws, operational variances, and system faults [3] [4], any undetected faults can cause an additional 103 to 500 TWh of energy consumption annually in the U.S. [5]. To mitigate these issues, fault detection and diagnosis (FDD) has been employed as a cost-effective solution for identifying potential system failure or deficiencies, and addressing these faults [6].

A fan coil unit (FCU) is the "last mile" in many building HVAC systems, providing zone-level heating and cooling. FCUs typically consist of a heating and/or cooling coil, fan, and filter, and are controlled with local thermostats to achieve desired space conditions. Large buildings may contain dozens – even hundreds – of FCUs installed in a building, thus ineffective or faulty FCUs operation can result in occupant discomfort and energy waste. Limited research has explored FDD for FCUs and the majority rely on simulation-based methods using simulated data [7, 8], or unsupervised [9] [10] methods. Further, existing FDD solutions are mostly developed specifically for particular, limited fault categories (e.g. valve stuck, fan failure) [10] or a particular FCU type. There is still a lack of generalized solutions for characterizing FCU operation and FDD for the FCUs, in lieu of designing different FDD solutions for different types of FCUs for cooling or heating. Even if machine learning or neural network-based solutions were to be implemented, they require retraining for at least three different types of FCUs (one for each type), which incurs significant additional resources.

In this paper, we propose an unsupervised multimodal learning approach for developing and designing novel algorithms for fault detection of the FCUs. A methodology for developing a generalized autoencoder (AE) model through multimodal and transfer learning is first presented. This methodology facilitates unsupervised learning for fault detection at the equipment-level (for a specific FCU type), and system-level (across different FCU types). Then, a global model and specific models are developed for fault detection of multiple FCU types. Fault detection performance of the global model and specific models are evaluated to demonstrate the generalizability of the AE across different FCU types. These are achieved through cross-testing and transfer learning based on real building automation system data. In a nutshell, based on unsupervised multimodal learning, only a single global AE model is required for fault detection through unsupervised learning from historical data of cooling-only, heating-only, cooling & heating FCUs.

2 Background

Figure 1 shows a schematic of an FCU, consisting of a heating coil (H/C), a cooling coil (C/C), and a fan. It is

part of an HVAC system and is widely used in multi-unit residential, commercial, and industrial buildings.

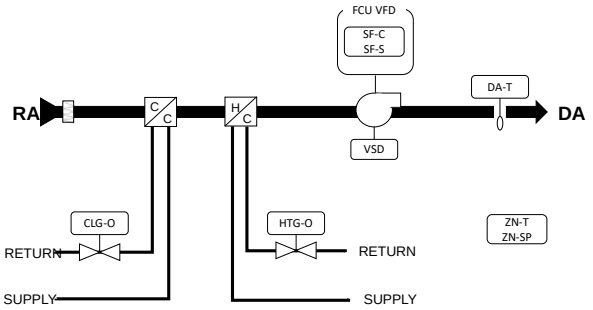


Figure 1: Schematic diagram of a FCU; VSD denotes variable speed drive.

Each component of a FCU and their functionalities are described in Table 1.

Table 1. FCU components and functionalities.

Component	Description
Cooling coil command (CLG-O)	The cooling coil is controlled by a valve controlled by a command CLG-O (of continuous value between 0 and 1), which indicates the percentage the valve is open. The valve is for controlling the flow of chilled water to regulate the coil's cooling capacity as needed.
Heating coil command (HTG-O)	The heating coil is controlled by a valve controlled by a command HTG-O (of continuous value between 0 and 1), which indicates the percentage the valve is open for supplying hot water from the building's central heating plant. When heating is needed, this coil introduces warmth to the air drawing into it, effectively raising the temperature. The flow of hot water through the heating coil is regulated by a valve for controlling the amount of heat to the air.
Supply air (DA)	Conditioned air to the space. The temperature of the supply air is indicated by the DA-T sensor.
Zone sensors (ZN-T, ZN-SP)	Sensors placed within the conditioned zone. ZN-T is the zone temperature; ZN-SP is the set-point of the temperature.
Fan sensors (SF-C, SF-S)	The fan draws air through the coils and then discharges it into the room. Supply Fan Command (SF-C) is a control output which commands the operation of the supply fan. Supply Fan Status (SF-S) indicates the supply fan's on/off status.
Occupancy schedule (OCC-SCHEDULE)	The system will operate based on a time of day schedule adjustable by the user.
Unit enable signal (UNITEN-MODE)	A network unit enable signal (UNITEN-MODE) will control the mode of the unit.

In this work, real building automation system datasets of cooling-only, heating-only, and cooling & heating FCUs in a building HVAC system are adopted. The data information covers the room temperature (ZN-T), supply air temperature (DA-T), as well as command data such as cooling valve output (CLG-O), heating valve output (HTG-O), and temperature set point (ZN-SP), etc., as summarized in Table 1.

2.1 Principle of an Autoencoder

AEs are a type of neural network model that enable unsupervised features extraction [11], due to their strong representation learning capabilities, simple structure, and ease of training. As illustrated in Figure 2, the architecture is formed by an encoder and a decoder.

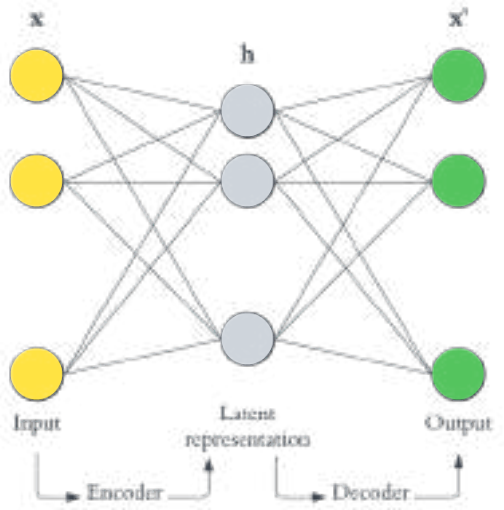


Figure 2: Neural network architecture of an AE.

The mathematical formulation of the AE architecture is as follows:

$$\boldsymbol{h} = f(\boldsymbol{W}_e\boldsymbol{x} + \boldsymbol{b}_e) \quad (1)$$

$$\boldsymbol{x}' = g(\boldsymbol{W}_d\boldsymbol{h} + \boldsymbol{b}_d) \quad (2)$$

where $\boldsymbol{x}$ and $\boldsymbol{x}'$ denote the input and reconstructed output, respectively, $\boldsymbol{h}$ denotes the latent space; $\boldsymbol{W}_e$ and $\boldsymbol{b}_e$ denotes the encoder's weights and bias respectively, while $\boldsymbol{W}_d$ and $\boldsymbol{b}_d$ denote the decoder's weights and bias respectively; $f()$ and $g()$ denote each of their respective activation functions.

In an AE, the encoder section first compresses high-dimensional input data into a compact latent representation. This compact latent representation is then used by its decoder section to reconstruct the original data.

During training phase, the AE's objective is to minimize the error between the original input data and

its reconstructed data. Given a training data set of size N, the neural network weights of an AE are optimized by minimizing a loss function $\mathcal{L}$ using gradient descent algorithm. The loss function $\mathcal{L}$ quantifies the reconstruction error as,

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}\left|x^{(i)} - x'^{(i)}\right|^2 \quad (3)$$

After successful training, the latent space $\boldsymbol{h}$ can be adopted as the feature representation for any given input $\boldsymbol{x}$.

This training process facilitates the neural network to learn and capture essential relationships among input variables, while filtering out any anomalies and noise, enabling the latent space to characterize any hidden patterns in the original input data. In general, the above encoder-decoder framework also facilitates integration of other deep learning networks like recurrent neural networks and convolutional neural networks for more complex representation learning [12].

A larger latent space that is close to the input size could lead to overfitting, as the AE model merely copies the input data without extracting essential patterns or features from the original input data. Conversely, a latent space that is too small, such as one dimension could cause underfitting and poor reconstruction of input data, as the AE model fails to extract any features or features from the original input data.

3 Multimodal AE for Fault Detection

The primary goal of this work is to design a generalized AE that captures the dynamics and interaction of unimodal (cooling-only, heating-only) and bimodal (heating & cooling) operations of three different FCU types, while also models each of them separately. Previous works in multimodal research has proposed feature concatenation and tensor fusion approaches for multimodal learning. In this work, the tensor fusion approach is adopted for multimodal modeling and learning of three types FCUs, by first defining the following vector field using a 2-fold Cartesian product:

$$\{(\boldsymbol{z}_c, \boldsymbol{z}_h) | \boldsymbol{z}_c \in \begin{bmatrix} CLG_0 \\ 1 \end{bmatrix}, \boldsymbol{z}_h \in \begin{bmatrix} HTG_0 \\ 1 \end{bmatrix}\} \quad (4)$$

where the subscripts $_c$ and $_h$ denote, respectively, cooling and heating; CLG-O denotes the cooling valve output command to a FCU for cooling; HTG-O denotes the heating valve output command to a FCU for heating. Note that the extra constant dimension with value 1 is for generating the three distinct input feature variables that capture, respectively, the dynamics and interaction unimodal (cooling-only, heating-only) and bimodal (heating & cooling) operations. As a result, for the designed AE neural network, each coordinate $(\boldsymbol{z}_c; \boldsymbol{z}_h)$ can be seen as a 2-D point in the 2-fold Cartesian space defined by the embedding vector $\begin{bmatrix} CLG_0 \\ 1 \end{bmatrix}$ for cooling, and the embedding vector $\begin{bmatrix} HTG_0 \\ 1 \end{bmatrix}$ for heating. Mathematically, this is equivalent to an outer product between the representation for cooling, $\boldsymbol{z}_c$, and the representation for heating, $\boldsymbol{z}_h$,

$$\boldsymbol{z}_m = \begin{bmatrix} CLG_0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} HTG_0 \\ 1 \end{bmatrix} \quad (5)$$

where $\otimes$ denotes the outer product between two vectors.

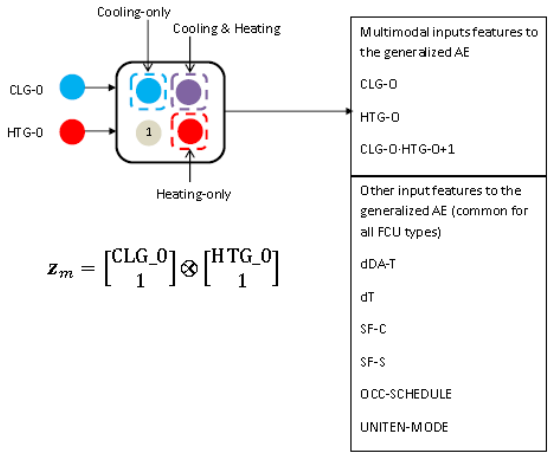


Figure 3: Proposed fusion approach with unimodal (cooling-only, heating-only) and bimodal (heating & cooling) features as additional inputs to the multimodal AE

As illustrated in Figure 3, the resulting matrix $\boldsymbol{z}_m$ can be viewed as a 2D surface that can be divided into 3 sub-regions, corresponding to three combinations of the data variables CLG-O, HTG-O and 1. In a nutshell, these three sub-regions characterize unimodal and bimodal dynamics, as well as interactions between three operational modes (cooling-only, heating-only, heating & cooling) of these three different FCU types.

3.1 Multimodal Learning for Fault Detection

Based on the design in Figure 3, the new input feature variable, G = CLG-O · HTG-O+1, is created, where CLG-O · HTG-O denotes the element-wise product of command points of cooling valve output command (CLG-O) and heating valve output command (HTG-O). Essentially, this variable is created for modeling the multimodal AE's behavior in learning and capturing the overall dynamics and interaction across the cooling-only, heating-only, heating & cooling FCUs as follows:

- For cooling-only FCUs, CLG-O ≠ 0 and HTG-O = 0; G = 1. HTG-O does not contribute to the error for fault detection.
- For heating-only FCUs, CLG-O = 0 and HTG-O ≠ 0; G = 1. CLG-O does not contribute to the error

for fault detection.

- For heating & cooling FCUs, CLG-O ≠ 0 and HTG-O ≠ 0; $G > 1$. Both CLG-O and HTG-O contribute to the error for fault detection.

The value of G also provides an indication to the multimodal AE that the FCU is operating for any of the cooling-only, heating-only, or heating & cooling purpose. In particular, the G enables the AE to capture the instances with hot water flowing through the heating coil, while chilled water flowing through the cooling coil (simultaneously), which is likely a fault event. On top of the above, all input variables are standardized using a min/max scaler to ensure uniform scaling across the dataset.

In addition, at the HVAC controllers, the temperature difference between the observed temperature and the set point is used for regulating the heating and cooling. Due to the temperature set-points are adjusted by different users and according to different cooling/heating requirements at different rooms in the building, there is a lack of consistency (rule) across different FCUs for set-point setting. It is challenging for a small neural network to capture this dynamic due to such inconsistency in setting set-points. To address this limitation, new feature variables are created [13]: (i) dT (room temperature ZN-T minus its set-point ZN-SP); (ii) dDA-T (discharge air temperature DA-T minus the room temperature set-point ZN-SP) and used as input to the AE models. These new variables emulate the controller's behavior, by taking account only the difference relative to the temperature set-point.

The other input feature variables to the AE, including CLG-O, HTG-O, SF-C, SF-S, etc., are points on the building automation system sensors, which are evaluated to identify unusual behavior, which could indicate an equipment fault. Based on the encoder and decoder outputs of each of the AE model for FCUs, the mean absolute error (MAE), which quantifies the error between each model's predictions and the actual observations, can be computed. The MAE is used for evaluation of both global model and specific models for FCUs, as to be explained in Section 3.3.

3.1.1 Effects of Faults on AE Input Variables

Potential fault examples of the FCUs are summarized in Figure 4. In Table 2, we present potential fault examples and their effects on the input variables to the AE. Note that these cover a subset of potential faults that can be inferred from existing sensors and not all possible fault combinations.

Note how these potential faults affect the input feature variables to the AE models. As an example, for fault CV1 (cooling valve stuck open), the valve command signal CLG-O could be at 0% (or any low value), while the room temperature (ZN-T) and supply air temperature (DA-T) are low due to full cooling conditions



Figure 4: Classification of FCU faults [13].

Table 2: Potential faults examples [13] and their effects on the input feature variables to the AE.

Fault codes	Condition	Effects of fault
CC1, CC2, CC3, CV2	CLG-O = 100%	High DA-T
HC2, HV2, HC1, HC3	HTG-O = 100%	Low ZN-T & DA-T
CV1	CLG-O = 0%	Low ZN-T & DA-T
HV1	HTG-O = 0%	High ZN-T & DA-T
SF2, SF4	CLG-O = 100%	High ZN-T

3.2 Development of Generalized AE Models

Two generalized AE models are developed and proposed: (i) Specific models (cooling-only, or heating-only, or heating & cooling); (ii) A global model that characterizes all three types of FCUs.

Proposed unsupervised multimodal AE for FCUs: The created feature G is adopted as a new input to the multimodal AE model. The proposed multimodal AE model that covers all cooling-only, heating-only, heating & cooling FCUs is illustrated in Figure 5. Following the abovementioned fusion design approach, the multimodal AE characterizes dynamics of these three FCU-types for fault detection.

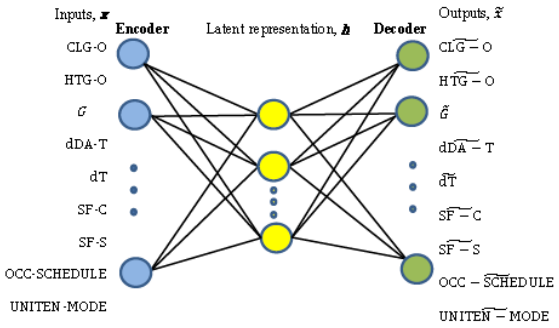


Figure 5: Proposed multimodal AE model

3.2.1 Global Model and Specific Models for FCUs

In this subsection, training steps for the multimodal AE to develop a global model and specific models for fault detection of FCUs are presented.

Specific models: The multimodal AE model in Figure 5 is trained to be ('specifically') applicable for fault detection for a specific FCU type (e.g., cooling-specific FCU). To that end, each specific model for a given FCU is trained with two-third of its own (or type) data. For example, the specific model of cooling-only FCU-C1 can be trained by its own historical data. By extension, we can also develop models for each type of FCUs, for e.g., cooling-only FCU, using data from multiple cooling-only FCUs. The specific model of any FCU is trained and used for fault detection of that particular FCU. Consequently, a specific model is expected to learn and characterize specific operational patterns which are limited to each type of FCU (as it is trained for), with a tradeoff for limited generalization capability.

Global model: The multimodal AE model in Figure 5 is trained to be (globally) applicable for fault detection across all three types of FCUs (cooling-only, heating-only, or heating & cooling). To that end, the (single) global model is trained with two-third of the data from a combination of all three types of FCUs. This global model is developed through a combination of training data from cooling-only FCUs (labeled as FCU-C1, FCU-C2, FCU-C3), heating-only FCUs (FCU-H1, FCU-H2, FCU-H3) as well as heating & cooling FCUs (FCU-HC1, FCU-HC2, FCU-HC3).

The single global model is trained for fault detection of three types of FCUs. Consequently, the global model is expected to learn and characterize more general operational patterns across all three types of FCUs due to its generalization capability, with a tradeoff for less accurate prediction of finer details.

As an example, consider a cooling-only FCU conditioning a server room. Due to constant heat emission from servers, this FCU is likely to operate according to more extreme cooling requirements. The specific (AE) model that is trained (using its cooling-only FCU data) is better at detecting anomalies during the FCU's normal or usual operational setting, but risks false alarms in new operational setting, such as during server maintenance shutdowns.

Alternatively, the above specific model can be substituted with the global AE model that has been trained with data from all three types of FCUs. Due to the global (AE) model's generalization capability, we expect that the global model's is less likely to give false alarms in new operational setting, while it may miss certain finer pattern shifts in the data information.

3.3 Quantifying Reconstruction Error

After learning a latent representation, the AE can be used to predict expected values for all variables at the output of its decoder section. To quantify the reconstruction error, individual MAEs are first evaluated by calculating the difference between the input values and outputs values of each feature variable. The average of all individual MAEs is called the global MAE. Then, fault detection is performed by comparing the global MAE against a predetermined threshold. This threshold can be prescribed through observing deviations of data, for example, by setting it as 10% variation in normalized data. During normal operation of any FCU, the MAE (or reconstruction error) is low. In the presence of any faults or anomalies, a significant deviation of any inputs from their normal values results in a larger MAE for the affected variables. A global MAE exceeding the prescribed threshold indicates that a potential fault occurs, as the FCUs operation is likely deviating from its normal operation. The input feature variable(s) associated with the highest individual MAE(s) are typically implicated as the cause of the fault.

Moreover, the MAE can also be used as a relative performance score for: (i) monitoring and comparing performance across different or same types of FCUs; (ii) tracking the operational health of a single FCU over time. The latter is due to the increasing MAE values could provide indication for potential deteriorating operation and system health of any FCUs.

4 Results and Discussion

In Figure 6 and Figure 7, fault detection results are presented for a heating & cooling FCU using the global model and its specific model with both real and predicted data. As shown in Fig. 6, both the global model and specific model detect similar anomalies patterns of CLG-O·HTG-O > 0, while hot water was flowing through the heating coil, and chilled water was flowing through the cooling coil simultaneously, which are likely fault events.

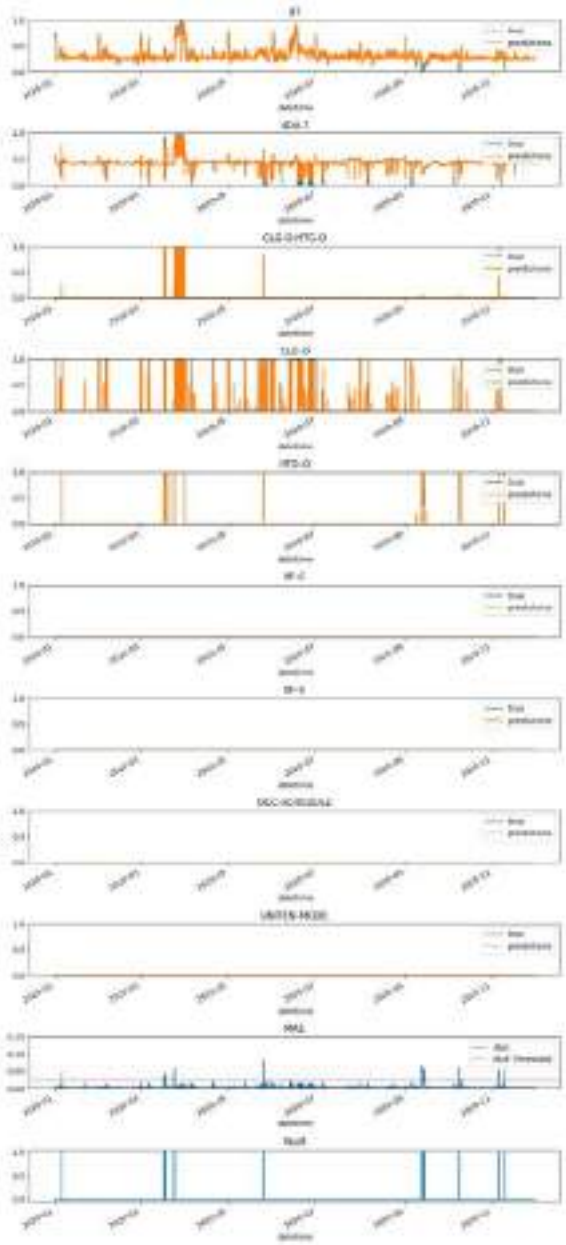


Figure 6: Prediction by the developed specific model for the heating & cooling FCU-HC1.

In particular, in April 2020, as both the supply air temperature (DA-T) and the room temperature (ZN-T) were increasing and causing spikes in dDA-T (= DA-T minus ZN-SP) and dT (= ZN-T minus ZN-SP), the FCU continued to supply hot water through the heating coil (HTG-O $\neq$ 0). Similarly, in Fig. 7, the global and specific models each detect the instance of CLG-O$\cdot$HTG-O > 0, which contribute to the peak of the overall MAE for fault detection. The threshold for MAE is calibrated such that that 0.5% of faults of data points is faulty. As expected, the global model predicts finer fluctuations less accurately than the specific model.

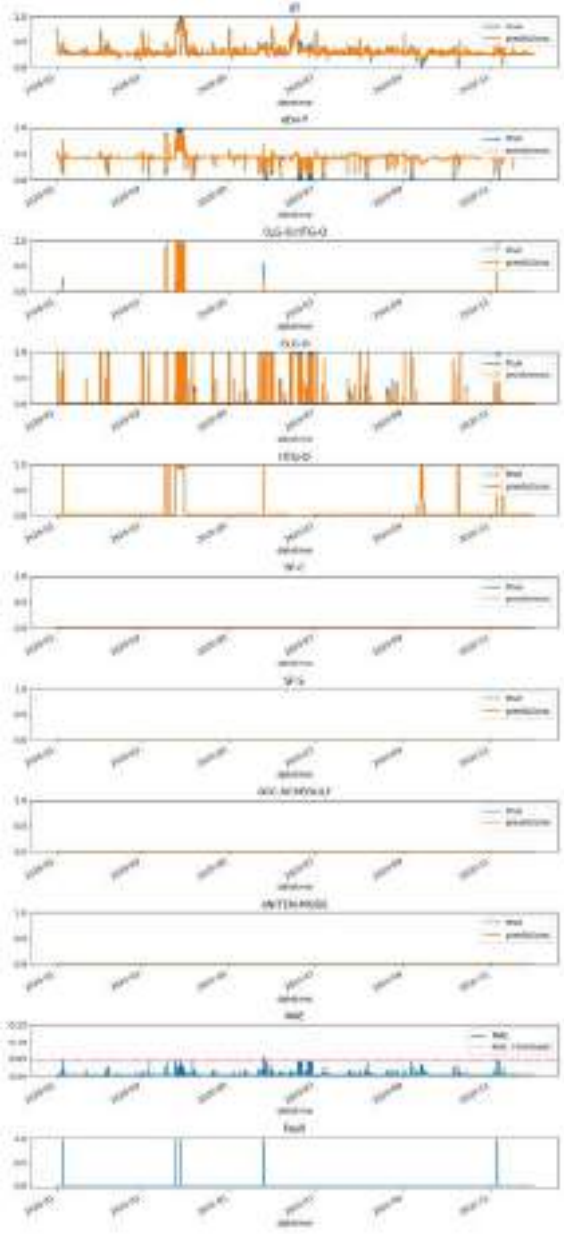


Figure 7: Prediction for the heating & cooling FCU-HC1 by the developed global model.

In Figure 8, fault detection results of the global model are presented for a heating-only FCU using real data and prediction. During the winter period of January to April 2020 in Canada, both the supply air temperature (DA-T) and the room temperature (ZN-T) regularly dropped below the setpoint, likely due to envelope heat loss. The heating valve output command (HTG-O) was regularly activated to supply hot water to the FCU's heating coil. Both the global model and the specific model (not shown) detect similar anomalies patterns of the FCU operation as faulty events, which lead to increasing total MAE above the detection threshold.

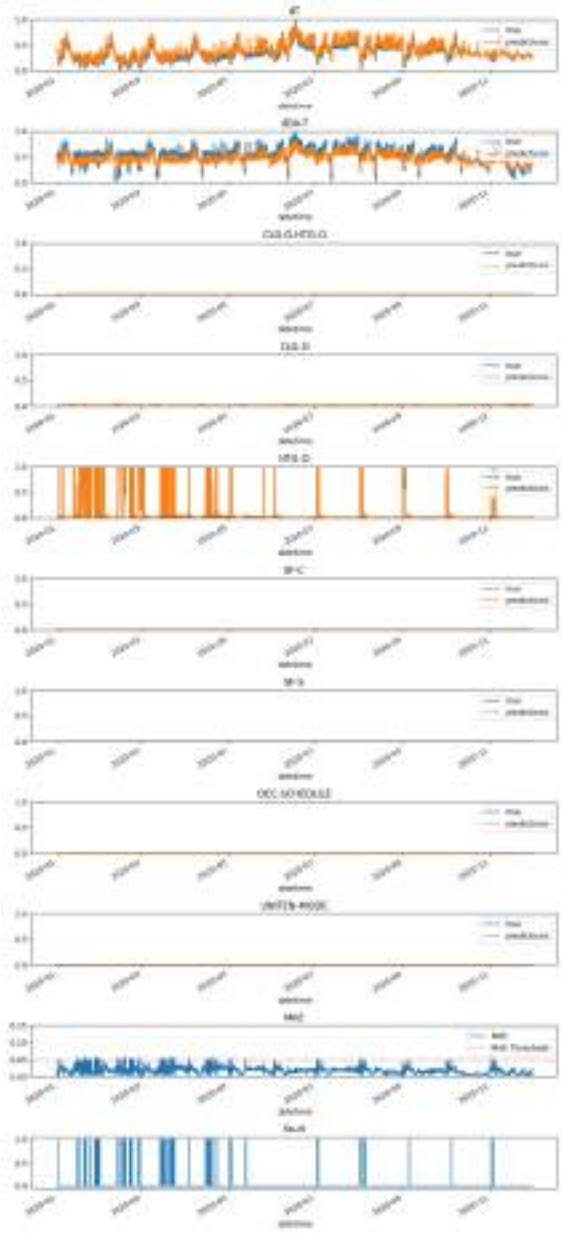


Figure 8: Prediction by the global model for the heating-only FCU-H2.

Data instances of normal operation as well as faulty ones are required for testing the developed models and evaluating the MAE for fault detection. Due to their scarcity, faulty data have been emulated by incorporating expected changes in appropriate sensor measurement data points for each FCU type [13].

In the left column of Figure 9, test results of the global model are presented for the heating-only FCU-H1 in detecting an emulated fault. The test dataset consists of 200 samples of normal data followed by 200 samples of emulated fault data of the heating valve being stuck open, despite not being commanded on (HTG-O = 0).

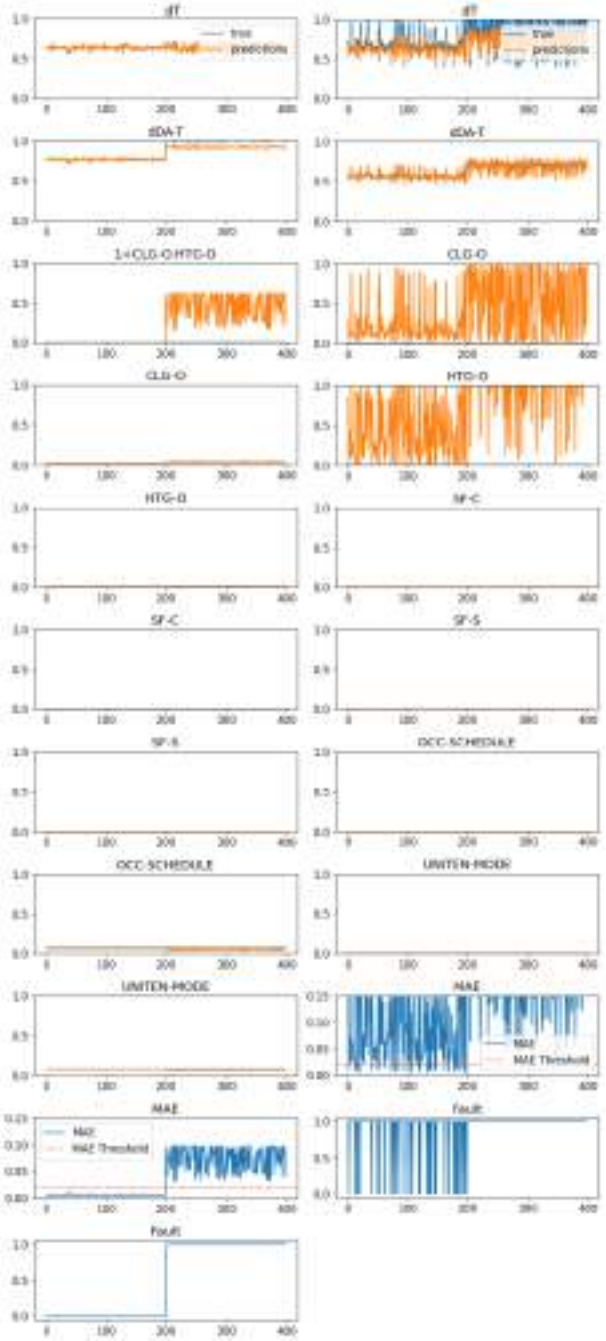


Figure 9: Detection of emulated fault by the: (left) global model with unsupervised multimodal learning, and (right) AE model without multimodal learning [13] for the heating-only FCU-H1.

The room temperature (ZN-T) and supply air temperature (DA-T) were modified accordingly to emulated full heating conditions, exhibiting higher supply air temperature (DA-T) values for the 200th to 400th samples, which the global model correctly identifies as faulty events, with *G* (= CLG-O·HTG-O+1)

and dDA-T (= DA-T minus ZN-SP) being the variables that lead to increasing total MAE above the detection threshold.

In the right column of Figure 9, fault detection results are presented for a recently published AE model in [13] without multimodal learning capability, after training it using historical data of 6 FCUs of the heating & cooling type. Without loss of generality, in-slab heating is omitted. In stark contrast, the AE model in [13] predicts that both the heating and cooling valve output commands (HTG-O and CLG-O) are to be activated throughout the normal and emulated fault intervals, in spite of FCU-H1 being a heating-only FCU. Due to lacking of multimodal learning capability, it fails to identify the actual emulated fault event, as the error between its predicted values and true values of dT, CLG-O and HTG-O significantly increase the total MAE above the detection threshold for all samples.

5 Conclusion

A multimodal AE is proposed for developing the global model and specific models for different types of FCUs through unsupervised multimodal learning: (i) the specific models are trained using data of a specific (or type) of FCU(s); (ii) the global model is trained using a combination of data from three types of FCUs. The applicability of unsupervised multimodal AE for fault detection across different types of FCUs is achieved through multimodal learning and transfer learning.

Acknowledgments

This research was funded by the Natural Science and Engineering Council of Canada (ALLRP 597645-24).

References

[1] H. Lee and J. Romero, "Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change," IPCC, Geneva, 2023.

[2] L. Perez-Lombard, J. Ortiz, R. Gonzalez and I. R. Maestre, "A review of benchmarking, rating and labelling concepts within the framework of building energy certification schemes," *Energy and Buildings,* 41, 272-278, 2008.

[3] A. Mahdavi, C. Berger, H. Amin, E. Ampatzi, R. Andersen, E. Azar, V. Barthelmes, M. Favero, J. Hahn, D. Khovalyg and H. Knudsen, "The role of occupants in buildings' energy performance gap: myth or reality?" *Sustainability,* 13(6), 3146, 2021.

[4] K. H. Andersen, S. B. Holøs, A. Yang, K. Thunshelle, Ø. Fjellheim and R. Jensen, "Impact of typical faults occurring in demand-controlled ventilation on energy and indoor environment in a Nordic climate," in *E3S Web of Conferences*, 2020.

[5] K. W. Roth, D. Westphalen, Llana and M. Feng, "The energy impact of faults in US commercial buildings," in *International Refrigeration and Air Conditioning Conference*, Purdue, 2004.

[6] A. Isazadeh, R. Kamal, C. Yagua, S. Eluvathingal and D. Claridge, "Detecting deficiencies using building performance data in healthcare facilities: Improving operational efficiency with Continuous Commissioning®," *Energy and Buildings,* 241, 110953, 2021.

[7] Y. Chen, G. Lin, Z. Chen, J. Wen and J. Granderson, "A simulation-based evaluation of fan coil unit fault effects," *Energy and Buildings,* 263, 112041, 2022.

[8] S. Pourarian, J. Wen, D. Veronica, A. Pertzborn, X. Zhou and R. Liu, "A tool for evaluating fault detection and diagnostic methods for fan coil units," *Energy and Buildings,* 136, 151-160, 2017.

[9] F. Rajabi and J. McArthur, "Applying OPTICS with and without PCA for fault detection of fan coil units using building automation system data," *Energy and Buildings,*114368, 2024.

[10] S. Dudley, M. Dey and S. Rana, "Smart building creation in large scale HVAC environments through automated fault detection and diagnosis," *Future Generation Computer Systems*, 108, 950-966, 2018.

[11] Y. Bengio, A. Courville and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 35(8), 1798-1828, 2013.

[12] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai and T. Chen, "Recent advances in convolutional neural networks,"*Pattern Recognition*, 77, 354-377, 2018.

[13] K. El Mokhtari and J. McArthur, "Autoencoder-based fault detection using building automation system data," *Advanced Engineering Informatics*, 62, 102810, 2024.

IFC Schema Extension for BIM-based Automated Code Checking of Korean Design Standards for Bridges

W. Lee[a], Y. Yu[b], B. Koo[a*], B. Choi[b], and S. Lee[b]

[a]Department of Civil Engineering, Seoul National University of Science and Technology, Republic of Korea
[b]Korea Construction Standards Center, Korea Institute of Civil Engineering and Building Technology, Republic of Korea
E-mail: wonbok@seoultech.ac.kr, youngsu@seoultech.ac.kr, *bonsang@seoultech.ac.kr, bhchoi@kict.re.kr, seunghwanlee@kict.re.kr

Abstract –

In Korea, research initiatives have commenced to digitize engineering and construction codes, and automate their compliance using BIM models for mandatory submission in public projects. The initiatives require employing IFC, a neutral and standard data format, to associate individual codes with relevant BIM elements. However, the IFC does not provide explicit extensions to represent the varied codes, which include the Korean Design Standards (KDS) and the Korean Construction Specifications (KCS).

To address this limitation, this study developed an extended IFC schema called 'IFC-KR', focusing initially on bridges. The extensions define subtypes to represent various types of bridges, structures, and elements, along with property sets that capture the relevant attributes and properties of 1,437 KDS codes. Moreover, a specialized extension module named 'IFC-KR Toolkit' has been developed to automate the addition of KDS data relevant to each element within the IFC Physical File.

To validate the functionality of the module, KDS items were applied to an Extradosed (ED) cable-stayed bridge model. The results confirmed that all data were accurately integrated, and the extended data was consistently verified using a third-party IFC viewer, demonstrating its accuracy and interoperability.

Keywords –

BIM; IFC Schema Extension; Automated Code Checking; Korean Design Standards; Bridges

1 Introduction

In Korea, engineering and design submittals for all public projects must be checked for their conformance to Korean Design Standards (KDS) and Construction Specifications (KCS) [8]. However, the current code checking process remains highly inefficient as it relies on engineers manually verifying KDS presented in textual format. Moreover, this manual process significantly hinders the ability to promptly respond to design changes or updates of these codes [17].

The Korean Ministry of Land, Infrastructure, and Transport (MOLIT) is making significant efforts to overcome these limitations by developing a framework for BIM-based automated code checking (ACC) [12]. At the core of these initiatives is the digitalization of KDS, which involves the conversion of individual provisions into digital rules and parameters that can be seamlessly integrated with BIM model data, thereby enable the ACC process.

Associating BIM models with KDS data requires employing the Industry Foundation Classes (IFC), the international standard data format, which allows customized extensions of its schema for local or regional data sharing requirements [11]. However, the standard IFC schema lacks formal definitions for the entities and property sets (Psets) required to integrate KDS tailored to Korea.

The Building and Construction Authority (BCA) in Singapore developed an extended IFC schema, known as 'IFC-SG[1]'. IFC-SG defines custom entities and Psets to ensure that data within IFC models aligns with Singapore's information mandates of BIM models, and is used as the medium for information exchange in CORENET-X [2] [3]. However, IFC-SG is designed exclusively for the building permit domain and is not suitable for accommodating KDS related data that relates to infrastructure assets and engineering specifications.

This study aimed to develop an extended IFC schema, termed 'IFC-KR', specifically tailored to accommodate data related to Korea's design standards while integrating

[1] https://www1.bca.gov.sg/regulatory-info/building-control/corenet-x/resources/ifc-sg-resource-toolkit
[2] https://www1.bca.gov.sg/regulatory-info/building-control/corenet-x

specialized engineering knowledge pertinent to the infrastructure domain. To enhance practical applicability, the study also developed 'IFC-KR Toolkit', a module that enables the insertion of required data into the IFC Physical File (IPF) in accordance with the IFC-KR schema.

This study focuses on bridges, a representative infrastructure facility type, and is structured as follows: Section 2 reviews the development status and limitations of the standard IFC schema and previous studies on IFC schema extensions. Section 3 presents the research methodology. Section 4 details the development process of the IFC-KR schema, and Section 5 outlines the methodology for developing the IFC-KR Toolkit. Section 6 validates the extension results in IFC-KR Toolkit.

2 Research Background

2.1 Current Status and Limitations of the Standard IFC Schema

buildingSMART International (bSI) developed the IFC, a neutral standard model, to enable seamless data exchange between BIM applications. The standard IFC schema provided by bSI encompasses the specification of objects, attributes, and relationships within BIM models and undergoes continuous updates to ensure its adaptability and relevance [10].

The initial IFC schema was primarily designed to represent the architectural domain [9]. However, the release of IFC4 in 2013 introduced support for infrastructure-related objects, focusing primarily on bridge facilities. In particular, bridge-related elements were formally integrated, with the addition of the *IfcBridge* and *IfcBridgePart* entities, geometric representations for horizontal structures and structural connectivity, and data attributes for load conditions. This was further expanded with IFC4.3 in 2021, which added entities and attributes to accommodate a broader range of infrastructure facilities, including railways, and roads [4].

Despite the continuous updates to the IFC schema, several factors still hinder its practical application. The main challenges include: 1) limited scope of application, 2) difficulties in adopting national and regional requirements, and 3) incomplete representation of domain-specific data requirements. The detailed issues for each challenge are as follows:

1) Limited scope of application

While the IFC schema has been updated to include new entities, its initial development was heavily focused on the architectural domain. Consequently, other domains, such as infrastructure, often had to adopt architectural entities to represent their specific objects. This limitation has been identified as a major barrier to achieving interoperability in non-architectural domains [14, 18]. In particular, Korean infrastructure projects frequently include tunnels, yet the IFC schema does not provide explicit definitions for tunnel-related entities. As a result, architectural entities must still be adopted to represent tunnels, which limits their practical application and interoperability.

2) Difficulties in adopting national and regional requirements

Although the IFC schema is designed as a standardized data structure, it has limitations in adopting the diverse requirements of various nations and regions. Furthermore, updates to the schema must undergo a rigorous review process by the bSI Standards Committee, which is time-consuming and makes it challenging to promptly incorporate necessary data into the schema [2]. This limitation is particularly evident in Korean infrastructure projects, where the KDS require detailed structural analysis and geometric attributes. However, the current IFC schema lacks predefined entity types and attributes for these requirements, making it difficult to directly integrate KDS compliance into BIM-based workflows.

3) Incomplete representation of domain-specific data requirements

bSI has continuously expanded the scope of IFC to address the evolving requirements of the construction industry. Initially, structural elements and analysis domains were introduced, followed by MEP domains, and more recently, infrastructure domains such as roads and railways. However, throughout this process, many newly defined entities have failed to fully capture the semantic representations and attributes required by their respective domains. Such limitations in representing essential information hinder the full utility of IFC, posing significant constraints on its applicability. For instance, while IFC4 introduced bridge-related entities, essential elements widely used in Korean bridge projects, such as joints and approach slabs, are still missing. This absence limits the accuracy and completeness of bridge models within the IFC schema, necessitating extensive customization to meet Korean infrastructure project requirements. Additionally, gaps in representation, interoperability, and compliance with engineering standards were identified based on practical application cases [13]. These limitations highlight the need for extensions such as IFC-KR to enhance adaptability while maintaining compatibility with the existing IFC framework.

2.2 Approaches for IFC Schema Extension

To address the limitations of the standard IFC, many studies have proposed to extend its schema. These

extensions are broadly classified into two methods: Entity Extension, which involves the definition of new entities, and Property Extension, which focuses on integrating and associating additional data with existing entities.

2.2.1 Entity Extension

This method involves defining new entities within the existing schema to represent previously unsupported elements and processes. By enabling the detailed representation of domains and entities that were not previously addressed, it provides the advantage of supporting tailored and highly precise data modeling [6]. However, the outcomes of this method are often not officially incorporated into the standard IFC schema. As a result, the added entities can only be utilized in BIM authoring tools and IFC viewers specifically modified to recognize these custom entities, presenting a significant limitation.

2.2.2 Property Extension

This method involves adding new properties, attributes, or subtypes to existing entities, primarily utilizing Type entities (e.g., *IfcBuiltElementType*) or Property sets (Psets). For example, the *IfcColumn* entity can be further specified using *IfcColumnType* entity, a subtype of *IfcBuiltElementType* entity. Additionally, when further detailed classifications are required, new types can be added using the USERDEFINED type.

Psets are employed to define general attributes, such as dimensions or materials, and are linked with *IfcObject* entities using *IfcRelDefinesByProperties* entity. This method preserves the original IFC schema while allowing new data to be added, minimizing compatibility issues across BIM authoring tools and IFC viewers.

Recently, the 'Linked Data Approach' has emerged as a method to effectively connect external data to the IFC schema while maintaining data integrity. This method links external data stored in the buildingSMART Data Dictionary (bSDD) and Object-Type Libraries (OTLs) to the IFC schema, enabling connectivity between IFC models and various data sources [16].

bSI does not recommend creating new entities that may be limited to specific use cases. Rather, it prioritizes the use of property extensions by using predefined classifications, types, and Psets within the standard IFC schema [1].

3 Research Methodology

This study aims to develop an extended IFC schema and extension module to facilitate BIM-based automated code checking in the bridge domain. As illustrated in Figure 1, the research is structured into three phases: 1) development of an extended IFC schema, 'IFC-KR', 2) development of a module for creating IFC-KR extensions, and 3) validation of the extended IFC using the extension module.

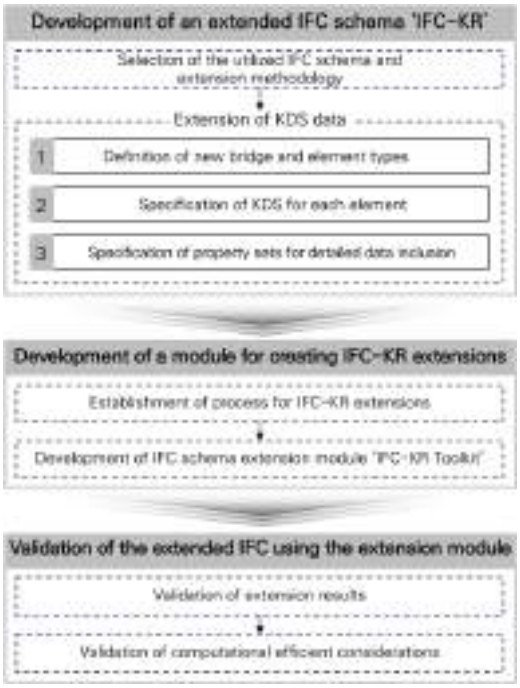


Figure 1. Research process

4 Development of 'IFC-KR'

4.1 Selection of the Utilized IFC Schema and Extension Methodology

IFC-KR is an extended IFC schema specialized for incorporating data related to the KDS. We developed the optimal IFC schema and extension method based on the extension approaches discussed in Section 2. Additionally, appropriate software was selected to implement the IFC-KR Toolkit, a module designed to extend the IFC schema and incorporate KDS data for individual elements into the IPF.

The standard IFC schema targeted for extension was selected as IFC4.3, which effectively represents bridge structures and infrastructural characteristics. IFC4.3 includes the *IfcBridge* entity for representing bridges, and bridge types such as Arch, Truss, and Girder bridges can be defined using the *IfcBridgeTypeEnum* entity. Additionally, structural elements of bridges, such as Abutments, Piers, and Superstructures, can be defined using the *IfcBridgePartTypeEnum* entity.

The property extension approach was adopted to incorporate the KDS data. Individual properties served as the basic units of data for this extension, while Pset was

utilized to organize and extend the schema when multiple KDS data items needed to be included.

This approach allows additional data to be incorporated without altering the standard IFC schema, ensuring that IFC-KR remains compatible with existing IFC-supported applications. Since the extensions are structured as additional Psets rather than modifications to core schema entities, software that supports the base IFC schema can still process IFC-KR schema without interoperability issues.

In the process of applying the Property Extension methodology, the Data Linkage Approach, inspired by the Linked Data Approach, was utilized. Similar to conventional methods, this approach connects external data to Psets. However, instead of relying on bSDD or OTLs, it is specifically designed to directly link to external files (e.g., .xml, .csv) containing digitized KDS. For this study, Autodesk Revit 2025, equipped with an IFC Exporter capable of extracting in the IFC4.3 format, was utilized as the BIM authoring tool. For the IFC viewer, KBim Assess-Lite[3], an open-source application with a proprietary parser enabling fast data exchange, was employed. The IFC-KR Toolkit, the module developed specifically for this study, was developed as an add-in in KBim Assess-Lite.

4.2 Extension of KDS Data

4.2.1 Definition of New Bridge and Element Types

This study focused on three bridge types most commonly utilized in Korea: PSC girder bridges, Steel girder bridges, and Extradosed (ED) cable-stayed bridges. In IFC4.3, the *IfcBridge* entity defines bridge types using the predefined type, *IfcBridgeTypeEnum* entity. However, the aforementioned three bridge types are not included within the predefined options. Thus, these bridge types were newly defined under the USERDEFINED type in *IfcBridgeTypeEnum* entity.

For elements, the *IfcElement* entity was used, with detailed subtypes defined by the enumeration type of the *IfcElementType* entity. For this study, 16 subtypes already defined in the standard IFC schema and applicable to the three bridge types were utilized without modification. For the nine element types not covered by the standard IFC schema, new subtypes were introduced as USERDEFINED types (Table 1).

Table 1. Subtypes by Elements in IFC-KR

Element (IFC4.3 entity)	Subtypes (*=USERDEFINED)
Anchor (*IfcTendonAnchor*)	TENSIONING_END
Approach slab (*IfcSlab*)	*APPROACH_SLAB
Base plate (*IfcPlate*)	BASE_PLATE
Bearing (*IfcBearing*)	ELASTOMERIC
Bracing (*IfcMember*)	BRACE
Bracing wall (*IfcWall*)	*BRACING_WALL
Cable (*IfcMechanicalFastener*)	*CABLE
Cast in place pile (*IfcPile*)	BORED
Column (*IfcColumn*)	PIERSTEM
Continuous slab (*IfcSlab*)	*CONTINUOUS_SLAB
Covering (*IfcCovering*)	COPING
Cross beam (*IfcBeam*)	*CROSSBEAM
Deck plate (*IfcPlate*)	BASE_PLATE
Diaphragm (*IfcBeam*)	DIAPHRAGM
Girder (*IfcBeam*)	GIRDER_SEGMENT
Parapet wall (*IfcWall*)	PARAPET
Pile foundation (*IfcPile*)	SUPPORT
Pylon (*IfcElementAssembly*)	PYLON
Steel pipe pile (*IfcPile*)	*STEEL_PIPE
Wall (*IfcWall*)	STANDARD
Wing wall (*IfcWall*)	RETAININGWALL
Joint (*IfcMember*)	*JOINT
Shear connector (*IfcMechanicalFastener*)	SHEARCONNECTOR
K-braced Diaphragm (*IfcBeam*)	*KBRACED_DAPHRAGM
PC_HOUSE (*IfcBuildingElementProxy*)	*PCHOUSE

4.2.2 Specification of KDS for Each Element

[15] developed a ‘standards map’ based on bridge design and construction processes that visualizes the code checking sequence and conditions for each element, and links them to the applicable KDS. This study utilized these standards maps to identify the required KDS for each element type and categorized KDS items by considering both element units and code checking procedures (Table 2).

Table 2. Classification of KDS items by elements and code checking procedures

Bridge type	Element type	Code checking procedure	KDS item
PSC girder bridge	Cross beam	Selection of design conditions	KDS 24 14 21_ 4.2.1.2_Table 4.2-1
			KDS 24 14 21_ 4.2.1.2_Table 4.2-2
		Load calculation and combination	KDS 24 12 21
			KDS 24 12 11
		Review of cross beam cross-section	KDS 24 17 11_ 4.6.3.3
			KDS 24 14 21_ 4.2.3.3
ED cable-stayed bridge	Pylon	Selection of design conditions	KDS 24 14 21_ 4.4.4.2_Table 4.4-3
			KDS 24 14 21_ 4.4.1_Table 4.4-1
		Load calculation and combination	KDS 24 12 22
			KDS 24 12 12
		Design of pylon cross-section	KDS 24 14 21_ 4.1.2.1
			KDS 24 14 21_ 4.6.2.6 (5)

4.2.3 Specification of Psets for Detailed Data Inclusion

[5, 7] classified the specific KDS codes for bridge

[3] Open-source IFC viewer developed by the national research project (https://www.inno-lab.co.kr/Home/kr/product-KBim_Assess- Lite.html)

structures based on quantitative and qualitative characteristics. They identified the input variables necessary for performing code checking for each item and the output variables required to represent the results.

KDS items were categorized as: 1) quantitative items requiring formula-based calculations and 2) qualitative items requiring engineering decisions (Table 3). Ultimately, only the quantitative items suitable for digitalization were selected, resulting in 1,437 KDS items being utilized.

Table 3. Examples of quantitative/qualitative classification items

Category	Contents of KDS
Quantitative	KDS 11 50 15_4.1.1.1 (2) [Allowable axial displacement of pile foundation] : The allowable axial displacement of the pile foundation shall be within the allowable displacement of the superstructure.
Qualitative	KDS 11 50 15_4.2.3.4 (4) [Consolidation settlement] : Other consolidation settlement can be calculated through appropriate numerical analysis methods such as finite element analysis.

4.3 Detailed Schema of IFC-KR

The IFC-KR schema, designed to incorporate the newly defined items, was developed with a three-tier structure comprising 'Object taxonomy, KDS mapping, and Psets'. The schema was further divided into separate sub-schemas to align specifically with KDS requirements.

1) Object taxonomy tier

The object taxonomy tier defines the IFC entities required to represent the different types of bridge, structural composition, and element. It specifies the structures and elements associated with each bridge type. Specifically, it consists of (1) bridge type (*IfcBridge*), (2) structural composition (*IfcBridgePart*), (3) element name, (4) IFC4.3 entity (*IfcElement*), and (5) subtypes of each element (*IfcElementType*).

The taxonomy was developed to include the three bridge types and nine element types defined in Section 4.2.1. An excerpt of the KDS object taxonomy tier is provided in Figure 2. For example, 'PSC Girder Bridge' is defined using the *IfcBridge* entity, while one of its sub-elements 'ABUTMENT' is defined using the *IfcBridgePart* entity. The specific element 'Wall' is assigned using IFC4.3 entity *IfcWall*, while its subtypes such as the 'STANDARD' and 'RETAININGWALL' are defined using the enumeration type of the *IfcElementType* entity.



Figure 2. Example of object taxonomy tier (KDS)

2) KDS mapping tier

The KDS mapping tier defines the KDS items that need to be checked for individual elements previously specified through the object taxonomy tier. For example, the Abutment wall needs to be checked for its load calculation and combinations. These correspond to 'KDS 24 14 21_4.6.6.2 (1)' and 'KDS 24 14 21_4.6.6.2 (2)'. A sample of the KDS mapping tier is shown in Figure 3.



Figure 3. Example of KDS mapping tier

3) Psets tier

The Psets tier define the data required for the quantitative code checking of the 1,437 KDS items. Each Pset contains the following items: (1) Pset name (KRPset Name), (2) KDS code, (3) BIM design values, and (4) Results. The detailed structure of the Pset tier for the 'Load calculation and combination' item (KDS 24 14 21_4.6.6.2 (2)) related to the wall element of a PSC girder bridge is given in Table 4, with an example input shown in Figure 4.

In Table 4, 'KRPset name' indicates the name of the design code associated with the element, while 'Property name' indicates the names of the input/output variables. 'Property value' represents the detailed values of these variables, and 'Property type' indicates the specific data type. Data types are expressed as *IfcLabel* for strings, *IfcReal* for real numbers, and *IfcBoolean* for boolean values. Finally, the 'Property unit' defines the unit of the detailed values.



Figure 4. Example of Psets tier (KDS)

Table 4. Example of Pset for 'KDS 24 14 21_4.6.6.2 (2)'

KRPset name	Property name	Property value	Property type	Property unit
KRPset_ KDS 24 14 21_ 4.6.6.2 (2)	KDS code	24 14 21_ 4.6.6.2 (2)	*IfcLabel*	-
	Axial reinforcing bar cross-sectional area in normal section, not at lap joint		*IfcReal*	mm^2
	Cross-sectional area of a column		*IfcReal*	mm^2
	Cross-sectional area of prestressing steel		*IfcReal*	mm^2
	Design tensile strength of prestressing steel		*IfcReal*	Mpa
	Design basis yield strength of axial reinforcing bar		*IfcReal*	Mpa
	Specified compressive strength of concrete		*IfcReal*	Mpa
	Results		*IfcBoolean*	-

5 Development of Module for creating IFC-KR extensions

5.1 Process of creating IFC-KR extensions

A four-step process for inserting KDS data into individual elements within an IPF using the extended IFC structure, IFC-KR, is presented in Figure 5.

The first step involves specifying the bridge type and individual structure types for the IFC4.3 model extracted from a BIM authoring tool. The second step determines the appropriate entity and subtype according to the specified element type. The third step selects the design checking procedures applicable to the specified element type and displays the relevant KDS items. Finally, detailed data is displayed for each checking item, including 'KDS code, BIM design values, and results' with the option to manually enter detailed BIM design values if necessary.

5.2 IFC Schema Extension Module 'IFC-KR Toolkit'

The process was implemented by developing the 'IFC-KR Toolkit'. This module was integrated as an add-in within KBim Assess-Lite. The User Interface (UI) was designed to display appropriate KDS provisions for the selected element by allowing users to specify consecutively, the Object taxonomy, KDS mapping, and Psets tier of IFC-KR. The module operates on IPFs extracted from BIM authoring tools (e.g., Autodesk Revit) which need to be extracted using version IFC4.3 and Reference View. The module also allows for the modification of incorrectly entered element types and enables the selection of unmapped additional items from the complete KDS list.

6 Validation of Results

6.1 Validation of extension results

To verify the effectiveness of the developed module, the validation process was conducted in three stages: 1) data integration accuracy, 2) property set consistency, and 3) IFC standard compliance.

1) Data integration accuracy

The IFC-KR was applied to an ED cable-stayed bridge model, where five KDS items were mapped to three key elements (pylon, girder, and pier). The IFC-KR Toolkit was used to insert the required KDS properties into the IPF (.ifc). The extended IPF file was then inspected to confirm that the correct KDS properties were assigned to each corresponding element.

2) Property set consistency

To ensure the accuracy of the inserted data, the generated IFC file was loaded into KBim Assess-Lite, where the extended KDS properties were checked within the 'Property Set' tab. Each property was manually verified to confirm that its naming, data types (e.g., *IfcReal*, *IfcBoolean*), and data structure matched the intended KDS format (Figure 6).

3) IFC standard compliance

Finally, the extended IFC-KR model was exported and tested using BIMvision, a commercial IFC viewer supporting IFC4.3. This step ensured that the inserted KDS properties were correctly interpreted and displayed without loss of data or formatting inconsistencies. Additionally, interoperability with third-party IFC viewers was assessed by verifying that the modified IFC schema maintained compatibility with existing BIM workflows.

The results of this validation confirmed that all KDS data was successfully embedded into the IFC model and could be accurately visualized and interpreted in external IFC software (Figure 7).

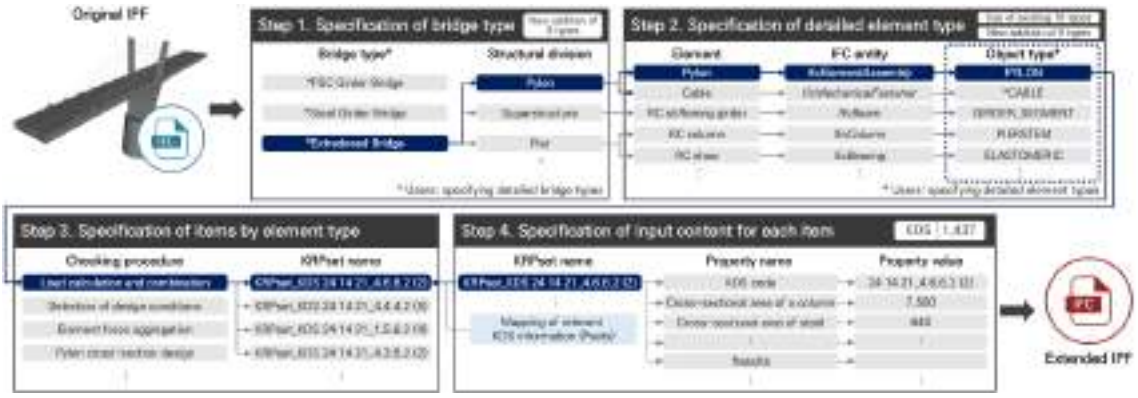


Figure 5. Process of extending KDS data using IFC-KR

This study primarily focused on the validation of an ED cable-stayed bridge as an example because ED bridges have a more complex structural configuration compared to PSC girder bridges and steel girder bridges. Given the higher complexity of ED bridges, it serves as a more rigorous test case for validating the IFC-KR schema. Additionally, the same validation process was applied to PSC girder bridges and steel girder bridges, and the results confirmed that the IFC-KR schema was equally effective for these bridge types as well.

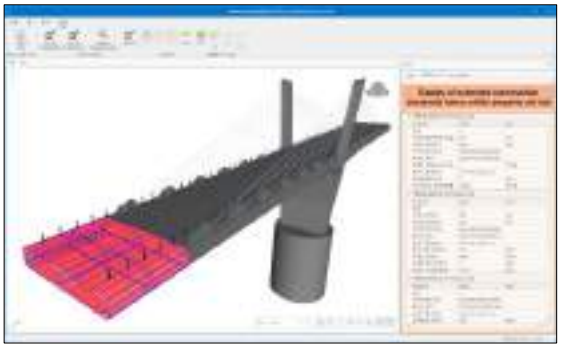


Figure 6. Extension results in IFC-KR Toolkit



Figure 7. Validation of extension results in BIMvision

6.2 Computational Efficiency Considerations

The IFC model used in this validation had a file size of 80MB, and the extension process using the IFC-KR Toolkit took approximately 15 seconds. However, it is important to note that the validated IFC model contained only a limited number of reinforcing bars. For high-LOD models with a significant number of reinforcing bars, the processing time may increase substantially. As such, further research will focus on evaluating computational efficiency under different model complexities and exploring potential optimization strategies to enhance performance.

7 Conclusion

This study introduced the extended IFC schema, 'IFC-KR', and 'IFC-KR Toolkit', a module dedicated to creating IFC-KR extensions, to support BIM-based ACC for Korean design standards.

IFC-KR was built on the IFC4.3 standard, with three bridge types and nine detailed element types newly defined as USERDEFINED types. Additionally, 1,437 KDS items were mapped, and the required data for each item, including input/output variables and BIM design values, was extended in the form of Psets.

The IFC-KR Toolkit was developed as a tool that enables the creation of IFC-KR compliant IPF's. The module allows the input and display of KDS items and Pset data for each element during the extension process.

To validate the module, KDS items were extended for the pylon, girder, and pier elements of an ED cable-stayed bridge model. The results confirmed that all KDS data was successfully populated, which was further verified by inspecting the data exported to a commercial IFC viewer, BIMvision.

This study formalized a framework for incorporating KDS data into IFC, and thereby enabled extended BIM-based ACC to include engineering specification checking for the infrastructure domain. This work is regarded as an important step in establishing a foundational framework for future BIM-based ACC implementation.

The stability of the IFC-KR schema in relation to future IFC standard updates was also considered during

development. Since IFC-KR does not directly modify the core IFC schema but extends it through additional property sets, it remains stable within the same IFC version (IFC4.3). Significant changes within the same version are unlikely, ensuring that the proposed extensions remain functional in the foreseeable future. However, future iterations of the IFC standard may introduce updates to infrastructure-related entities, which could affect certain extensions. Therefore, continuous monitoring and adaptation will be necessary to ensure long-term compatibility and maintain interoperability across evolving BIM standards.

8 Acknowledgement

This research was supported by the "Research of Construction Codes Library and Ontology for Digital Transformation of Construction Codes" project, part of the "Research on the Development of Construction Standards Operated by the Korea Construction Standards Center" project of the Korea Institute of Civil Engineering and Building Technology (KICT).

References

[1] Borrmann, A., Amann, J., Chipman, T., Hyvarinen, J., Liebich, T., Muhic, S., Mol, L., Plume, J., and Scarponcini, P. IFC Infra Overall Architecture Project Documentation and Guidelines, buildingSMART International. On-line: https://www.buildingsmart.org/wp-content/uploads/2017/07/08_bSI_OverallArchitecture_Guidelines_final.pdf, Accessed: 19/11/2024.

[2] Borrmann, A., Muhic, S., Hyvarinen, J., Chipman, T., Jaud, S., Castaing, C., Dumoulin. C., Liebich. T., and Mol. L. The IFC-Bridge Project–Extending the IFC Standard to Enable High-Quality Exchange of Bridge Information Models, In *Proceedings of the 2019 European Conference on Computing in Construction*, pages 377-386, Chania, Greece, 2019.

[3] Building and Construction Authority (BCA). CORENET X. (Report No. URA/PB/2023/01-DCG). On-line: https://www. corenet.gov.sg/media/2391818/appbca-2023-02.pdf, Accessed: 02/12/2024.

[4] buildingSMART International (bSI). IFC Specifications Data base. On-line: https://technical.buildingsmart.org/standards/ifc/ifc-schema-specifications, Accessed: 22/11/2024.

[5] Choi, J. W., Kim, J. Y., Im, M. I., Hwang, J. B., and Chi, S. H. Development of a Construction Specification Digitalization Methodology for Smart Construction, In *Proceedings of KSCE 2023 Convention*, pages 303-304, Yeosu, Republic of Korea, 2023.

[6] Eastman, C. *BIM Handbook: A Guide to Building Information Modeling for Owners, Managers, Designers, Engineers and Contractors.* John Wiley & Sons, Hoboken, NJ, 2011.

[7] Jang, J. G., Seo, S. H., Lee, S. J., and Zi, G. S. Digital Code Transformation for Establishment of Digital Utilization System, In *Proceedings of KSCE 2023 Convention*, pages 402-403, Yeosu, Republic of Korea, 2023.

[8] Kim, S., Kim, T. S., and Park, H. P. Development of Korean code system for construction specifications and design standards. *KSCE Journal of Civil Engineering*, 20, 1605-1612, 2016.

[9] Lee, S. H., Park, S. I., Kwon, T. H., and Seo, K. W. Civil Infrastructure Information Modeling Method Based on Extended IFC Entities using BIM Authoring Software. *Journal of the Computational Structural Engineering Institute of Korea*, 30(1), 77-86, 2017.

[10] Liu, L., Li, B., Zlatanova, S., and Oosterom, P.V. Indoor Navigation Supported by the Industry Foundation Classes (IFC): A survey, *Automation in Construction*, 121, 1-20, 2021.

[11] Ministry of Land, Infrastructure and Transport. BIM Implementation Guidelines for the Construction Industry, On-line: https://www.molit.go.kr/USR/policyData/m_34681/dtl.js p?id=4634, Accessed: 19/11/2024.

[12] Ministry of Land, Infrastructure and Transport. In the Future, Design Errors in Construction Projects will be Automatically Reviewed by Computers, On-line: https://www.molit.go.kr/USR/NEWS/m_71/dtl.jsp?lcmspage=1&id=95089150, Accessed: 22/11/2024.

[13] Park, S. I., Park, J., Kin, B. G., and Lee, S. H. Improving applicability for information model of an IFC-based steel bridge in the design phase using functional meanings of bridge components, *Applied Sciences*, 8(12), 2531, 2018.

[14] Park, S. I., Lee, S. H., Almasi, A., and Song, J. H. Extended IFC-based Strong form Meshfree Collocation Analysis of a Bridge Structure, *Automation in Construction*, 119, 1-21, 2020.

[15] Park, E. W., Sung, W. K., and Kang, D. J. Establishment of Bridge Design & Construction Process and Their Reference Map for Digitalization of Korea Construction Standards. In *Proceedings of KSCE 2022 Convention*, pages 84-85, Busan, Republic of Korea, 2022.

[16] Pauwels, P., McGlinn, K., Törmä, S., and Beetz, J. Linked data. *Building information modeling: Technology foundations and industry practice*, 181-197, 2018.

[17] Preidel, C., and Borrmann, A. BIM-based code compliance checking. *Building information modeling: Technology foundations and industry practice*, 367-381, 2018.

[18] Won, J., Kim, T., Yu, J., and Choo, S. Development of the IFC Schema Extension Methodology for Integrated BIM, In *Proceedings of eCAADe 2022 Conference*, pages 339-346, Ghent, Belgium, 2022.

A Mixed Reality Dataset and Benchmarking of Deep Learning Models for MEP Classification

Boan Tao[1], Jiajun Li[1] and Frédéric Bosché [1]

[1]School of Engineering, University of Edinburgh, Edinburgh, Scotland, UK

boan.tao@ed.ac.uk, Jiajun.Li@ed.ac.uk, f.bosche@ed.ac.uk,

Abstract -

This paper introduces a Mixed Reality (MR)-captured dataset and a benchmark evaluation of deep learning models for Mechanical, Electrical, and Plumbing (MEP) component classification. The dataset, collected using Microsoft HoloLens 2, comprises eight distinct building environments containing pipes and ducts, presenting unique challenges characteristic of MR-captured data, including lower point density and increased noise compared to laser scans. In addition to presenting the dataset, this study provides a comparative evaluation of nine state-of-the-art deep learning architectures, assessing their performance and computational efficiency through rigorous cross-validation. Our analysis reveals that PointNN, a non-parametric approach, achieves superior classification accuracy (90.21%) while maintaining minimal computational requirements (1.34 GFLOPs). This dual contribution of this paper aims to advance automated MEP inspection systems and provide insights for developing efficient deep learning models for construction applications.

Keywords -

Mixed Reality; Point Cloud Classification; Deep Learning; MEP Systems; Construction Inspection; Building Information Modeling;

1 Introduction

Mechanical, Electrical, and Plumbing (MEP) systems form intricate networks essential for providing services such as heating, ventilation, air conditioning (HVAC), water distribution, and waste management. The inherent complexity of these systems introduces significant challenges in their design, installation, and maintenance, which result in inefficiencies, conflicts, and costly rework during construction [1]. Recent advancements in digital technologies have offered promising solutions to address these challenges. Notably, the integration of Mixed Reality (MR) and Artificial Intelligence (AI) can contribute to improving MEP system construction, monitoring and management.

Mixed Reality (MR) devices, such as the Microsoft HoloLens 2, provide unprecedented capabilities to overlay digital information onto the physical world, enabling real-time visualisation and interaction with virtual MEP components in their actual context [2]. These visualisation capabilities can be enhanced by deep learning algorithms that efficiently process and analyse complex three-dimensional data, including point clouds representing MEP systems [3]. For example, mobile scanning solutions like Apple's RoomPlan API [4] integrate device sensors with machine learning to create parametric 3D models of interior spaces. Such systems leverage camera feeds and LiDAR sensors to automatically detect and classify structural elements like walls, windows, and openings, while simultaneously recognising room features and fixtures, providing a comprehensive digital representation of the built environment. However, while such tools exist for capturing and modelling architectural elements, there is currently no comparable automated solution for MEP system mapping. Even with high-quality laser scanning data, the automated detection and classification of MEP components remains a significant technical challenge. This is due to several factors, including the inherent complexity and diversity of MEP systems, as well as occlusions commonly encountered in densely packed service areas.

This paper presents two key contributions:

1. It introduces a novel MR-captured dataset for MEP component classification, detailing its acquisition, preprocessing, and characteristics.
2. It benchmarks multiple 3D deep learning architectures on this dataset, providing insights into their classification performance and computational efficiency.

The remainder of this paper is organised as follows: Section 2 provides a comprehensive review of existing 3D datasets in construction, challenges in MR-based data capture, and recent advances in point cloud classification methods. Section 3 presents our dataset acquisition and preparation methodology, including detailed scanning protocols, dataset statistics, and cross-validation strategies. Section 4 introduces the selected deep learning models and evaluation metrics, along with implementation specifications. In Section 5, we present a detailed analysis of model performance and computational efficiency, accompanied by insights into the relative strengths and limitations of different architectural approaches. Within this section, we also discuss the implications of our findings for practical deployment and outline future research directions.

Finally, Section 6 summarises our key contributions and their significance for advancing MR-based MEP inspection in the Architecture, Engineering, and Construction (AEC) industry.

2 Literature Review

2.1 3D Datasets in Construction

Many point cloud datasets used in research are not fully representative of the challenges faced in construction environments. For instance, the ModelNet dataset [5] is a popular benchmark for 3D shape classification and segmentation, but it consists of CAD models converted to point clouds, lacking the noise and incompleteness typical of real-world scans. Similarly, ShapeNet [6] is a large-scale dataset of 3D shapes, but it also consists of clean, complete 3D models rather than actual scanned point clouds. These datasets are valuable for developing and testing algorithms, but they don't capture the complexity, imperfections and overall challenges of real-world construction scenes.

Recognising this limitation, Uy et al. [7] introduced ScanObjectNN, a real-world point cloud dataset for object classification. ScanObjectNN includes point clouds extracted from scanned indoor scenes, preserving real-world characteristics like background, occlusion, and missing parts. While more realistic than synthetic datasets, ScanObjectNN focuses on common indoor objects (e.g., chairs, tables, beds, cabinets, displays) rather than construction-specific or MEP components.

For construction-specific applications, several datasets have been developed to address the unique challenges of the built environment. The S3DIS dataset [8] provides large-scale, real-world point clouds of indoor spaces, including structural elements and some building components (e.g., doors, windows, columns, beams). The Paris-Lille-3D dataset [9] offers large-scale outdoor point clouds of urban environments, which can be relevant for certain construction tasks.

In more direct relation to MEP components (the focus of our paper), Yin et al. [10] established a benchmark industrial LiDAR dataset containing approximately 80 million data points from four industrial scenes. Their dataset, captured using a Leica BLK360 scanner, achieved accuracies of 4 mm to 7 mm and included five categories of MEP components: pipes, pumps, tanks, I-shape, and rectangular beams. Agapaki and Brilakis [11] also introduce construction and MEP point cloud datasets: CLOI (Channels, L-shapes, circular sections, I-shapes). It is a large-scale repository of labelled point clusters specifically designed for industrial and MEP components. It contains over 140 million hand-labelled points from five industrial facilities, covering ten classes of the most frequent construction and MEP shapes. The dataset was collected using static terrestrial laser scanners, capturing the real-world noise and occlusions typical in complex industrial environments.

2.2 Challenges in MR-based Data Capture

Despite the potential of MR devices for on-site data capture, several challenges remain. Hübner et al. [12] evaluated the HoloLens (version 1) for indoor mapping applications, providing insights into its depth sensing capabilities and limitations. They found that the device's accuracy was highly dependent on distance and surface inclination, with optimal performance at distances less than 2.5 m and inclinations below 20 degrees. Beyond these limits, the sensor's accuracy deteriorated significantly, with noise exceeding 2 cm at greater distances or inclinations.

MR devices like the HoloLens produce point clouds that are significantly less dense and more prone to noise compared to traditional LiDAR scans. This is particularly problematic in MEP environments, where occlusions, reflective surfaces, and hard-to-reach installation locations can further degrade data quality.

The accuracy of MR-captured point clouds decreases with distance, which poses challenges for capturing tall or expansive spaces often encountered in industrial and commercial buildings. These limitations have implications for the applicability of existing deep learning models, which are typically trained on high-quality LiDAR datasets, as discussed in Section 2.1.

2.3 Point cloud Classification

Point cloud classification has advanced significantly with deep learning techniques. PointNet [13] pioneered direct processing of unordered point sets using MLPs but struggled with capturing local features. PointNet++ [14] improved upon this by introducing hierarchical feature learning to model local geometric structures and handle noise.

Recent works have built on this foundation with various improvements. PointNeXt [15] modernises PointNet++ with better training strategies, while PointMetaBase [16] introduces a modular meta-architecture to optimize computational efficiency. PointVector [17] enhances local feature learning through vector-based representations.

Graph-based models like DGCNN [18] apply dynamic edge convolutions to capture local relationships and adapt to structural changes, which is especially useful for noisy MR data. DeepGCNs [19] extend this concept with deeper architectures and residual connections for more complex feature hierarchies.

Attention mechanisms, such as those used in Point Transformer [20], help models focus on the most informative regions of sparse or noisy point clouds, improving robustness.

Further, PointMLP [21] simplifies the architecture using residual MLPs while achieving strong performance on real-world datasets like ScanObjectNN, making it suitable

for MR data.

Finally, Point-NN [22] introduces a non-parametric approach using simple operations (FPS, k-NN, pooling) with trigonometric spatial encoding. This method excels in MR environments due to its inherent robustness to varying point densities and distributions without requiring trainable parameters.

2.4 Gap in Existing Research

While there has been significant progress in both point cloud analysis and the application of MR in construction, there remains a critical gap in datasets with characteristics representative of data acquisition using typical MR-mounted sensing solutions, in particular to support applications related to MEP mapping. Existing datasets predominantly focus on high-precision LiDAR scans, which do not adequately capture the unique challenges posed by MR-based data collection. The sparse, noisy, and gap-prone nature of MR-captured meshes is not represented in these datasets. There is a need for datasets that reflect the real-world conditions encountered in MR-based inspections, including lower point density, increased noise, and more frequent occlusions due to the limitations of MR devices and the challenging nature of MEP environments.

3 Dataset Acquisition and Preparation

3.1 Research Framework

To clearly present the overall workflow of this research, we summarise the main steps, objectives, methods, and equipment involved:

- **1. MR Data Acquisition**
  Objective: Capture raw 3D spatial meshes of real-world MEP systems (pipes and ducts) under realistic MR conditions, while managing challenges such as noise and sparsity.
  Device: Microsoft HoloLens 2
  Method: On-site scanning of building environments with multi-angle coverage and careful motion control. The operator tilted the device upwards to capture ceiling-mounted MEP components, ensuring maximum possible coverage.

- **2. Dataset Preparation**
  Objective: Generate a structured and labelled MR point cloud dataset (MR-MEP) suitable for training and evaluating deep learning models.
  Process: Convert raw spatial meshes to point clouds, clean the data (removing large architectural elements and noise), segment MEP components using BIM-derived Oriented Bounding Boxes (OBBs), label points as 'pipe' or 'duct', and apply data augmentation to balance class distribution.

- **3. Model Benchmarking**
  Objective: Assess and compare the classification performance and computational demands of state-of-the-art models on MR-captured MEP data.
  Models: Nine deep learning architectures, including PointNet, PointNet++, DGCNN, PointMLP, PointNeXt, PointTransformer, PointMetaBase, PointVector, and PointNN.
  Method: Perform 6-fold cross-validation with scene-based splits, applying pre-training on ScanObjectNN where applicable and balancing training data with augmented samples.
  Evaluation Metrics: Overall Accuracy (OA), Mean Class Accuracy (mAcc), class-specific accuracy (pipes and ducts), and computational efficiency (parameters, FLOPs, throughput).

- **4. Performance Analysis**
  Objective: Determine the optimal model for practical MR-based MEP classification, balancing high accuracy with computational efficiency for real-world deployment, including potential on-device processing.
  Process: Analyse cross-validation results, evaluate accuracy and stability, and compare computational efficiency. Identify models most suitable for MR data constraints, such as sparsity and noise.

3.2 Scanning Procedure and Protocols

The Microsoft HoloLens 2 MR headset was selected as the primary data collection device for this study, owing to its sophisticated sensing and processing capabilities [23]. This advanced device incorporates a comprehensive sensor array specifically designed for spatial mapping. Hübner et al. [12] conducted an extensive evaluation of the HoloLens 2 for indoor mapping tasks, revealing a correlation between sensor-to-object distance and measurement accuracy. Their findings indicate that as the distance between the sensor and the object increases, the associated measurement noise also tends to rise. Specifically, for distances under 2.5 m, the noise level remains below 5 mm. However, beyond this threshold, the error exhibits a rapid escalation, with noise levels reaching 2 cm or more at greater distances.

In our study reported here, the scene mapping functionality of the HoloLens 2 (HL2) was utilised to capture the environment as a triangle mesh. The mesh resolution was set to the finest level, and a comprehensive scan of the environment was conducted. Given that the HL2's capture sensors are front-facing, it was necessary to tilt the device upwards to scan certain areas. Additionally, since some Mechanical, Electrical, and Plumbing (MEP) components were located on the ceiling, which exceeded a height of 2.5 m, at times the device had to be handheld at arm's length above the operator's head, with both hands gripping the sides of the headset for stability. The camera was oriented upward to directly face the MEP components,

and the operator maintained a steady, controlled motion with movement speed kept below approximately 0.3 meters per second to ensure complete coverage.

Data acquisition was conducted across 8 distinct building environments to ensure a diverse and representative dataset. A standardised scanning protocol was implemented to maintain consistency and data quality across all sites. This protocol comprised a preliminary site walkthrough to identify and delineate key MEP areas of interest, followed by systematic and comprehensive scanning of the identified MEP areas. Particular attention was given to capturing each component thoroughly from multiple angles to optimise coverage. The duration of each scanning session was approximately 15-20 minutes, with variations depending on the complexity and scale of the individual construction site.

3.3 Dataset Overview and Statistics

The dataset primarily focuses on MEP components like pipes and ducts, featuring both mixed reality captured environments and their corresponding BIM models. Data were collected from eight distinct scenes, six of which are classified as mixed environments containing both pipes and ducts: bike parking, apartment room, Garage 1, Garage 2, Garage 5, and Garage 6. The remaining two scenes, Garage 3 and Garage 4, are pipe-only environments.

To establish the ground truth, BIM was employed to generate precise digital representations from laser scan data acquisition of those environments.

3.4 Dataset Preparation for Point Cloud Classification Use Case

In the context of scan-vs-BIM comparison for automated construction progress monitoring, we developed a component-wise detection strategy utilising oriented bounding boxes (OBBs) derived from the corresponding BIM elements. Similarly to [24, 25], the idea is to explore whether the point cloud in the vicinity of each BIM object confirms the presence of that element. This systematic decomposition approach resulted in 382 distinct OBB regions across all environments, where each OBB region encompasses point cloud data that potentially corresponds to a specific MEP component segment.

Table 1 presents the key characteristics of the eight scenes in our dataset, detailing the number of vertices, segmented pipes, segmented ducts, and the dimensions of OBB enclosing each MEP component in each scene.

3.5 Class Balancing Strategy

Analysis of the dataset composition reveals a significant class imbalance between ducts and pipes. Of the total 382 elements, 44 (11.5%) represent ducts, while the remaining 338 (88.5%) correspond to pipes. This imbalance is arguably representative of the reality of building environments [26], where piping networks often dominate due to their essential role in various building services including water supply, drainage, heating, cooling, and fire protection systems.

To mitigate the issue of class imbalance, a data augmentation pipeline consisting of rotation, scaling, and mirroring was applied to the duct class within the mixed scenes. This augmentation process was repeated until a 1:1 ratio between duct and pipe samples was achieved, ensuring a balanced dataset for model training.

3.6 Cross-Validation Strategy

A 6-fold cross-validation strategy was employed to robustly evaluate the performance of the selected and deployed classification models. Each fold ensures that every mixed scene serves as the test set once, while the model is evaluated on a combination of pipes and ducts. During training, the model is exposed to both mixed scenes and pipe-only scenes.

The cross-validation folds were structured such that in each iteration, one scene was held out for testing while all others were used for training. The test scenes for the six folds were: Bike Parking, Apartment Room, Garage 1, Garage 2, Garage 5, and Garage 6, respectively.

4 Methods and Implementation

4.1 Selected Deep Learning Models

We select a diverse set of classification models, including traditional deep learning architectures, graph-based methods, transformer-inspired models, and a non-parametric method. This range enables an assessment of trade-offs between model complexity, computational efficiency, and performance, particularly for point cloud processing tasks involving noisy or sparse data captured by MR devices. These models have been pre-trained on ScanObjectNN, which closely resembles Mixed Reality captured data as discussed in Section section 2.1, making them particularly suitable for real-world MR applications. The selected architectures include:

- **PointNet** [13] serves as a fundamental architecture for point cloud classification and segmentation tasks.
- **PointNet++** [14] builds on PointNet and introduces hierarchical feature learning through Set Abstraction modules, enhancing the model's capability to capture local structures.
- **DGCNN** [18] takes a graph-based approach, implementing dynamic edge convolution to address noise and sparsity challenges by effectively leveraging local neighborhood structures.
- **PointMLP** [21] advances the field by implementing a sophisticated residual MLP framework designed for robust performance.

Scene	Vertices	Pipes	Ducts	Size (m^2)
Bike Parking	32,962	51	3	10.4 × 8.6
Apartment Room	82,139	96	11	12.5 × 10.8
Garage 1	95,416	24	7	18.6 × 14.8
Garage 2	85,814	40	4	20.9 × 10.9
Garage 3	47,177	67	0	12.2 × 4.3
Garage 4	14,029	24	0	13.2 × 8.0
Garage 5	35,967	3	5	12.7 × 3.9
Garage 6	39,447	33	14	30.5 × 15.9
Total	432,951	338	44	131.0 × 77.2

Table 1. Scene-level characteristics of the captured environments

- **PointNeXt** [15] revisits the PointNet++ architecture, incorporating enhanced training strategies and introducing a more scalable model design.
- **PointMetaBase** [16] offers a unified analysis framework that enables efficient point cloud processing.
- **PointVector** [17], which employs vector-based processing techniques to capture spatial relationships within point clouds;
- **PointTransformer** [20], which adapts transformer principles to emphasise spatial relationships in point cloud data;
- **Point-NN** [22] takes a unique approach as a non-parametric, training-free method that utilises a point-memory bank for similarity matching.

4.2 Evaluation Metrics

We employ a diverse set of metrics across six-fold validation to rigorously assess model performance and efficiency.

The evaluation of classification accuracy encompasses three distinct metrics:

- **Class-specific Accuracy (Acc)** quantifying the classification accuracy for pipe and ducts independently, enabling granular performance assessment and identification of potential classification biases.
- **Overall Accuracy (OA)** quantifies the proportion of correctly classified samples across the entire dataset, providing a global performance indicator.
- **Mean Class Accuracy (mAcc)** averages individual class accuracies, thereby ensuring equitable evaluation across all categories, regardless of class imbalance issues.

The evaluation of computational efficiency also comprises three fundamental metrics:

- **M, the total number of trainable parameters** quantifying the model complexity and expressed in millions.
- **G, the FLOPs** quantifying the computational demand during inference and expressed in billions of floating-point operations.
- **Throughput measurements** quantifying the system's processing speed by measuring the number of complete inference requests the model can process per second under realistic deployment conditions.

For statistical robustness, we report the mean and standard deviation of each accuracy and efficiency metrics across all six folds, thereby providing comprehensive insights into both performance magnitude and consistency.

4.3 Hardware and Software Specifications

The experiments were conducted on a system equipped with an NVIDIA GeForce RTX 3090 GPU and an AMD Ryzen 9 3900X 12-core processor. The software environment included Ubuntu 20.04.6 LTS as the operating system. The deep learning framework used was PyTorch 2.0.0, which was integrated with CUDA 11.8 for GPU acceleration. Python version 3.9.0 was used as the programming language to run the experiments.

5 Results and Discussion

The experimental results are presented in table 2 and table 3 that provide a comprehensive evaluation of the selected models across both accuracy and computational efficiency metrics. Table 2 summarises the cross-validation accruacy results.Among all models, PointNN demonstrates clear superior performance, achieving by far the highest scores across all accuracy metrics (90.21% overall accuracy), and with lower standard deviations, indicating more stable performance across different test scenes. Table 3 presents the computational efficiency metrics.The results show that PointNN achieves optimal efficiency with minimal parameters (~0.0M[1]) and highest processing speed (11,932.60 instances per second), while PointMetaBase requires the least computational operations at 0.55G FLOPs, demonstrating its theoretical computational efficiency.

To visualise the trade-off between model performance and computational demands, fig. 1 plots the Overall Accuracy against FLOPs for all evaluated models. The figure clearly demonstrates that PointNN (marked with a star) achieves an optimal balance, reaching the highest accuracy

[1]because it is a non-parametric method

Model	Overall Accuracy (%)	Mean Class Accuracy (%)	Pipe (%)	Duct (%)
PointNN	**90.21 ± 2.12**	**89.45 ± 2.34**	**91.33 ± 2.45**	**85.67 ± 3.56**
PointMLP	81.55 ± 8.23	74.00 ± 12.45	67.76 ± 37.91	80.23 ± 21.54
PointNet++	80.85 ± 13.21	78.22 ± 13.45	77.42 ± 23.45	79.02 ± 24.12
PointTransformer	80.52 ± 13.45	77.49 ± 14.56	89.20 ± 11.23	75.78 ± 17.89
PointNext	80.45 ± 14.23	75.37 ± 17.34	80.53 ± 34.56	70.22 ± 31.23
DGCNN	78.57 ± 7.89	72.36 ± 9.21	70.44 ± 31.82	74.28 ± 25.85
PointNet	78.43 ± 8.56	70.24 ± 9.78	62.80 ± 33.45	77.68 ± 24.56
PointMetaBase	78.35 ± 13.65	74.79 ± 11.83	81.80 ± 18.32	67.78 ± 32.41
PointVector	68.79 ± 15.67	63.79 ± 12.45	71.85 ± 27.89	55.74 ± 37.89

Table 2. Cross-validation Results (Mean ± Std)

Model	Parameters (M)	FLOPs (G)	Throughput (ins./sec)
PointNN	**~0.0**	1.34	**11,932.60**
PointMetaBase	1.417	**0.55**	2,313.80
PointNext	1.364	1.64	2,776.64
PointNet++	1.463	1.71	2,373.52
PointVector	1.544	1.68	1,069.91
DGCNN	1.800	4.82	802.59
PointNet	3.470	0.91	5,620.63
PointMLP	13.228	31.35	199.37
PointTransforme	22.647	29.16	432.74

Table 3. Computational Efficiency Metrics

while maintaining moderate computational complexity. Most deep learning models cluster in the middle range of both metrics, with FLOPs between 0.5G and 5G and accuracies between 75% and 85%. Notably, more complex models like PointMLP and PointTransformer, despite their substantially higher computational demands (>25G FLOPs), do not yield proportional improvements in accuracy.

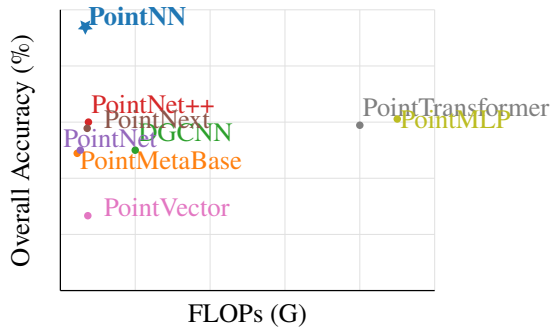


Figure 1. Model accuracy versus computational complexity (FLOPs). PointNN (marked with a star) achieves the best accuracy while maintaining minimal computational requirements.

5.1 Performance Analysis

PointNN achieves superior performance across all accuracy metrics, with an overall accuracy of 90.21% (±2.12%) and mean class accuracy of 89.45% (±2.34%). PointNN's consistently low standard deviations (±2.12% to ±3.56% across different metrics) strongly suggest superior generalization capability and resistance to overfitting, particularly valuable for MR applications where data distributions can vary significantly between deployments. Most notably, it maintains remarkably consistent performance across both Pipe (91.33% ±2.45%) and Duct (85.67% ±3.56%) classes, indicating stable learning across different geometric features.

The second tier of performers exhibits concerning signs of potential overfitting, as evidenced by their high cross-validation standard deviations. PointMLP (±8.23%), PointNet++ (±13.21%), PointTransformer (±13.45%), and PointNext (±14.23%) show comparable mean accuracy but their high variances across different folds suggest that these models may be overspecialising to training set characteristics rather than learning generalisable features, with their class-specific performance showing dramatic variations (up to ±37.91% for PointMLP's pipe classification) indicating inconsistent feature learning.

More traditional architectures like DGCNN (78.57% ±7.89%) and the original PointNet (78.43% ±8.56%) demonstrate moderate performance with relatively lower cross-validation standard deviations compared to the second tier. While their balanced standard deviations suggest more stable learning, their lower overall accuracy makes them less suitable for precision-critical MR applications. These models' moderate standard deviations indicate better generalisation than more complex models but still fall short of PointNN's stability.

From a computational perspective, PointNN's perfor-

mance is particularly noteworthy for MR applications, given its minimal parameter count (~0.0M) and high throughput (11,932.60 instances/second). The absence of learnable parameters inherently prevents traditional overfitting, as reflected in the consistent cross-validation results. This stands in stark contrast to more complex models like Point-Transformer and PointMLP, which require substantially more parameters (22.647M and 13.228M respectively) and show signs of overfitting in their high standard deviations despite achieving reasonable mean accuracy.

Overall, the non-parametric approach of PointNN provides several key advantages for MR point cloud processing: First, the trigonometric encoding effectively captures high-frequency geometric patterns, which remain stable even when surface details are inconsistently captured by MR devices. Second, the point-memory bank's similarity-based classification provides flexible matching rather than rigid decision boundaries, crucial for handling varying point densities and object appearances. Third, PointNN shows remarkable immunity to common MR data challenges including partial occlusions and sensor noise through its local geometry aggregation mechanism. Combined with its computational efficiency and inherent regularization from the non-parametric design, these characteristics make PointNN especially valuable for real-world MR deployments where robustness, consistency, and real-time processing are critical requirements.

6 Conclusion

This work presented a new dataset MR-MEP and an investigation into the performance of point cloud classification deep learning models on that type of data. Specifically, we present MR-MEP, a specialised dataset that pairs mixed reality captured point clouds with corresponding BIM models for MEP components in building environments. The dataset was collected using Microsoft HL2 across 8 diverse building scenes, containing 432, 951 vertices spanning a total area of 131.0×77.2 m^2. What distinguishes MR-MEP is its dual-source nature - combining precisely captured mixed reality point cloud data with corresponding BIM models - making it valuable for multiple point cloud understanding tasks including classification, object detection, and semantic/instance segmentation. For our specific investigation into MEP component classification, the dataset is split into 382 annotated MEP components (338 pipes and 44 ducts), segmented using BIM-derived oriented bounding boxes (OBBs). To ensure robust evaluation, we implemented a 6-fold cross-validation strategy and addressed the natural class imbalance through careful data augmentation.

Our experimental analysis of the different point cloud classification models demonstrates that the recently proposed non-parametric approach, PointNN, achieves state-of-the-art performance with 90.21% classification accuracy while maintaining minimal computational overhead of 1.34 GFLOPs. The model's ability to maintain high accuracy despite the quality limitations of MR-captured data suggests potential for robust deployment in real-world construction scenarios. Moreover, the minimal computational requirements make it particularly suitable for on-device processing in MR devices, addressing a critical need in mobile construction inspection applications.

Although the proposed approach demonstrates promising results in MEP component classification from MR-acquired point cloud data, several critical areas merit further investigation. The current implementation, while effective, has been primarily evaluated on a binary classification task distinguishing between pipes and ducts. This should be extended to support multi-class classification capabilities, in order to accommodate the diverse range of MEP components encountered in real-world construction environments. Additionally, while our work focuses on classification, further evaluation across other crucial point cloud understanding tasks is needed. For instance, in real construction scenarios, MEP components need to be precisely localized through object detection and semantically segmented for detailed analysis. For such tasks, while Point-NN's non-parametric approach demonstrates strong performance in classification, its effectiveness in more complex scenarios like instance or semantic segmentation remains to be thoroughly investigated. It is possible that other architectures with learned parameters (including the *Point-PN* approach proposed by the same authors [22]) might prove more suitable for tasks requiring fine-grained spatial understanding or complex feature hierarchies, as these tasks may benefit from learned, task-specific feature representations. To bridge the gap between theoretical advances and practical implementation, future work should focus on integrating deep learning-based detection and classification models like those into comprehensive MEP construction monitoring frameworks such as the one proposed by [24]. Such integration presents several technical challenges that warrant investigation, including: the optimisation of real-time processing capabilities within MR-based progress monitoring systems, and the development of robust verification mechanisms for construction geometry inspection.

References

[1] Jun Wang, Xiangyu Wang, Wenchi Shou, Heap-Yih Chong, and Jun Guo. Building information modeling-based integration of mep layout designs and constructability. *Automation in Construction*, 61:134–146, 2016. ISSN 0926-5805. doi:10.1016/j.autcon.2015.10.003.

[2] Ahlem Assila, Amira Dhouib, Ziad Monla, and Mourad Zghal. Integration of augmented, virtual and mixed reality with building information modeling: A systematic review. In Jessie Y. C. Chen and Gino Fragomeni, editors, *Virtual, Augmented and Mixed Reality: Design and Development*, pages 3–19, Cham, 2022. Springer International Publishing.

ISBN 978-3-031-05939-1. doi:10.1007/978-3-031-05939-1_1.

[3] Hongzhe Yue, Qian Wang, Hongxiang Zhao, Ningshuang Zeng, and Yi Tan. Deep learning applications for point clouds in the construction industry. *Automation in Construction*, 168:105769, 2024. ISSN 0926-5805. doi:10.1016/j.autcon.2024.105769.

[4] Apple Inc. *RoomPlan*, 2024. URL `https://developer.apple.com/documentation/roomplan`. Accessed: 2024-11-09.

[5] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3D ShapeNets: A deep representation for volumetric shapes . In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1912–1920, Los Alamitos, CA, USA, June 2015. IEEE Computer Society. doi:10.1109/CVPR.2015.7298801.

[6] Angel X. Chang, Thomas A. Funkhouser, Leonidas J. Guibas, Pat Hanrahan, Qi-Xing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiong Xiao, L. Yi, and Fisher Yu. Shapenet: An information-rich 3d model repository. *ArXiv*, abs/1512.03012, 2015.

[7] Mikaela Angelina Uy, Quang-Hieu Pham, Binh-Son Hua, Thanh Nguyen, and Sai-Kit Yeung. Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1588–1597, 2019. doi:10.1109/ICCV.2019.00167.

[8] Iro Armeni, Ozan Sener, Amir R. Zamir, Helen Jiang, Ioannis Brilakis, Martin Fischer, and Silvio Savarese. 3d semantic parsing of large-scale indoor spaces. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1534–1543, 2016. doi:10.1109/CVPR.2016.170.

[9] Xavier Roynard, Jean-Emmanuel Deschaud, and François Goulette. Paris-lille-3d: A large and high-quality ground-truth urban point cloud dataset for automatic segmentation and classification. *Int. J. Rob. Res.*, 37(6):545–557, May 2018. ISSN 0278-3649. doi:10.1177/0278364918767506.

[10] Chao Yin, Boyu Wang, Vincent J.L. Gan, Mingzhu Wang, and Jack C.P. Cheng. Automated semantic segmentation of industrial point clouds using respointnet++. *Automation in Construction*, 130:103874, 2021. ISSN 0926-5805. doi:https://doi.org/10.1016/j.autcon.2021.103874.

[11] Eva Agapaki and Ioannis Brilakis. Cloi-net: Class segmentation of industrial facilities' point cloud datasets. *Advanced Engineering Informatics*, 45:101121, 2020. ISSN 1474-0346. doi:https://doi.org/10.1016/j.aei.2020.101121.

[12] Patrick Hübner, Kate Clintworth, Qingyi Liu, Martin Weinmann, and Sven Wursthorn. Evaluation of hololens tracking and depth sensing for indoor mapping applications. *Sensors*, 20(4):1021, 2020. doi:10.3390/s20041021.

[13] R. Qi Charles, Hao Su, Mo Kaichun, and Leonidas J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 77–85, 2017. doi:10.1109/CVPR.2017.16.

[14] Charles R. Qi, Li Yi, Hao Su, and Leonidas J. Guibas. Pointnet++: deep hierarchical feature learning on point sets in a metric space. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 5105–5114, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.

[15] Guocheng Qian, Yuchen Li, Houwen Peng, Jinjie Mai, Hasan Abed Al Kader Hammoud, Mohamed Elhoseiny, and Bernard Ghanem. Pointnext: revisiting pointnet++ with improved training and scaling strategies. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22, Red Hook, NY, USA, 2024. Curran Associates Inc. ISBN 9781713871088. doi:10.5555/3600270.3601955.

[16] Haojia Lin, Xiawu Zheng, Lijiang Li, Fei Chao, Shanshan Wang, Yan Wang, Yonghong Tian, and Rongrong Ji. Meta Architecture for Point Cloud Analysis . In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17682–17691, Los Alamitos, CA, USA, 2023. IEEE Computer Society. doi:10.1109/CVPR52729.2023.01696.

[17] Xin Deng, WenYu Zhang, Qing Ding, and XinMing Zhang. Pointvector: A vector representation in point cloud analysis. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9455–9465, 2023. doi:10.1109/CVPR52729.2023.00912.

[18] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E. Sarma, Michael M. Bronstein, and Justin M. Solomon. Dynamic graph cnn for learning on point clouds. *ACM Trans. Graph.*, 38(5), October 2019. ISSN 0730-0301. doi:10.1145/3326362.

[19] Guohao Li, Matthias Muller, Ali Thabet, and Bernard Ghanem. DeepGCNs: Can GCNs Go As Deep As CNNs? . In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9266–9275, Los Alamitos, CA, USA, November 2019. IEEE Computer Society. doi:10.1109/ICCV.2019.00936.

[20] Hengshuang Zhao, Li Jiang, Jiaya Jia, Philip Torr, and Vladlen Koltun. Point transformer. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 16239–16248, 2021. doi:10.1109/ICCV48922.2021.01595.

[21] Xu Ma, Can Qin, Haoxuan You, Haoxi Ran, and Yun Fu. Rethinking network design and local geometry in point cloud: A simple residual mlp framework, 2022.

[22] Renrui Zhang, Liuhui Wang, Yali Wang, Peng Gao, Hongsheng Li, and Jianbo Shi. Starting from non-parametric networks for 3d point cloud analysis. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5344–5353, 06 2023. doi:10.1109/CVPR52729.2023.00517.

[23] Dorin Ungureanu, Federica Bogo, Silvano Galliani, Pooja Sama, Xin Duan, Casey Meekhof, Jan Stühmer, Thomas J. Cashman, Bugra Tekin, Johannes L. Schönberger, Pawel Olszta, and Marc Pollefeys. Hololens 2 research mode as a tool for computer vision research. *CoRR*, abs/2008.11239, 2020. doi:10.48550/arXiv.2008.11239.

[24] Boan Tao, Frédéric Bosché, and Jiajun Li. Mixed reality-based mep construction progress monitoring: Evaluation of methods for mesh-to-mesh comparison. *Automation in Construction*, 168:105852, 2024. ISSN 0926-5805. doi:10.1016/j.autcon.2024.105852.

[25] Zhiqi Hu and Ioannis Brilakis. Matching design-intent planar, curved, and linear structural instances in point clouds. *Automation in Construction*, 158:105219, 2024. ISSN 0926-5805. doi:https://doi.org/10.1016/j.autcon.2023.105219.

[26] Viktor Drobnyi, Zhiqi Hu, Yasmin Fathy, and Ioannis Brilakis. Construction and maintenance of building geometric digital twins: State of the art review. *Sensors*, 23 (9):4382, 2023. doi:10.3390/s23094382.

Online data acquisition and building modelling for building renovation

Danya Liu[1], Kepa Iturralde[1,2] and Christoph Holst[1]

[1]Chair of Engineering Geodesy, School of Engineering and Design,
Technical University of Munich, 80333 Munich, Germany
[2]Chair of Digital Transformation in Construction, Institute of Construction Management,
Faculty of Civil and Environmental Engineering, University of Stuttgart, 70569 Stuttgart, Germany

danya.liu@tum.de, kepa.iturralde@ibl.uni-stuttgart.de, christoph.holst@tum.de

Abstract -

The issue of renovating a significant number of buildings has emerged as a prominent concern within the architecture field in recent years. The renovation of existing buildings represents a significant opportunity for reducing energy consumption and greenhouse gas emissions at minimal economic cost. However, achieving effective renovation outcomes can be challenging, particularly for customers who lack technical expertise. To address this gap, this paper presents an online data acquisition platform designed to streamline the integration of building information through a standardized process. Furthermore, this paper presents the optimization of algorithms to address common challenges such as fragmentation, an over-reliance on manual input, and the utilization of disconnected tools in the renovation process. Finally, the feasibility and effectiveness of the platform are demonstrated through case studies featuring 3D building models processed through the platform. These case studies highlight the potential of the platform as a valuable tool for building renovation.

Keywords -

Data acquisition; renovation; energy saving; modeling.

1 Introduction

Since the intensification of extreme weather and climate events resulting from global climate change represents an escalating risk to human society [1][2][3]. Over the past two decades, greenhouse gas emissions have emerged as a primary driver of the climate crisis [4]. The construction industry, in particular, has emerged as a significant contributor to both greenhouse gas emissions and energy consumption [5], thereby exacerbating environmental concerns. As the urgency to mitigate these issues grows, the need for innovative solutions that enhance energy efficiency in the built environment becomes increasingly evident.

In response to this pressing need, this study proposes an online platform aimed at improving energy efficiency and standardizing renovation practices in the construction industry. The prevailing renovation methodologies are characterized by fragmentation, an over-reliance on manual input, and the utilization of disconnected tools. This results in inefficiencies and suboptimal energy assessments. To address these challenges, the proposed platform implements a semi-automated, structured workflow to ensure consistency and accuracy throughout the renovation process.

The platform operates by guiding users through a systematic process, beginning with the input of basic building data, which generates an initial model devoid of openings. Subsequent processing and integration of existing openings generates a comprehensive and normalized model. This structured approach not only eliminates errors but also streamlines energy assessments, improving the efficiency and reliability of renovation projects.

This paper commences with an overview of the prevailing circumstances of the renovation of buildings. Thereafter, the functionality of the online platform is elucidated, whereby all processes are rendered uniform. The processes include the user input of existing basic building information, the creation of an initial simple model based on the user input (without any openings on the walls), the pre-processing of the existing openings information, the combination of the simple model with the existing information, and the generation of the final model, normalization process of the obtained model. Subsequently, the optimization section, which encompasses several technically challenging elements, is addressed. This study presents a series of case studies based on the developed platform. Concurrently, a set of criteria is provided to complete the evaluation of the platform. The findings indicate that the online platform is both effective and straightforward to operate, thereby potentially streamlining the process of building renovation.

2 Related work

2.1 Existing platforms for building renovation

As previously stated in the introduction, the integral functional data acquisition platform designed in this

project is not based on any other existing platforms. However, there are some existing renovation systems proposed by other researchers.

For instance, the RINNO project [6] provides a comprehensive solution for the full-lifecycle management of renovations, encompassing the entire project from initial planning and design to the end-of-life phases. The platform incorporates advanced analytical capabilities, interoperability with BIM [7], and a repository for renovation resources, thereby optimizing time, costs, and energy efficiency while enhancing stakeholder collaboration.

The SUNShine platform [8] is an innovative digital tool designed to enhance efficiency, transparency, and standardization in the deep renovation of multi-family buildings. By facilitating effective communication among diverse stakeholders throughout the renovation process, the platform streamlines project management and reduces costs through digitalized and standardized documentation.

It is evident that the majority of contemporary research endeavors are oriented towards the development of a comprehensive renovation platform. However, there is a notable absence of a standardized platform for data acquisition.

2.2 Challenges in building renovation

A substantial body of evidence [9] indicates that the rate of building renovation in the Europe Union (EU) is relatively low, falling far short of the EU's stated targets [10][11][12][13] for such work. Nevertheless, the research [14] indicates that accelerating the renovation of buildings in the near term is a formidable challenge. In addition to identifying the optimal channels for reaching both the demand and supply sides of renovation technology, it is crucial to assess the user-friendliness and efficiency of these channels.

The challenge mentioned above was also proposed by other researchers [15]. In order to achieve the implementation of Nearly Zero-Energy Buildings (NZEB) [16] renovation packages in Europe, it is necessary to shift building renovation strategies towards solutions that incorporate a multitude of passive and active components. This will inevitably result in an increase in the complexity and costs associated with the execution of such strategies.

Besides, in comparison to new construction projects, renovations of existing buildings present a number of challenges, including high design difficulty, demanding quality and safety management, complex construction processes, poor site conditions, and dynamic project management requirements. The design phase of existing building renovation projects is often lengthy and inefficient [17].

Based on these existing challenges, a fully functional online data acquisition platform is required in order to effectively increase the renovation rate of buildings and reduce energy loss.

3 Methodology

3.1 The stepwise data acquisition platform

The data acquisition process comprises several regularized phases. During this process,

1. the user is required to input the relevant information pertaining to the target building, as illustrated in Figure 1.
2. the creation of an initial simple model based on the user input, which does not include any openings (windows, balconies, doors...) on the facades.
3. the pre-processing of the existing openings information. The openings information here can be generated by another platform, which is the culmination of our research group's efforts [18]. Its primary objective is to facilitate the selection of openings on a facade from images captured from disparate viewpoints, subsequently storing them in a normalized `.json` file.
4. the combination of the model with the processed existing openings information.
5. the generation of the final model in the form of a formatted `.json` file, which can be utilized to visualize the comprehensive model with all the openings clearly delineated. This format is particularly well-suited for 3D modeling and BIM applications, offering a streamlined process for architects and designers [19][20][21][22][23][24].
6. the normalization of the obtained final model for being transferred into an `.ifc` file.

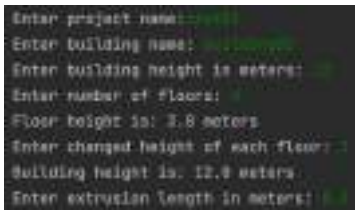


Figure 1. The illustration of user input

3.2 The illustration of the data acquisition process

The provided flowchart illustrated in Figure 2 presents a comprehensive overview of the data acquisition platform, delineating the sequence of operations and offering a more intuitive representation of its efficacy.

4 Optimization

As previously stated in the section on related work, the current challenges make the renovation process a signifi-

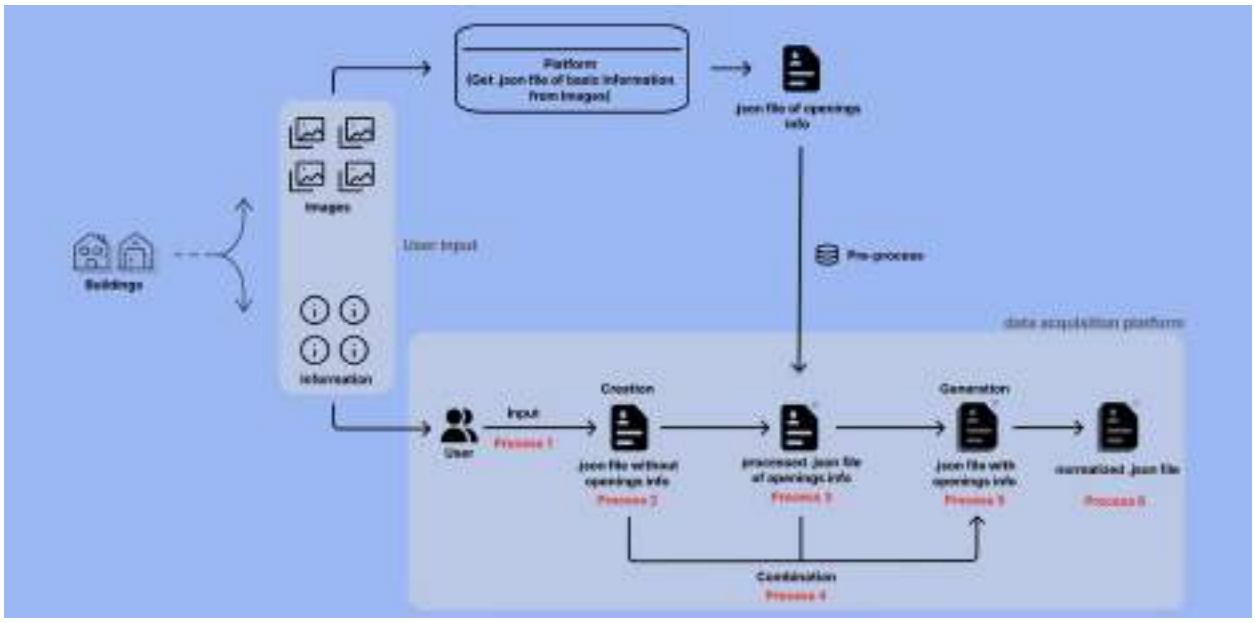


Figure 2. The flowchart of the designed platform

cant undertaking. To surmount these obstacles, it is imperative to augment the efficacy and sophistication of the data acquisition methodology, thereby paving the way for the comprehensive automation of the operational procedure. It is of paramount importance to ensure that the generated model is as close as possible to the actual building in the real world. In light of the aforementioned considerations, a number of optimizations can be implemented with a view to enhancing the platform's automation capabilities, the precision and reliability of the model simulation.

4.1 Roof implementation

The majority of contemporary algorithms fail to consider the impact of a building's roof on its renovation. This is an important shortcoming, as it results in the generation of buildings that bear little resemblance to those found in the real world. The data acquisition platform proposed in this paper addresses this issue by allowing the addition of a variety of roof styles to buildings.

In order to guarantee that the roof appears as an organic extension of the building, it is essential to ensure that there are no gaps between the facades and roof elements. The roof is therefore treated as an additional layer, seamlessly integrated with the rest of the building's geometry. And the structure of the model should be like the structure shown in Figure 3.

4.2 Data filtering

The filtering function was developed with the objective of filtering and processing building data from an external



Figure 3. The structure of the building with a roof

`.json` file, incorporating user input for the targeted model. The process commences with the loading of data from a `.json` file, followed by the handling of any potential errors that may arise.

For example, in the majority of cases, only a single building will be processed; however, the `.json` file will contain a series of buildings. The user is prompted to select a specific building by its unique identifier, which is assumed to be an integer. This allows the program to filter the `.json` file and retain only the relevant building data. In the event that no matching building is identified, the script will notify the user and terminate without implementing any changes to the file.

To illustrate, if a `.json` file contains information about multiple buildings, but the user only wishes to process one (e.g., the building with ID= 3), the only thing required to do is to enter the ID of the desired building (3) in the user input field. The platform will then automatically filter out all other buildings with IDs other than 3. This

feature also facilitates greater automation of the platform. The illustration of this optimization is shown below in the Figure 4.

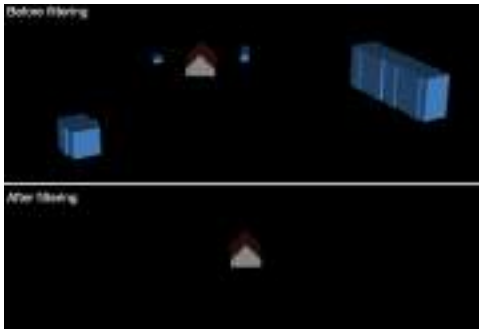


Figure 4. Before and after using the filter function

4.3 Data combination

As previously stated in the `Methodology` section, the file generated externally and the file obtained by user input must be combined. However, since the fundamental data regarding openings is not always consistent, the algorithm is optimized by loading the building data from two sources and identifying the pertinent facade objects (openings) and their geometries.

Subsequently, the data regarding the openings within the aforementioned facades is extracted and stored for subsequent utilization. By iterating through the building structures and their components, the script effectively isolates key data related to facades and their openings, which is essential for accurate building modeling and analysis.

In addition, this phase of the process entails the processing of facade names, whereby they are extracted from the building data and modified to ensure uniformity in format. This guarantees that walls from both data sources can be readily identified and matched, despite minor discrepancies in their nomenclature. And in the event of slight discrepancies in nomenclature between the two files, such as differences in case and initial count, it is imperative to normalize the naming conventions in order to ensure consistency. It plays a pivotal role in aligning the data between the two sources, thereby facilitating a seamless merging process.

4.4 Height estimation

In light of the fact that users are not always able to provide precise building heights, the height estimation was incorporated into the design of the platform as an additional feature. This results in a more precise final model, which the user can then utilise for the renovation of the building with the objective of reducing energy consumption.

The optimized algorithm based on [25] for estimating building height in the absence of external data, specifically addressing the scenario where the camera is angled relative to the building facade. The diagonal distance (D'), defined as the straight-line distance from the camera to the building base, serves as a preliminary reference point. However, it necessitates correction to account for perspective distortions induced by the camera's angle of observation. This is accomplished through the application of the formula:

$$D = D' \cdot \cos(\theta), \tag{1}$$

wherein θ represents the angle between the camera's line of sight and the perpendicular to the facade. The angle θ can be determined through image analysis or by reference to the camera setup data. The correction of D' to D ensures an accurate representation of the horizontal distance between the camera and the building base. With the corrected horizontal distance (D), the real-world height of the building (H_{building}) is calculated using the principles of projective geometry.

$$H_{\text{building}} = \frac{H_{\text{image}} \cdot D \cdot H_{\text{sensor}}}{f \cdot R_h}, \tag{2}$$

where:

- H_{building}: Real-world height of the building (in meters).
- H_{image}: Building height in the image (in pixels).
- D: Corrected horizontal distance from the camera to the building base.
- H_{sensor}: Physical height of the camera sensor (in millimeters).
- f: Camera focal length (in millimeters).
- R_h: Image resolution (vertical pixel count).

In light of the aforementioned details, the user is only obliged to input the letter D; the remainder of the information will be automatically detected and extracted by the platform. This has the potential to markedly enhance the precision and efficacy of our modelling procedures.

5 Case study - Pilot buildings

A number of pilot buildings have been selected for the purpose of evaluating the designed online data acquisition platform. The designed platform was subjected to an examination of its functionality on over 20 different buildings in various locations across Europe. In this section, two models, one relatively simple and the other more complex, have been presented for the purpose of demonstrating the feasibility of the data acquisition platform.

To be mentioned, the input of this platform should be from users as shown in Figure 1 and the output should be the final normalized .json file which can be transferred into the .ifc file for the visualization.

5.1 Pilot building 1 (6 facades)

Initiating the process, the first pilot building presented here is comprised of six facades, each exhibiting a limited number of openings. Subsequent to the data acquisition platform's processing, the resulting model, illustrated in Figure 5, was generated. And the approximate external shape of the model and the address information (Munich, Germany) are provided in the attachment to Figure 6.



Figure 5. Example of the simple building



Figure 6. Basic information of the simple building

5.2 Pilot building 2 (22 facades)

To illustrate the capacity of the designed platform to address the complexities inherent in more sophisticated buildings, an additional pilot building is presented here for evaluation. Given the 22 facades comprising this pilot building, a comprehensive comparison for each facade is not feasible. For purposes of reference, the images of this building can be accessed via this URL: `https://www.ensnare.eu/#pilotsites`. The illustration of this pilot building is shown below in Figure 7.



Figure 7. Example of the complicated building

6 Future work

6.1 More roof types

A significant area for future research is the expansion of the methodology to accommodate a broader range of roof types in building height estimation. Although the existing methodology is effective in handling standard roof designs, real-world applications frequently entail more intricate geometries, such as gable, hip, mansard, or even curved roofs. This improvement will facilitate the system's ability to process a more expansive range of building types, thereby enhancing its robustness and versatility in real-world scenarios.

For example, there are a number of different configurations as shown in Figure 8, including flat (or shed), gabled, hipped, arched, and domed. Additionally, there is a wide variety of other configurations. Copyright information can be found at this URL: `https://commons.wikimedia.org/wiki/File:Roof_diagram.jpg`

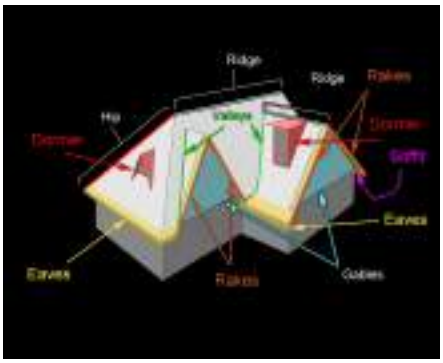


Figure 8. An illustration of different types of roofs

6.2 Multi-Platform Integration

Another crucial objective is the integration of the proposed platform with existing platform [18], with the aim of creating a comprehensive and unified system. Many platforms are currently designed for specific tasks in building renovation, such as initial modeling, data acquisition or simulation. However, they often operate independently. This integration will entail the harmonization of data formats, the optimization of interoperability between modules, and the implementation of automated workflows with the objective of minimizing manual intervention. Such a unified system will not only enhance usability for non-expert users but also provide a seamless end-to-end solution for building renovation projects, aligning with the prevailing industry trends toward automation and digitization.

7 Conclusion

The designed data acquisition and building modeling platform represents a substantial advancement in the ability to address the complexities inherent to building renovation projects. The platform's modular design incorporates features such as standardized data input, pre-processing of structural details, the generation of detailed 3D models and so on, thereby ensuring accuracy and efficiency across renovation workflows. Furthermore, the optimization of key algorithms addresses common technical challenges, thereby enhancing the platform's performance and reliability.

Besides, the feasibility and effectiveness of the platform are demonstrated by the generation of accurate three-dimensional pilot building models, which validate its capability to facilitate complex renovation tasks. The platform not only simplifies data workflows but also enhances the quality of decision-making and project outcomes. Its innovative approach provides a robust foundation for future enhancements, including the incorporation of advanced modeling for diverse building types and integration with comprehensive renovation management systems. In conclusion, the platform represents a crucial instrument for modernizing and simplifying the building renovation process, in accordance with the increasing demand for digital transformation in the fields of architecture and construction.

References

[1] Haolan Liao and Qianwang Deng. Ees-eoq model with uncertain acquisition quantity and market demand in dedicated or combined remanufacturing systems. *Applied Mathematical Modelling*, 64:135–167, 2018.

[2] Neng Shen, Haolan Liao, Rumeng Deng, and Qunwei Wang. Different types of environmental regulations and the heterogeneous influence on the environmental total factor productivity: empirical analysis of china's industry. *Journal of Cleaner Production*, 211:171–184, 2019.

[3] Haolan Liao, Qingyu Zhang, and Lu Li. Optimal procurement strategy for multi-echelon remanufacturing systems under quality uncertainty. *Transportation Research Part E: Logistics and Transportation Review*, 170:103023, 2023.

[4] Haolan Liao, Rong Ren, and Lu Li. Existing building renovation: a review of barriers to economic and environmental benefits. *International Journal of Environmental Research and Public Health*, 20 (5):4058, 2023.

[5] Sonia I Seneviratne, Markus G Donat, Andy J Pitman, Reto Knutti, and Robert L Wilby. Allowable co2 emissions based on regional and impact-related climate targets. *Nature*, 529(7587):477–483, 2016.

[6] Theo Lynn, Pierangelo Rosati, Antonia Egli, Stelios Krinidis, Komninos Angelakoglou, Vasileios Sougkakis, Dimitrios Tzovaras, Mohamad Kassem, David Greenwood, and Omar Doukari. Rinno: Towards an open renovation platform for integrated design and delivery of deep renovation projects. *Sustainability*, 13(11):6018, 2021.

[7] Zahra Pezeshki and Syed Ali Soleimani Ivari. Applications of bim: a brief review and future outline. *Archives of Computational Methods in Engineering*, 25:273–312, 2018.

[8] Charlotte Eloise Stancioff, Laila Māra Pesoa, Plamen Penev, and Karine Jegiazarjana. The sunshine platform: efficiency, transparency and standardization in the deep renovation process of multi-family buildings. *Open Research Europe*, 1:86, 2021.

[9] Faidra Filippidou, Nico Nieboer, and Henk Visscher. Are we moving fast enough? the energy renovation rate of the dutch non-profit housing using the national energy labelling database. *Energy Policy*, 109:488–498, 2017.

[10] Energy Efficiency Directive. Eu energy efficiency directive 2012/27/eu. *Journal reference*, 50(315): 1–56, 2012.

[11] Nina Holck Sandberg, Igor Sartori, Oliver Heidrich, Richard Dawson, Elena Dascalaki, Stella Dimitriou, Tomáš Vimm-r, Faidra Filippidou, Gašper Stegnar, Marjana Šijanec Zavrl, et al. Dynamic building stock modelling: Application to 11 european countries to support the energy efficiency and retrofit ambitions of the eu. *Energy and Buildings*, 132:26–38, 2016.

[12] D Staniaszek, O Rapf, F Mariottini, R Lottes, S Kunkel, S Danciu, M Fabeer, C Marian, and M Dumitru. Renovation strategies of selected eu countries: a status report on compliance with article 4 of the energy efficiency directive. *Buildings Performance Institute Europe (BPIE): Brussels, Belgium*, 2014.

[13] Charles Wilson, Hazel Pettifor, and George Chryssochoidis. Quantitative modelling of why and how homeowners decide to renovate energy efficiently. *Applied energy*, 212:1333–1344, 2018.

[14] Shaohua Jiang, Meng Wang, and Ling Ma. Gaps and requirements for applying automatic architectural design to building renovation. *Automation in Construction*, 147:104742, 2023.

[15] Alejandro Prieto, Tatiana Armijos-Moya, and Thaleia Konstantinou. Renovation process challenges and barriers: addressing the communication and coordination bottlenecks in the zero-energy building renovation workflow in european residential buildings. *Architectural Science Review*, 67(3): 205–217, 2024.

[16] Rajan Kumar Jaysawal, Suprava Chakraborty, D Elangovan, and Sanjeevikumar Padmanaban. Concept of net zero energy buildings (nzeb)-a literature review. *Cleaner Engineering and Technology*, 11: 100582, 2022.

[17] Yuan Zhan, Yu Peng, and Changrong Xiong. Application of bim in renovation design of existing buildings. In *International Conference on Urban Climate, Sustainability and Urban Design*, pages 263–271. Springer, 2023.

[18] Kepa Iturralde, Sathwik Amburi, Sandhanakrishnan Ravichandran, Samanti Das, Danya Liu, and Thomas Bock. Online modelling and prefab layout definition for building renovation. In *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, pages 605–608. IAARC Publications, 2023.

[19] Jose Ferrandiz, Abdulaziz Banawi, and Enric Peña. Evaluating the benefits of introducing "bim" based on revit in construction courses, without changing the course schedule. *Universal Access in the Information Society*, 17:491–501, 2018.

[20] AY Mainicheva, VV Talapov, and Guanying Zhang. Principles of the information modeling of cultural heritage objects: the case of wooden buddhist temples. *Archaeology, Ethnology & Anthropology of Eurasia*, 45(2):142–148, 2017.

[21] Vladimir V Sharmanov, Tatyana L Simankina, and Anton E Mamaev. Bim in the assessment of labor protection. *Magazine of civil engineering*, (1 (69)): 77–88, 2017.

[22] Hanbin Luo and Peisong Gong. A bim-based code compliance checking process of deep foundation construction plans. *Journal of Intelligent & Robotic Systems*, 79:549–576, 2015.

[23] Amarnath Chegu Badrinath, Yun-Tsui Chang, and Shang-Hsien Hsieh. A review of tertiary bim education for advanced engineering communication with visualization. *Visualization in Engineering*, 4:1–17, 2016.

[24] Bernardette Soust-Verdaguer, Carmen Llatas, and Antonio García-Martínez. Critical review of bim-based lca method to buildings. *Energy and Buildings*, 136:110–120, 2017.

[25] Richard C Gerum, Sebastian Richter, Alexander Winterl, Christoph Mark, Ben Fabry, Céline Le Bohec, and Daniel P Zitterbart. Cameratransform: A python package for perspective corrections and image mapping. *SoftwareX*, 10:100333, 2019.

Acknowledgement

The research presented in this paper was originally carried out in the context of the H2020 project **ENSNARE**. **ENSNARE** has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. **958445**.

Construction Industry Vision Alberta Dataset (CIVAD): Developing a Comprehensive Object Detection Dataset for Diverse Construction Applications

Mohamed Sabek[1], [2], Qipei Mei[1], [3], Gaang Lee[1], [4], Ali Golabchi[1], [5], and Vicente Gonzalez[1], [6]
1 Infrastructure and Human Tech Lab (https://www.iht-lab.com/), Department of Civil and Environmental Engineering, Faculty of Engineering, Donadeo Innovation Centre for Engineering, University of Alberta, 116 Street NW, Edmonton T6G 1H9, Alberta, Canada
2 sabek@ualberta.ca , **3** qipei.mei@ualberta.ca , **4** gaang@ualberta.ca , **5** alireza1@ualberta.ca ,**6** vagonzal@ualberta.ca

Abstract

Integrating computer vision technologies into the construction industry has the potential to revolutionize site monitoring, safety management, and quality control. However, a critical gap remains in the availability of specialized datasets tailored to construction sites' distinct conditions and complexities while including sufficient classes representing most items in construction sites. Existing Computer Vision (CV) models often rely on generic training datasets, which limit their effectiveness for specific construction-monitoring-related tasks. Consequently, there is a pressing need for comprehensive domain-specific datasets that can capture the full spectrum of construction-related objects and activities.

This study addresses this gap by developing a foundational training dataset called the Construction Industry Vision Alberta Dataset (CIVAD), specifically designed for CV applications in the construction sector. Our dataset included over 50 classes with more than 86,905 images of different objects, such as tools, machinery, safety equipment, and construction materials, to support diverse CV tasks. It utilizes a combination of web scraping, inclusion of existing open-source datasets, and direct data captured from construction sites.

A set of novel methods, such as semiauto-labeling with advanced models, such as Grounded SAM and Grounding DINO, were used with our custom algorithms. These models were utilized to process parts of the dataset imagery with humans in the loop. This approach facilitated an efficient and accurate dataset creation process. The CIVAD dataset and methods employed in this study represent a significant step forward in integrating CV technologies across various construction-related applications.

Keywords – Construction Industry, Computer Vision, Foundational Training Dataset, Object Detection, Auto-labeling, YOLO v11, Site Monitoring

1 Introduction

The construction industry is one of the largest and most complex sectors in the world and has a significant impact on the global economy [1]. It builds and maintains physical infrastructure, including roads, bridges, buildings, and other critical structures that support modern society. It remains a highly diverse and dynamic sector characterized by inherent challenges, including constantly evolving project environments, diverse stakeholder requirements, and the interplay of numerous variables and uncertainties that directly impact project outcomes [2]. These complexities require innovative solutions to enhance efficiency, ensure safety, and maintain quality standards under ever-changing conditions.

Although digital technologies have gradually penetrated the construction industry, they are yet to catalyze a fundamental transformation in their practices. The industry remains manual and fragmented, grappling with longstanding inefficiencies and challenges that underscore the nascent stage of it is digital evolution and highlight the gap in integrating Industry 4.0 [3]. However, in recent years, the construction industry has experienced several changes, primarily fueled by the growing implementation of artificial intelligence technologies, particularly in subfields such as computer vision.

CV has the potential to revolutionize various aspects of construction [4], including site monitoring, progress tracking, and safety compliance monitoring. CV technologies use cameras and other sensors to capture and analyze visual data, thereby providing valuable insights and information that can be used to improve construction processes and outcomes [12]. Despite the potential of CV technologies, their adoption in the construction industry has been limited. One of the most notable barriers is the lack of standardized datasets. Developing field-applicable computer vision models entails datasets tailored to the unique and dynamic conditions of construction sites, including diverse construction-related objects and environmental variability [13].

However, creating these datasets is challenging. First, collecting and annotating large amounts of visual data is labor-intensive and costly, and requires specialized knowledge of construction processes and CV techniques. In addition, the fragmented nature of the construction industry, in which projects often vary widely in scale, location, and standards, makes it difficult to develop universal datasets that replicate the complex environments of construction sites [4].

2 Related Work

Computer vision technologies are increasingly applied in the construction industry to help perform specific tasks [5], such as monitoring and tracking construction equipment at construction sites. This cannot be achieved without proper training datasets to train the CV models on these activities, which is essential for building any CV model. The training datasets help the model to understand it as a deployment environment, which in this case is a construction site. To support these efforts, researchers have developed several domain-specific datasets that can be used to train and evaluate CV models in construction-related applications. This study examines four widely utilized and cited publicly available construction datasets: MOCS [6], SODA [7], ACID [8], and CIS [9], highlighting their contributions and limitations and demonstrating how these shortcomings highlight the need for the development of more comprehensive and versatile datasets.

The MOCS (Moving Objects in Construction Sites) dataset contains roughly 41,668 images sourced from 174 construction sites. It consists of 13 classes of moving objects commonly found in construction environments, including workers, various types of cranes, vehicles, and machinery. Although MOCS has proven valuable for safety analysis, activity recognition, and productivity monitoring, the limited number of object classes restricts it is applicability to a narrow range of elements commonly encountered in traditional construction settings. In addition, it focuses on dynamic entities and overlooks the material, environmental, and structural aspects, making the trained models less adaptable to complex or unconventional construction scenarios.

Although MOCS has broader coverage, the Site Object Detection Dataset (SODA) features 19,846 images across 15 classes, offering a more balanced approach to object selection. However, the overall scale and scope of the dataset remain limited because of the relatively small number of classes and limited extent of scene variability. Models trained merely on SODA are not sufficiently generalizable to cover many items found within the construction environment, such as different types of construction equipment tools and materials, because of the limited scope of the SODA dataset in terms of the number of classes. A significant limitation of both the MOCS and SODA datasets is their geographic bias, as most of the annotated data were collected exclusively from construction sites in Asian countries. This narrow scope restricts the dataset's ability to generalize to North American or European contexts, where differences in color schemes, signage, safety banner languages, and weather conditions, such as snow, can significantly affect the performance and accuracy of the trained models. Another existing dataset is the Alberta Construction Image Dataset (ACID), which consists of 10,000 images across 10 classes with instance segmentation and image captioning. It primarily focuses on heavy construction equipment. Despite it is usefulness in machinery identification and detection, ACID is narrow and has only a few categories, limiting it is applicability to more integrated applications in construction [23].

Another important dataset is the Construction Instance Segmentation (CIS) dataset, which has many unique features that differentiate it from other construction-specific datasets. The dataset contained 50,000 images with over 100,000 instances annotated across 10 different classes. One of the most notable aspects is the inclusion of Prefabricated Building Construction (PBC) objects such as precast components (PCs) and PC delivery trucks, which are required for computer vision-based studies on PBC management. Moreover, the dataset distinguishes between workers who wore safety helmets and those who did not, making helmet-wearing detection more efficient than separately detecting helmets and workers. The CIS dataset spans a large set of construction sites, viewpoints, and weather, providing a variety of construction-specific object segmentation, helmet detection, and large instance segmentation data, thereby addressing gaps in the current PBC-related object segmentation research. However, the CIS dataset still suffers from the same limitations that exist in all the previously discussed datasets, which is the limited number of classes included in these datasets, which makes them only cover a specific portion of machinery, tools, workers, and materials existing in traditional construction sites.

These limitations make it challenging to build and train accurate CV models that are capable of handling the diverse needs of traditional construction sites. These sites typically involve multiple vehicle types, various tools, numerous workers, and assorted materials that must be tracked and classified. Consequently, tasks beyond equipment-related recognition, such as integrated site monitoring, schedule compliance checks, and resource allocation, are incredibly challenging without broader scene coverage.

Although merging existing datasets to mitigate their individual gaps could be considered, in practice, combining them with CV models such as YOLO is difficult because of potential issues such as data misalignment, differing annotation formats, inconsistent image resolutions, varied augmentation strategies, and ensuring that all datasets use the same classes [10].

3 Addressing Shortcomings in Construction Datasets: Why CIVAD Is Necessary

Although pioneering datasets, such as MOCS, SODA, ACID, and CIS, have advanced CV research in construction, they share limitations that hinder broader industry adoption, highlighting the need for a new dataset. This dataset captures the unique complexities of construction sites, including variable lighting, occlusions, and diverse camera viewpoints. Unlike static environments, construction sites often face unpredictable conditions [11]. Additional factors, such as local safety regulations and specialized machinery in places such as Canada, further underscore the need for more comprehensive construction-specific datasets.

In response to these limitations, we created a new dataset called the Construction Industry Vision Alberta Dataset (CIVAD), which offers a more versatile training option for researchers looking into training computer vision models to implement in different types of applications within the construction industry. These applications can include productivity estimations, vehicle tracking and monitoring, safety monitoring of construction individuals, and identification of potential bottlenecks in different construction processes. The CIVAD contains more than 86,905 images across 50 classes, and Figure 1 shows a sample of the dataset. In terms of class numbers, diversity, and the general number of annotated images, CIVAD covers much more than any other available computer vision dataset, specifically for construction applications, as listed in Table 1.



Figure 1- a sample of the CIVAD Dataset shows different classes

Table 1. A comparison between CIVAD and other datasets

	Dataset name				
Feature/Metric	**MOCS**	**SODA**	**ACID**	**CIS**	**CIVAD**
Total Images	41,668	19,846	10,000	50,000	86,905
Number of Classes	13	15	10	10	50
Annotation Types	Bounding Boxes, Masks	Bounding Boxes	Bounding Boxes, Limited Segmentation	Bounding Boxes, Masks	Bounding Boxes, Multiple Formats
Regional Scope	Asia	Asia	Canada	Asia	Worldwide
Objects Covered	Moving Objects	Site Objects	Heavy Machinery	PBC Objects, Helmet Detection	Workers, Tools, Machinery, Materials, Safety Equipment
Specialized Classes	Vehicles, Workers	Equipment, Tools	Machinery	Workers (Helmet Detection), PBC	Machinery, Workers, Safety Equipment, Tools, Materials
Format ready to Fintune Vision-Language Models	No	No	No	No	Yes
Application Areas	Safety Analysis, Activity Recognition	Object Detection	Machinery Identification	PBC Management, Safety Compliance	Safety Monitoring, Productivity Analysis, Resource Management
Unique Features	Focus on dynamic objects	Balance between object types	Heavy machinery emphasis	Helmet detection, PBC-specific data	Comprehensive data diversity
Data Collection Method	Site Collection	Site Collection	Open-source, Site Collection	Site Collection	Web Scraping, Site Collection, Augmented

CIVAD is unique because it is data diverse, which includes machinery, tools, workers, safety equipment, materials, and infrastructure components; therefore, it broadly represents the construction environment. To improve the robustness of the trained CV model on CIVAD, the dataset covers a wide range of weather conditions, lighting conditions, and different camera angles, with different types of augmentations applied to the dataset to expand it is size and versatility. It is also important to mention that our dataset offers training data in more than five different formats, ranging from JSON, TFRecord, and Pascal VOC to the YOLO format in TXT. This diversity is particularly advantageous for fine-tuning Vision-Language Models (VLMs) and vision foundation models, such as Microsoft's Florence-2 [15], which rely on large and variable datasets to learn robust representations.

By incorporating multiple object types under different environmental scenarios, CIVAD helps these models capture subtle visual cues that generic datasets may overlook, such as the unique features of construction equipment or the contextual interplay of workers, machinery, and materials. VLMs demonstrate notable zero-shot detection abilities and can detect and classify objects without explicit training data [16]. CIVAD enables researchers and practitioners to continuously refine the accuracy and generalization of VLM in construction environments, allowing the development of real-world applications. CIVAD fills a gap in the literature by providing a standardized and comprehensive dataset that bridges the gap between academic research and practical implementation.

This can accelerate innovation in computer vision technology and encourage the development of resilient AI models that can reshape construction safety, productivity, and efficiency.

4 CIVAD Dataset Creation Process

The CIVAD dataset was created using a well-defined process with different steps and techniques to efficiently and accurately develop this large dataset. This includes new approaches for making this dataset as efficient as possible, as shown in Figure 2, which shows all steps from data collection to the final dataset composition.

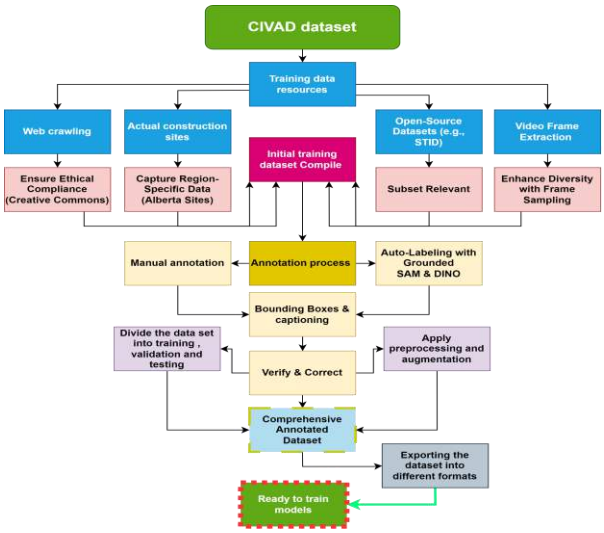


Figure 2. CIVAD dataset creation steps

4.1 included classes

CIVAD currently includes 50 object classes organized into five major categories, as shown in Table 2.

Table 2. CIVAD classes according to their categories

Categories	Classes names				
Machinery	Articulated Haulers	Bulldozer	Concrete Mixer	Concrete Pump	Dump Truck
	Excavator	Forklift	Mobile Crane	Motor Graders	Pile Boring Machine
	Scissor Lift	Sheepsfoot Roller	Skid Steer Loader	Smooth Roller	Static Crane
	Tower Crane	Wheel Loader			
Workers	White Hat Engineers	Red Hat Firefighter	Yellow Hat Worker	Green Hat Safety Officer	-
Tools	Bolts	Wrench	Trowel	Bucket	Cutter
	Drill	Grinder	Hammer	Knife	Saw
	Shovel	Spanner	Tacker		
Safety Equipment	Safety Barrier	Safety Cone	Safety Helmet	Safety Shoes	Safety Sign
	Safety Vest	Traffic Barrel	Welding Helmets	Safety Goggles	
Materials	Wood Plank	Concrete Block	-	-	-

4.2 Image Collection

The process of collecting images for the CIVAD dataset was carefully designed to strike a balance between scale, diversity, and realism, all of which have been recognized as critical factors in developing strong and generalizable computer vision models [27]. For example, large and varied datasets, such as COCO [14], have demonstrated that increasing both the quantity and variety of training samples can significantly improve a model's ability to generalize to real-world scenarios. Additionally, realism ensures that data capture authentic conditions and further refines model performance when transitioning from laboratory to field applications [28].
The CIVAD dataset creation methodology utilizes a combination of approaches, including web scraping, subsetting open-source datasets, and manual data capture to assemble a large, varied, and contextually relevant dataset that aligns with the requirements of the construction industry.

By encompassing a wide range of conditions and subjects, CIVAD aims to mirror the complexity and unpredictability inherent in real construction sites, thereby enhancing the robustness and transferability of the trained models. All images and video sources used were ethically compliant (i.e., open access or licensed under Creative Commons). During scraping, specific criteria were used to ensure that the retrieved images contained relevant construction activity, machinery, workers, tools, and materials. The systematic analysis of many online archives, such as Pixabay [26], has enabled the collection of a large pool of high-resolution visuals that represent real-world construction environments. Frame extraction from videos was used to increase the size and diversity of the dataset.

Individual frames were extracted from high-resolution videos at regular intervals during construction activities. This technique increases the total number of images and brings about natural variations in lighting, angles, and viewpoints, as well as worker/machine interaction, making the photos more realistic and providing rich supervision [29]. Frame extraction ensures that the dataset contains the dynamic and ever-changing nature of construction environments while replicating real-world scenarios better than static imagery alone [30]. We also included a subset of 3,600 pictures from an existing open-source dataset called the Small Tool Image Database (STID), which is under Creative Commons license 4.0 [17]. The STID dataset consists of 12 classes of small tools used in the construction industry, which we used as a base to include these types of tools in our dataset and then expanded it with more images for each class.

This inclusion was motivated by the need to represent small, handheld tools that are critical to everyday construction tasks yet are often underrepresented in general-purpose image datasets. Incorporating STID data helped ensure that these tools, which can significantly influence safety, productivity, and site operations, were adequately covered in the CIVAD. Also, incorporating a pre-annotated subset of 3,600 images from the STID (Small Tool Image Database) significantly reduced the time and effort needed to include 12 classes of small tools in our dataset, sparing us from annotating all such data from scratch. However, many publicly available construction image datasets originate from or predominantly replicate Asian construction environments, which can differ markedly from those found in Canada, the U.S., and Europe; for instance, in the types of equipment used, signage design, and specific safety protocols.

Therefore, we conducted manual data capture at construction sites in Alberta, Canada, where we captured approximately 3,158 pictures representing approximately 10% of the CIVAD without augmentation, to address these regional discrepancies and enhance the applicability of the dataset in Canadian contexts. This approach ensures that our dataset reflects region-specific factors such as environmental conditions and safety standards, which may not be adequately represented in datasets with predominantly Asian or non-Canadian content. Researchers obtain permission from project managers at active construction sites, ranging from commercial developments to infrastructure projects, and adhere to strict guidelines to respect property rights and protect sensitive information. As a result, the collected images more accurately represented the construction machinery, materials, and safety measures typical of North American settings. This emphasis on North American conditions, combined with the diverse set of tools, weather conditions, and site layouts in the dataset, boosts the overall realism and transferability of computer vision applications across regions, with construction practices similar to those in Canada, the U.S., and potentially even parts of Europe.

The final dataset comprised 86,905 images, which were expanded from 32,380 images using different augmentation techniques and strategically split into training, validation, and test sets to support model development and evaluation: a training set of 81,849 images (94%), a validation set of 3,632 images (4%), and a test set of 1,424 images (2%) of the total dataset.

4.3 The Annotation Process

To annotate the CIVAD dataset, the Roboflow platform [18] was utilized to perform accurate and efficient construction image labeling. Because of the large size of the CIVAD dataset (over 32,380 images without augmentation), a semiautomatic annotation approach was used to ensure accuracy and efficiency. Manual annotation and integration of cutting-edge automated tools were used in a process that combined these approaches. Manual annotation was performed to ensure that key objects, such as machinery, tools, workers, and safety equipment, were adequately identified and labeled by applying bounding boxes to the objects of interest in the images to achieve accuracy in object localization. For annotation acceleration and dataset scale management, semi-automatic labeling techniques have used state-of-the-art foundational computer vision models such as Grounded SAM [19] and Grounding DINO [20], which generate automatic predictions of bounding box coordinates.

These models proved helpful for adding annotations to identifiable object classes based on their training data; however, not all objects were correctly identified. Therefore, a thorough manual review was performed to ensure that all the relevant classes were properly labeled in each image. This hybrid approach reduced the annotation time as much as possible without sacrificing the accuracy across the entire dataset. The result was a comprehensive dataset of 32,380 images annotated with over 58,789 instances representing 50 object classes. On average, there are 2.8 annotations per image, demonstrating the complexity and variety of construction site visuals. It's worth noting that the resolution of the images in the dataset ranged from 4 megapixels to 1 megapixel, and the median image ratio for the dataset was 1280x900 pixels, with an average of 1.08 megapixels throughout the whole dataset. Figure 3 shows the number of individual annotation count distributions for images in the CIVAD dataset.

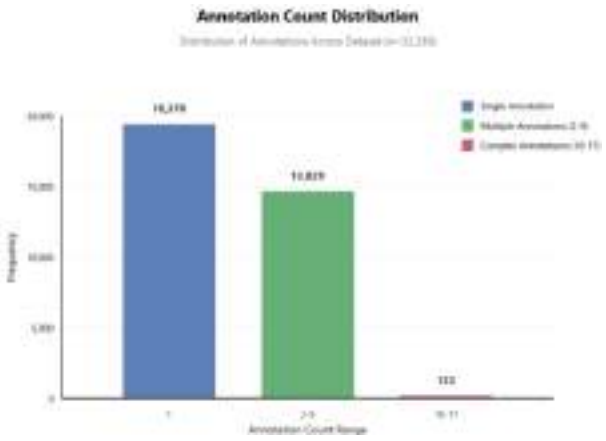


Figure 3. Annotation count distributions for images in the CIVAD dataset.

The attributes of this dataset, such as high-resolution images, diverse object classes, and realistic coverage, make it ready to train robust CV models for object detection and tracking tasks. Higher-resolution images in CIVAD capture subtle details, such as tool features and safety gear, thereby improving accuracy in complex environments. A variety of machinery, tools, and contexts help models learn richer features and generalize more effectively. The combination of manual precision with automated tools leads to annotations with a sufficient scale and the required quality accuracy to serve a range of computer vision applications in the construction industry, as proven by the high accuracy of the trained models on the dataset, which will be discussed in the next section.

4.4 The Applied Preprocessing and Augmentations Process on CIVAD

A series of preprocessing techniques was applied to standardize the images and enhance their usability in preparing the dataset for use in computer vision tasks. We began by performing an auto-orientation step to correct the image alignment by ensuring a consistent orientation based on the EXIF metadata, thereby preventing annotation mismatches caused by misaligned pixel coordinates. This preprocessing step eliminates common issues in computer vision pipelines, such as misrotation, ensuring that the images in the CIVAD dataset are uniformly oriented, have accurate annotations, and are ready for training [25]. All images were then resized to 640 × 640 pixels, which is a standard input size requirement for YOLO-based models [21], to ensure compatibility with state-of-the-art object detection frameworks and reduce computational overhead. Various augmentation techniques have been used to create camera augmentations for simulating conditions commonly seen in real-world construction sites, further diversifying the dataset and improving the model's generalization ability.

The following augmentations were applied:

- Crop: Variations in the scale and focal point of the objects were achieved by cropping the images with 0% minimum zoom and 20% maximum zoom, which gave us good results in our tests.
- Rotation: To simulate tilted or angled camera views, random rotations between -15° and +15° were performed.
- Exposure: Changes in lighting conditions were applied, by changing lighting conditions within a range of -15% to +15%, reflecting the effects of natural and artificial lighting on construction sites.

Augmentation was strategically applied to expand CIVAD from 32,380 original images to 86,905 images, significantly enhancing the dataset diversity without incurring additional annotation costs. The rationale for this augmentation is to simulate real-world variations, such as lighting changes, camera angles, and object perspectives, which are crucial for training robust and generalizable computer vision models, thus reflecting realistic construction site scenarios more accurately.

5 Trained Model Performance Evaluation

In this section, we discuss the training of a YOLOv11-M model [24] on CIVAD's 50 classes to confirm the dataset's quality and alignment with our research goals. The dataset was divided into training (81,849 images), validation (3,632 images), and test (1, 424 images) sets. Training was performed over 300 epochs with a batch size of 16, an initial learning rate of 0.001, and stochastic gradient descent optimization on an NVIDIA GeForce RTX 4090 GPU with 24 GB of VRAM [22], which required approximately 44 hours to complete. By applying preprocessing and strategic augmentation, we increased the dataset from the original size of 32,380 images to 86,905 images. The rationale behind this augmentation was to simulate real-world variability, such as changes in lighting, camera angles, and environmental conditions, thus enhancing the robustness and generalization capability of the trained computer vision models. To comprehensively evaluate the model's performance, we utilized the mean Average Precision (mAP), Precision, and Recall, which are recognized metrics for evaluating object detection model performance [31]. These metrics reflect the accuracy of the model and can be generalized across diverse construction scenarios, validating both the dataset size and appropriateness of augmentation techniques for simulating real-world variations.

These metrics demonstrate that the model is capable of accurately detecting and classifying objects from the 50 object classes of workers, machinery, tools, materials, and safety equipment across the dataset, as shown in Figure 4, which shows the trained model's performance across some of the CIVAD classes.

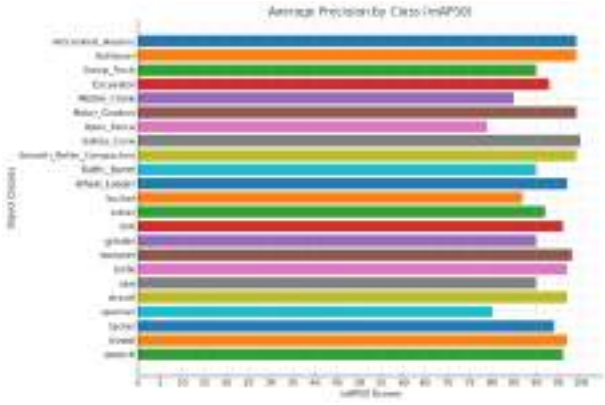


Figure 4. YOLOv11-M training results on common classes from the CIVAD dataset

6 Conclusion

To support the digital transformation of the construction industry, this study introduces some promising annotation techniques along with the CIVAD dataset, which comprises more than 86,905 images across 50 classes, capturing only a subset of the existing items in traditional construction sites. However, this dataset bridges critical gaps in existing datasets and provides vast diversity and applicability. It offers great flexibility through multiple annotation file formats and compatibility with various deep-learning frameworks, including YOLO and Vision-Language Models. CIVAD has great potential for use in safety monitoring, productivity analysis, and automated resource management, which can help improve the efficiency and safety of construction sites. We anticipate that these advancements will foster new opportunities for construction automation and that CIVAD will continue to catalyze innovation in computer vision applications related to the construction industry for many years.

7 Future Work

The CIVAD dataset provides a robust foundation for increasing the number of existing computer vision applications in construction. However, the evolution of the dataset is necessary to meet changing requirements and extend it is scope.

Future plans include:

- Expansion of Dataset Scope: We will increase the number of object classes from 50 to over 100 and the dataset size to more than 120,000 images. The dataset will include more sophisticated tools, machinery, materials, and safety equipment to better represent the different environments at construction sites. This is important for extending the generalization capabilities of the dataset.
- Enhancing Dataset Diversity: Introducing rare and region-specific scenarios of construction, such as nighttime activities, extreme weather, and geographic diversity, to enhance model robustness and generalization.
- Support for Vision-Language Models (VLMs): We will introduce additional captioning, metadata, and annotations tailored for VLM development. This multimodal focus will enable automated documentation, semantic scene understanding, and other tasks that require synergy between textual and visual data.

These points will expand the CIVAD dataset as a valuable tool to help researchers overcome barriers that hinder the integration of modern technologies, such as computer vision, into traditional industries, including the construction industry. This will enable researchers to build more reliable computer vision models in the construction industry.

Disclaimer: Specific portions of this dataset will be available upon request. Please contact Dr. Vicente Gonzalez at vagonzal@ualberta.ca to submit any requests.

References

[1] Alaloul W. S., Musarat M. A., Rabbani M. B. A., Altaf M., Alzubi K. M., and Al Salaheen M. Assessment of economic sustainability in the construction sector: Evidence from three developed countries (the USA, China, and the UK). Sustainability, 14(10):6326, 2022.

[2] Osobajo O., Oke A., Omotayo T., and Obi L. Systematic review of circular economy research in construction industry. Smart and Sustainable Built Environment, 11(1):39-64, 2020.

[3] Costin A., Hu H., and Medlock R. Building information modeling for bridges and structures: Outcomes and lessons learned from the steel bridge industry. Transportation Research Record Journal of the Transportation Research Board, 267:576-586, 2021.

[4] Pal A. and Hsieh S. H. Deep-learning-based visual data analytics for smart construction management. Automation in Construction, 131, 2021.
[5] Jiang Z. and Messner J. I. Computer vision applications in construction and asset management phases: Literature review. Journal of Information Technology in Construction, 28, 2023.
[6] Xuehui A., Li Z., Zuguang L., Chengzhi W., Pengfei L., and Zhiwei L. Dataset and benchmark for detecting moving objects in construction sites. Automation in Construction, 122, 2021.
[7] Duan R., Deng H., Tian M., Deng Y., and Lin J. SODA: Large-scale open-site object detection dataset for deep learning in construction. Automation in Construction, 142:104499, 2022.
[8] Xiao B. and Kang S. C. Development of an image dataset of construction machines for deep-learning object detection. Journal of Computing in Civil Engineering, 35(2):05020005, 2021.
[9] Yan X., Zhang H., Wu Y., Lin C., and Liu S. Construction instance segmentation (CIS) dataset for deep learning-based computer vision. Automation in Construction, 156:105083, 2023.
[10] Zhao J., Ou M., Xue L., Cui Y., Wu S., and Chen G. Joining datasets via data augmentation in the label space for neural networks. ArXiv, abs/2106.09260, 2021.
[11] Schuldt S., Nicholson M. R., Adams Y. A., and Delorit J. D. Weather-related construction delays in a changing climate: A systematic state-of-the-art review. Sustainability, 2021.
[12] Guo B. H., Zou Y., Fang Y., Goh Y., and Zou P. X. Computer vision technologies for safety science and management in construction: A critical review and future research directions. Safety Science, 135:105130, 2021.
[13] Salimi M., Loni M., Afshar S., Sirjani M., and Cicchetti A. ConstScene: Dataset and model for advancing robust semantic segmentation in construction environments. ArXiv, abs/2312.16516, 2023.
[14] Lin T. Y., Maire M., Belongie S., Hays J., Perona P., Ramanan D., and Zitnick C. L. Microsoft Coco: Common objects in the context. In Proceedings of the 13th European Conference on Computer Vision (ECCV), pages 740-755, Zurich, Switzerland, 2014.
[15] Xiao B., Wu H., Xu W., Dai X., Hu H., Lu Y., and Yuan L. Florence-2: Advancing a unified representation for various vision tasks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024.
[16] Al Rahhal M. M., Bazi Y., Elgibreen H., and Zuair M. Vision-language models for zero-shot classification of remote-sensing images. Applied Sciences, 13(22):12462, 2023.
[17] Lee K., Jeon C., and Shin D. H. Small-tool image database and object detection approach for indoor construction site safety. KSCE Journal of Civil Engineering, 27(3):930-939, 2023.
[18] Roboflow. Computer vision tools for developers and enterprises. On-line: https://roboflow.com/, Accessed: 19/12/2024.
[19] Ren T., Liu S., Zeng A., Lin J., Li K., Cao H., and Zhang L. Grounded Sam: Assembling open-world models for diverse visual tasks. ArXiv, abs/2401.10968, 2024.
[20] Liu S., Zeng Z., Ren T., Li F., Zhang H., Yang J., and Zhang L. Grounding Dino: Marrying a Dino with grounded pre-training for open-set object detection. In Proceedings of the European Conference on Computer Vision (ECCV), pages 38-55, 2024.
[21] Jiang P., Ergu D., Liu F., Cai Y., and Ma B. Review of the YOLO algorithm developments. Procedia Computer Science, 199, 2022.
[22] NVIDIA. NVIDIA GeForce RTX 4090 graphics card. On-line: https://www.nvidia.com/en-us/geforce/graphics-cards/40-series/rtx-4090/, Accessed: 19/12/2024.
[23] Park S. and Kim S. Semantic segmentation of heavy construction equipment based on point-cloud data. Buildings, 14(8):2393, 2024.
[24] Ultralytics. YOLO11, NEW. Ultralytics YOLO Docs. On-line: https://docs.ultralytics.com/models/yolo11/#overview, Accessed: 02/01/2025.
[25] Fritz J. Image rotation and Exif Data. On-line: https://www.jonathanfritz.ca/2016/07/25/image-rotation-and-exif-data/, Accessed: 01/01/2025.
[26] Pixabay open-source archive. On-line: https://pixabay.com/, Accessed: 05/01/2025.
[27] Sun P. et al. On the diversity and realism of distilled dataset: An efficient dataset distillation paradigm. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024.
[28] Michaelis C., Bethge M., and Ecker A. S. A broad dataset is all you need for one-shot object detection. ArXiv, abs/2011.04267, 2020.
[29] Feng Y., Zhang J., Li G., Togo R., Maeda K., Ogawa T., and Haseyama M. A novel frame-selection metric for video inpainting was used to enhance the urban feature extraction. Sensors, 2024.
[30] Bonetto E., Xu C., and Ahmad A. GRADE: Generating realistic animated dynamic environments for robotics research. ArXiv, abs/2310.17041, 2023.
[31] Padilla R., Netto S. L., and da Silva E. A. Survey of performance metrics for object-detection algorithms. In Proceedings of the International Conference on Systems, Signals, and Image Processing (IWSSIP), pages 237-242, 2020.

A Two-Phase AI-Driven Approach to Automated Construction Planning Using Small Language Models for Activity Sequencing and Missing Task Prediction

Akarsth Kumar Singh[1], Aritra Pal[2] and Shang-Hsien Hsieh[1]

[1]Department of Civil Engineering, National Taiwan University, Taiwan
[2]Department of Engineering, University of Cambridge, United Kingdom
akarsth@caece.net , ap2425@cam.ac.uk , shhsieh@ntu.edu.tw

Abstract

This research presents a two-phase AI-driven approach to enhance automated construction planning and scheduling, addressing common challenges such as project delays and budget overruns in the Architecture, Engineering, and Construction (AEC) industry, often caused by inefficient project management. Traditional AI methods rely heavily on static programming and manually annotated datasets, leading to inefficiencies. The proposed methodology uses prompt engineering to systematically extract and structure information from historical project schedules, creating datasets for training domain-specific Small Language Models (SLMs). In the first phase, fine-tuning SLMs, particularly mistral-7B, achieved 83.69% accuracy, outperforming traditional approaches. In the second phase, another fine-tuned SLM identified activity patterns, optimized sequencing, and predicted missing tasks, achieving a mean similarity score of 81.43%. The results demonstrate that structured prompt engineering significantly improves model accuracy and efficiency, reduces manual efforts, and addresses computational inefficiencies. This scalable, secure, and efficient approach sets a new standard in automated construction planning.

Keywords

Prompt Engineering; Domain-Specific Instruct Fine-Tuning; Small Language Models; Construction Planning and Scheduling; Activity Sequencing

1 Introduction

Construction delays and cost overruns are pervasive challenges in the AEC industry, significantly impacting the successful completion of projects [5]. Planning and scheduling are at the core of these issues and pivotal in determining project timelines and budgets. A deficiency in planning and scheduling processes can lead to delays, misallocated resources, and, ultimately, increased costs [7]. Inaccurate estimations, inadequate risk assessments, and failure to account for the complexities and interdependencies of construction activities often result in unrealistic and unachievable schedules. This lack of detailed and precise planning leads to inefficiencies in resource utilization, conflicts in task scheduling, and disruptions in the workflow, all of which contribute to extended project durations and escalated costs.

1.1 Challenges in AEC Project Planning and Scheduling

In the AEC industry, project planning and scheduling face significant challenges, notably in learning from documents, leveraging historical schedules, addressing fragmentation and interoperability issues, mitigating human errors in plan generation and updates, and implementing automation [4]. The intricate process of extracting actionable insights (key construction information and activity sequences) from extensive project documentation and past schedules is compounded by the diverse formats and standards, leading to fragmentation that hinders seamless data integration and interoperability across different project management tools. Furthermore, the reliance on manual input for generating and updating project plans introduces a high potential for human error, affecting the accuracy and reliability of schedules. These challenges highlight the need for advanced solutions to automate processes, improving efficiency and accuracy in project planning. Using sophisticated algorithms and machine learning to analyze historical data and automate schedule generation can greatly reduce these issues, leading to more streamlined and error-resistant project management in the AEC sector.

1.2 Need for Automated Planning and Scheduling

Automation in AEC project planning and scheduling is vital for optimizing project management practices [8].

Over the past three decades, researchers have explored various methods to reduce manual work and improve precision, including case-based reasoning, knowledge-based systems, model-based approaches, meta-heuristic algorithms, expert systems, and learning-based techniques [4]. However, challenges in data integration, platform interoperability, and adapting automation to project-specific needs persist. The complex and dynamic nature of construction projects demands more flexible solutions. Generative AI, with advanced Natural Language Understanding (NLU) and Natural Language Generation (NLG), presents a promising alternative [6], enhancing efficiency by interpreting complex documents and generating detailed schedules from large datasets. This marks a significant improvement in AEC automation in planning and scheduling.

2 Related Works

In the AEC industry, Building Information Modeling (BIM) and Natural Language Processing (NLP) have significantly enhanced project management by improving communication and information efficiency [11]. BIM revolutionizes the industry with digital building representations, facilitating automatic schedule generation and optimization, thereby reducing reliance on planners' expertise and using various data types like spatial and material information. Meanwhile, NLP, a sub-field of AI, focuses on computer interactions with human language through syntactic, semantic, and ontological analysis, accelerating information extraction from documents and interpreting construction processes. NLP's applications in the AEC sector include extracting data from BIM for safety compliance, analyzing contract documents, and automating Construction Schedule Management (CSM) [11], thereby improving schedule generation, validation, and accuracy.

2.1 Current Practices in Construction Planning and Scheduling

Recent efforts to automate construction scheduling have focused on leveraging BIM, with early methods combining BIM data with genetic algorithms [8] and incorporating graph-based indexing and AI for task scheduling [12], though BIM models often lack the granularity to capture project management aspects such as design approvals and procurement. More recent approaches employ NLP and deep learning to extract construction knowledge from historical schedules. Amer and Golparvar-Fard [2] introduced dynamic templates by learning from past schedules, while subsequent studies applied deep-learning-based NLP for sequence evaluation. Hong et al. [9] explored Graph Convolutional Networks to identify activity types and sequences.

2.2 Role of Language Models in the AEC Industry

LMs are revolutionizing planning and scheduling in the AEC industry by analyzing complex project data and identifying critical patterns for effective management [3]. They extract essential task details from project documents and past schedules [10], improving accuracy and reducing human error. Domain-specific training is essential [14], enabling LMs to handle construction-specific terminology and enhance project management applications. This training helps LMs identify tasks, understand resource needs, and predict challenges effectively. Integrating LMs into AEC workflows supports data-driven decisions, resource optimization, and delay reduction through historical data and predictive analytics [2][9]. Their performance improves with continued learning from real-world projects, promoting efficient project delivery.

This study addresses critical challenges in construction planning, particularly the extraction of structured information from historical project schedules and the generation of optimized activity sequences, including the prediction of missing tasks. To tackle these challenges, this research proposes a two-phase AI-driven framework integrating prompt engineering techniques with domain-specific fine-tuned SLMs. The paper is systematically organized, beginning with an introduction that outlines current industry challenges and emphasizes the need for automation in planning practices. It then reviews relevant literature, details the proposed methodology, presents experimental validations and analyses, and concludes by summarizing key findings with recommendations for future research directions.

3 Methodology

Figure 1 illustrates the framework proposed in this study for construction project planning using Generative AI. The approach leverages historical project schedules as the primary input. At its core is the extraction of key information, including detailed descriptions of the work breakdown structure, specific activities, tasks, and the network of predecessor-successor relationships. It captures the complex interplay of critical dependencies in construction planning. This study adopts a two-phase methodology that applies Generative AI and fine-tuned SLMs to enhance construction project planning and scheduling.

In Phase 1, prompt engineering is applied to extract and organize key information from existing and historical project schedules, resulting in a structured dataset. This dataset is then used to fine-tune an SLM for domain-specific information extraction in construction planning.

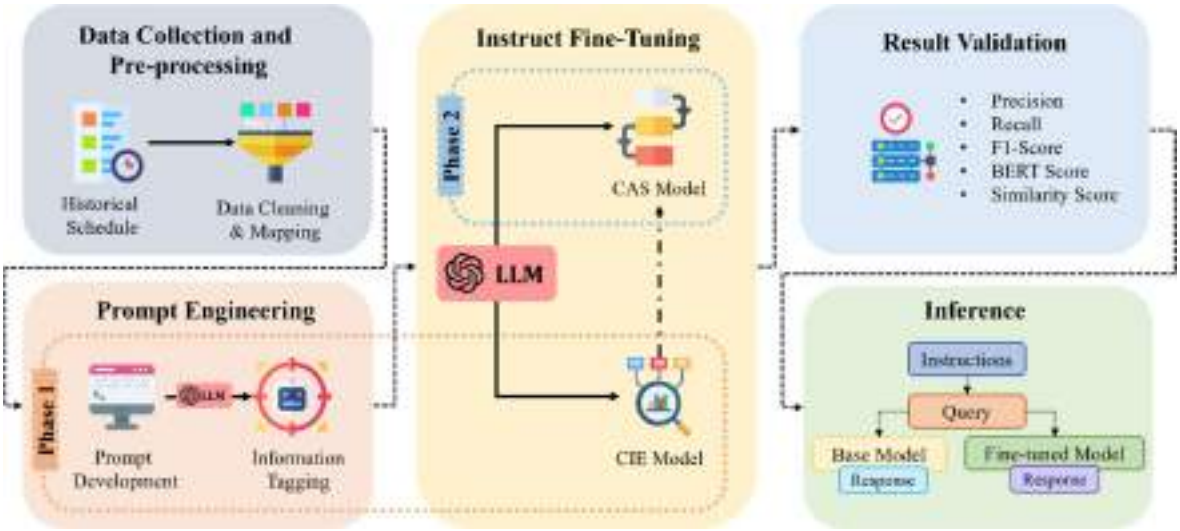


Figure 1. Overview of Proposed Methodology

In Phase 2, a separate SLM is trained on the organized data to learn activity patterns, enabling it to generate optimized activity sequences and identify missing tasks. The fine-tuned SLMs performance is evaluated against a benchmark model to assess its accuracy and reliability. Additionally, local deployment of the model ensures enhanced data privacy and security. This two-phase approach effectively reduces manual errors, accelerates processing, and improves the efficiency and accuracy of construction planning and scheduling.

3.1 Construction Activity Formalization for Data Preparation

In real-world construction projects, planning and scheduling have unique characteristics but share common structures, formats, and dependencies due to similar construction elements. Extracting information from historical schedules is challenging due to individualized approaches by schedulers, who use distinct terminologies based on specific project requirements [1]. For example, terms like "formwork" and "shuttering" serve the same function but differ in terminology. Work breakdown structures also vary; a cast-in-situ project may describe a superstructure task as "Construction-Super Structure-First Floor-Column-Reinforcement" under labels like "EPC Type-Work Item-Location-Object-Task," while a firefighting task may be labeled as "Construction-MEPF (Fire Fighting work)-Firefighting pipes-Fix or Installation-First Floor" under "EPC Type-Work Item-Object-Task-Location." This variation underscores the lack of uniformity in construction labels.

This study presents a structured approach to formalize construction activity definitions to improve data preparation for downstream processing. It consolidates details from diverse work breakdown structures and maps them to standardized labels "EPC Type, Work Items, Location, Object, and Task," emphasizing Location, Object, and Task. By harmonizing inconsistent terminologies and structures, this method improves the uniformity and usability of construction schedule data, enhancing the efficiency and accuracy of data analysis and model training.

3.2 Prompt Engineering and Key Construction Information Extraction

Prompt engineering is a vital technique for advanced models like GPT, involving the crafting of specific queries to guide the model in generating relevant outputs. It enhances precision and fine-tuning capabilities, enabling refined interaction and alignment with specific objectives. In the AEC industry, prompt engineering is a powerful tool for streamlining data retrieval [14], particularly when processing historical scheduling data. It automates the extraction and tagging of construction labels from past schedules, transforming a labor-intensive task into an efficient process. Effective prompt design requires understanding both model mechanics and domain-specific knowledge. In this study, prompts are strategically designed to capture relevant data and the complexities of construction project management. Prompt engineering, by targeting tasks, objects, and locations, ensures precise data extraction and categorization, forming a core part of our methodology.

3.3 Instruct Fine-Tuning SLMs for Construction Planning

In this study, we apply instruct fine-tuning to SLMs, tailoring these advanced AI systems to meet the

unique requirements of the construction industry. Although LMs are inherently powerful, their generalist training needs more depth for specialized tasks [13]. Our methodology strategically guides these models with carefully crafted prompts from prompt engineering data, enabling them to produce outputs that accurately align with the nuanced demands of construction management. This fine-tuning enhances the LMs proficiency in interpreting task-specific data [6], significantly improving their performance in complex construction tasks. The models are trained to parse detailed construction data, providing insights directly applicable to various aspects of construction management, such as extracting key information, generating activity sequences, and identifying missing activities. By refining these models to understand and process the unique aspects of construction scheduling and task interrelations, we transform SLMs into essential tools for delivering targeted, domain-specific solutions. This approach positions instruct fine-tuning at the core of our methodology, leveraging SLMs capabilities to advance construction project management.

4 Experimental Setup

4.1 Data Collection

In our approach to tagging construction constituents within comprehensive activities, we collected structured historical schedule data from two residential building projects, selecting 1500 activities for thorough analysis. To understand activity dependencies and sequences, we also compiled 1375 pairs of predecessor and successor activities, covering structural work, finishing, MEP (Mechanical, Electrical, and Plumbing), and commissioning domains. Using Python libraries, we efficiently extracted and parsed information from .xer files, the metadata files of Primavera, a widely used project management software in construction. This systematic organization prepared the data for subsequent analysis and model training in our study.

4.2 Materials and Methods

In this research, dataset preparation and model fine-tuning were conducted using Large Language Models (LLMs) via the OpenAI API and high-performance computational resources. Models were chosen for their domain-specific fine-tuning suitability, computational efficiency, and open-source availability. Initial evaluations were performed using the OpenAI API, while fine-tuning was done on Mistral-7B, Gemma-2-9B, and LLaMA-3.1-8B. Training utilized an NVIDIA RTX A6000 GPU with 48 GB VRAM, 50 GB RAM, and 8 virtual CPUs.

4.3 Prompt Template Preparation

Prompt templates are essential for effectively guiding and fine-tuning pre-trained models. In our approach, we use OpenAI's structured response prompt template with various instruction formats—zero-shot, one-shot, few-shot, chain-of-thought (CoT), and CoT with few-shot prompting. These formats help the model interpret and respond accurately by providing varying levels of contextual guidance. For fine-tuning, we adopt the Alpaca dataset format, illustrated in Figure 2, which standardizes instruction, user input, and assistant response. This ensures model outputs align with intended objectives. Fine-tuning is performed using open-source SLMs, allowing customization for domain-specific tasks. Combining OpenAI's prompt templates with the Alpaca format enhances the performance and adaptability of pre-trained models. This structured, context-rich data preparation enables the model to handle complex, domain-specific tasks with improved precision and efficiency.

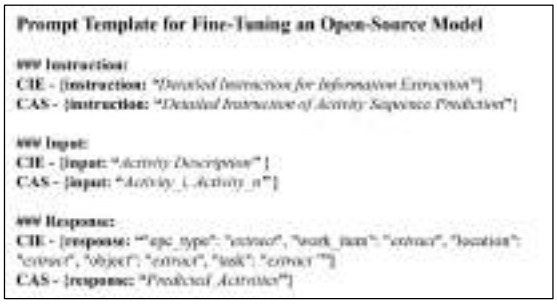


Figure 2. Overview of Prompt Template

4.4 Prompt Template for Key Construction Information Extraction (CIE)

This template is intricately designed to facilitate the extraction of construction constituents from activities. It includes explicit instructions that detail the specific constituents needed for each activity, like EPC Type, Work Item, Location, Object, and Task. Structured input is provided to ensure that all aspects of a construction activity are precisely captured and categorized. This categorized data is then used to fine-tune the Key Construction Information Extraction model. Initially, the template provides clear instructions for the information extraction task. Following this, the user input section allows for entering specific details regarding the construction activity. The model then processes this input and generates responses in specified categories. This structured interaction ensures the models accurate understanding and categorization of essential construction information, thereby enhancing information extraction accuracy.

4.5 Prompt Template for Construction Activity Sequencing (CAS)

For sequencing construction activities, the instruction section clearly outlines the sequencing task, guiding users to input detailed descriptions of construction activities. Users then provide specific sequences. The model processes this information and responds with a logically organized sequence of activities, capturing the necessary progressions and dependencies. This template is pivotal in accurately forecasting construction activity flows, which is crucial for effective project planning and execution.

Building upon the dataset preparation and fine-tuning processes described, the subsequent section presents a comprehensive evaluation of the selected language models. The results highlight their effectiveness in domain-specific tasks, particularly in information extraction and activity sequence generation, demonstrating their applicability and performance within the context of automated construction planning and scheduling.

5 Results and Discussion

In this research, LLM was first used to extract activity constituents for dataset preparation. The model's performance was evaluated using various instruction formats with a base LLM to identify the most effective strategy. The best-performing format was then used to prepare the dataset for fine-tuning, requiring only minor manual corrections to ensure accuracy and quality. The finalized dataset was split into 80% for training and 20% for testing. This structured approach ensured effective model learning and reliable performance evaluation on unseen test data.

5.1 Phase 1: Performance of Key Construction Information Extraction (CIE)

Figure 3 shows the impact of different prompting strategies on the GPT-4o-mini-base model's performance in extracting activity constituent information. The CoT Few-shot approach achieved the highest accuracy, proving that combining step-by-step reasoning with examples greatly enhances the model's predictive ability. Similarly, the CoT and Few-shot methods also improved accuracy, highlighting the importance of structured reasoning. In contrast, the Zero-shot strategy led to more non-predictions and hallucinations, indicating that lacking contextual guidance reduces the model's effectiveness. Overall, Figure 3 underscores that structured prompting with reasoning and examples are key to improving model performance in specialized tasks.

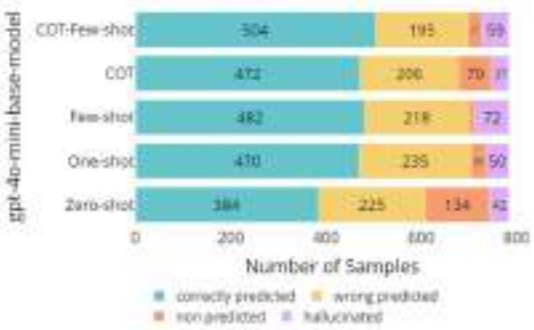


Figure 3. Performance Analysis of the CIE Model by Structured Prompt Techniques

The performance of fine-tuned SLMs, including mistral 7B, gemma 2 9B, and llama 3 8B, was evaluated for information extraction tasks. As shown in Figure 4, the mistral 7B model achieved the highest accuracy with the most correct predictions, followed closely by gemma 2 9B. In contrast, llama 3 8B exhibited lower accuracy with a higher number of non-predictions and hallucinated outputs. These models were fine-tuned using a consistent configuration across all models to ensure a fair comparison.

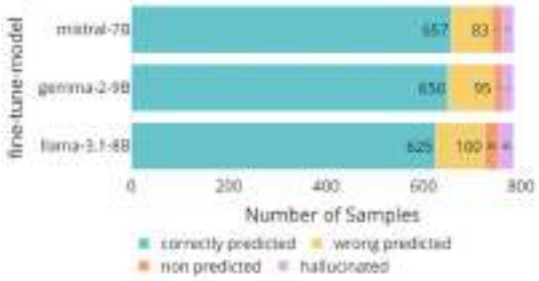


Figure 4. Performance Analysis of the CIE Model by Fine-tune model

Specifically, Low-Rank Adaptation (LoRA) adaptation was applied with a rank of 16, targeting key projection layers (e.g., q_proj, k_proj, v_proj) and optimized settings such as lora_alpha of 16 and lora_dropout set to 0. Training was conducted with a batch size of 2, gradient accumulation steps of 4, a learning rate of 2e-4.

To quantify model performance, all models were tested using these identical parameters, ensuring consistent evaluation. However, these parameters can be adjusted in future experiments to further enhance model performance.

Table 2 presents a performance comparison between fine-tuned open-source models and closed-source base models for information extraction tasks. Among the fine-tuned models, mistral 7B achieved the highest accuracy with 83.69% correct prediction, followed closely by

Table 2: Performance Comparison of Fine-Tuned Open-Source Models and Closed-Source Base Models for Information Extraction

Model Type	Model/Approach	Performance Gap to Best (%)	Correctly Predicted (%)
Fine Tune Model (Open Source)	mistral-7B	0.00	83.69
	gemma-2-9B	-0.89	82.80
	llama-3.1-8B	-4.07	79.61
Base Model (Closed Source)	gpt-4o-mini (CoT-Few-shot)	-19.49	64.20
	gpt-4o-mini (Few-shot)	-22.29	61.40
	gpt-4o-mini (CoT)	-23.56	60.12
	gpt-4o-mini (One-shot)	-23.82	59.87
	gpt-4o-mini (Zero-shot)	-34.77	48.91

gemma 2 9B and llama 3 1 8B, with slight performance gaps. In contrast, the closed-source gpt-4o-mini model, evaluated across various prompting strategies, showed noticeably lower performance. The CoT Few-shot approach led the closed-source models with 64.20% accuracy, while the Zero-shot strategy had the lowest accuracy at 48.91%. The performance gap to the best-performing model was smallest for mistral 7B and significantly larger for the base gpt-4o-mini models, especially with simpler prompting methods. These results highlight the superior performance of fine-tuned open-source models over closed-source base models in information extraction tasks.

5.2 Phase 2: Performance of Construction Activity Sequences (CAS)

As shown in Figure 5, the evaluation of construction activity sequence generation using BERTScores metrics Precision, Recall, and F1-score primarily focused on the performance of fine-tuned SLMs, specifically mistral-7B, gemma-2-9B, and llama-3.1-8B. Among these models, mistral-7B demonstrated the most balanced and consistent performance across all evaluation metrics, indicating its relative effectiveness in accurately generating and predicting construction activity sequences. In comparison, gemma-2-9B and llama-3.1-8B exhibited lower performance, suggesting that these models require further fine-tuning and optimization to improve their sequence generation capabilities. To provide a benchmark, the gpt-4o-mini model was included in the evaluation and achieved the highest scores across all metrics, with a precision of 0.89, recall of 0.87, and an F1-score of 0.88. While these results highlight the strong performance of larger models in complex sequence generations, this study primarily focused on assessing and improving fine-tuned SLMs for their computational efficiency and scalability.

The similarity scores of generated construction activity sequences were evaluated to compare the performance of fine-tuned SLMs and the larger gpt-4o-mini model. As shown in Figure 6, the gpt-4o-mini model achieved the highest mean similarity score of 91.59%, highlighting its superior ability to generate sequences closely aligned with the reference data. Among the fine-tuned models, mistral-7B performed the best with a mean similarity of 81.43%, while llama-3.1-8B and gemma-2-9B recorded mean similarity scores of 78.20% and 77.68%, respectively.

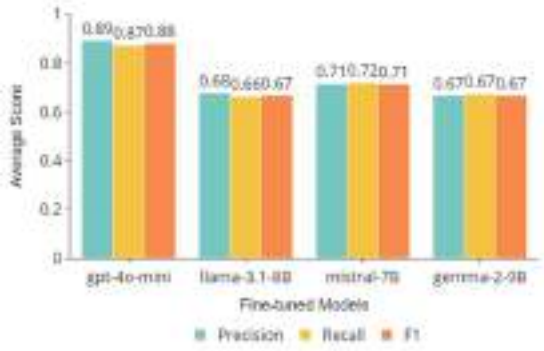


Figure 5. Evaluation of CAS Model: BERTScores (Precision, Recall, and F1) Analysis

These results demonstrate that while the fine-tuned SLMs performed reasonably well, they still lag the gpt-4o-mini in generating accurate and coherent construction activity sequences. However, the performance of the fine-tuned models, particularly mistral-7B, suggests that SLMs have strong potential for improvement.

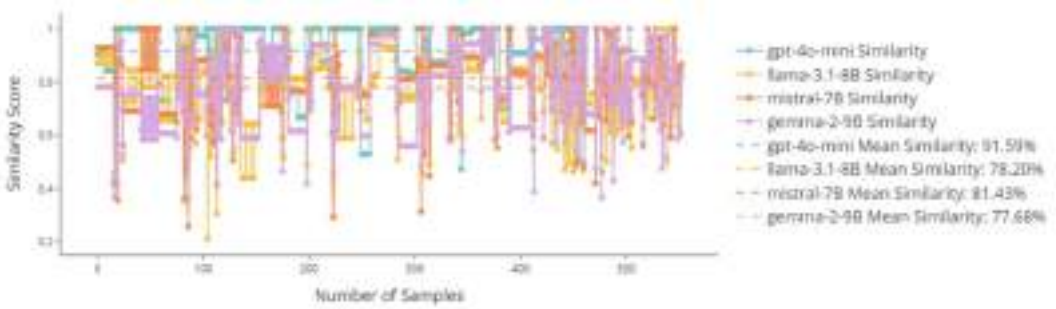


Figure 6. Evaluation of CAS Model: BERT Similarity Scores

5.3 Key Findings and Lesson Learned

In comparison to previous studies [1][9], which heavily rely on human-labeled schedule data and models such as LSTM, GNN, or BERT for learning activity patterns [9][2][3], this research adopts a more scalable and efficient approach using Small Language Models (SLMs) fine-tuned on minimally labeled data. Although pre-trained LLMs such as gpt-4o-mini were fine-tuned independently within this study and used as performance benchmarks, the primary emphasis remains on open-source SLMs. Among these, mistral-7B exhibited strong adaptability and effectiveness in handling construction planning tasks, offering a domain-aligned and practical alternative.

A key lesson learned is that targeted improvements in fine-tuning strategies such as advanced hyperparameter optimization and task-specific adaptation significantly enhance SLM performance. Emphasizing object and task identification during training allowed the models to infer contextual elements like location, improving the accuracy and relevance of generated sequences. The iterative process of hyperparameter tuning and model refinement proved crucial in balancing performance and computational efficiency.

These findings highlight the value of domain-specific data preparation, continuous model enhancement, and the combined potential of large and small models. With further optimization, fine-tuned SLMs can serve as effective, resource-efficient solutions for large-scale construction activity sequence generation.

5.4 Inferences of CIE and CAS model

Figure 7 showcases the optimal fine-tuned CIE model, mistral-7B, illustrating the differences in responses to user queries between the base model assistant and the fine-tuned model. Similarly, Figure 8 highlights the superior CAS fine-tuned model, mistral-7B, emphasizing the discrepancies in responses between the base and fine-tuned models. In both figures, errors are marked in red, while correct and newly identified sequences previously missed by the base model assistant are indicated in green. This comparison demonstrates the enhanced accuracy and predictive capabilities of the fine-tuned mistral-7B model.

System:
<Instruction_predicting_comprehensive_construction_activity_constituents>

User: "First Floor Column Shuttering Work"
Base Model Assistant: "EPC Type: Not specified, Work Item Domain: Construction, Location: First Floor, Object: Column, Task: Shuttering Work"
Fine-Tuned Model Assistant: first floor: location, column: object, shuttering: task, construction: epc type, structural: work item domain

Figure 7. Inference Outcomes Demonstrated by the CIE Model

System:
<Instruction_predicting_sequence_and_missing_activity>

Example: Same construction work item domain structural
User: "first floor column reinforcement,..., first floor slab formwork"
Base Model Assistant: "first floor column reinforcement, first floor column pouring, first floor beam reinforcement, first floor beam pouring, first floor slab formwork"
Fine-Tuned Model Assistant: first floor column reinforcement, first floor column formwork, first floor column pouring, first floor slab formwork

Example: Different construction work item domain finishing to mep
User: User: "Ceiling Plastering,..., HVAC testing"
Base Model: Based on the typical order of construction activities, ...could include:....."Ceiling Plastering, Wiring installation, Insulation installation, Drywall installation, Painting, Trim installation, Flooring installation, HVAC testing"
Fine-Tuned Model Assistant: Ceiling Plastering, Ceiling Priming, Ceiling Sanding, Ceiling Painting, HVAC Duct Installation, HVAC Duct Insulation, HVAC Duct Testing, HVAC Testing

Figure 8. Inference Outcomes Demonstrated by the CAS Model

6 Conclusion

This research presents a two-phase AI-driven approach to enhance automated construction project planning using SLMs for activity sequencing and missing task prediction. In the first phase, prompt engineering extracts and organizes key information from historical project schedules to produce structured outputs. Due to inconsistencies and lower accuracy, an SLM was fine-tuned with domain-specific data to improve extraction. In the second phase, another SLM, trained on the refined dataset, learns activity patterns to generate optimized sequences and identify missing tasks. The gpt-4o-mini model served as a benchmark and achieved the highest accuracy. However, fine-tuned models like mistral-7B delivered promising results with notable advantages in efficiency, scalability, and security.

A key advantage of SLMs is their enhanced data privacy and security. Once fine-tuned, these models can be deployed and used locally without relying on external servers or cloud-based services, ensuring sensitive construction data remain secure and protected. This makes SLMs particularly suitable for industries like construction, where data confidentiality is critical.

Future work will address unstandardized schedule data, generate synthetic data for diverse projects, and enhance reasoning in models for better understanding of scheduling logic. Expanding datasets, adopting open-source models, and integrating with BIM will further improve performance and efficiency in AEC project management.

Acknowledgements

The authors thank the National Science and Technology Council, Taiwan, for supporting this research through research grant Nos. NSTC 112-2221-E-002-078-MY3.

References

[1] Amer, F., & Golparvar-Fard, M. (2019). “Formalizing Construction Sequencing Knowledge and Mining Company-Specific Best Practices from Past Project Schedules.” Proceedings of ASCE International Conference on Computing in Civil Engineering, ASCE, Atlanta, GA, 215-223.

[2] Amer, F., & Golparvar-Fard, M. (2021). “Modeling Dynamic Construction Work Template from Existing Scheduling Records via Sequential Machine Learning.” Advanced Engineering Informatics, 47, 101198.

[3] Amer, F., Jung, Y., & Golparvar-Fard, M. (2021). “Transformer Machine Learning Language Model for Auto-alignment of Long-term and short-term Plans in Construction.” Automation in Construction, 132.

[4] Amer, F., Koh, H. Y., & Golparvar-Fard, M. (2021). “Automated Methods and Systems for Construction Planning and Scheduling: Critical Review of Three Decades of Research.” Journal of Construction Engineering and Management, 147(7), 03121002.

[5] Changali, S., Mohammad, A., & Van Nieuwland, M. (2015). “The Construction Productivity Imperative.” McKinsey Quarterly, 1-10.

[6] Chen, H., Zhang, Y., Zhang, Q., Yang, H., Hu, X., Ma, X., Yanggong, Y., & Zhao, J. (2023). “Maybe Only 0.5% Data is Needed: A Preliminary Exploration of Low Training Data Instruction Tuning.” arXiv, cs.AI, 2305.09246.

[7] Doloi, H. (2013). “Cost Overruns and Failure in Project Management: Understanding the Roles of Key Stakeholders in Construction Projects.” Journal of Construction Engineering and Management, 139(3), 267-279.

[8] Faghihi, V., Nejat, A., Reinschmidt, K. F., & Kang, J. H. (2015). “Automation in Construction Scheduling: A Review of the Literature.” International Journal of Advanced Manufacturing Technology, 81, 1845-1856.

[9] Hong, Y., Xie, H., Hovhannisyan, V., & Brilakis, I. (2022). “A Graph-based Approach for Unpacking Construction Sequence Analysis to Evaluate Schedules.” Advanced Engineering Informatics, 52, 101625.

[10] Pal, A., Lin, J. J., & Hsieh, S.H. (2023). “Schedule-Driven Analytics of 3D Point Clouds for Automated Construction Progress Monitoring.” Proceedings of ASCE International Conference on Computing in Civil Engineering 2023, ASCE, Reston, VA, 412.

[11] Singh, A. K., Pal, A., Kumar, P., Lin, J. J., & Hsieh, S. H. (2023). “Prospects of Integrating BIM and NLP for Automatic Construction Schedule Management.” Proceedings of the International Symposium of Automation and Robotics in Construction, Madras, TN, India, Vol. 40, 238-245.

[12] Singh, J., & Anumba, C. J. (2022). “Real-Time Pipe System Installation Schedule Generation and Optimization Using Artificial Intelligence and Heuristic Techniques.” Journal of Information Technology in Construction (ITcon), 27, 173-190

[13] Zhang, S., Dong, L., Li, X., Zhang, S., Sun, X., Wang, S., Li, J., Hu, R., Zhang, T., Wu, F., & Wang, G. (2023). “Instruction Tuning for Large Language Models: A Survey.” arXiv preprint arXiv:2308.10792.

[14] Zheng, J., & Fischer, M. (2023). “Dynamic Prompt-based Virtual Assistant Framework for BIM Information Search.” Automation in Construction, 155,105067.

Development of a Petrochemical Plant Pipe Counting Method Based on Improved Mask R-CNN and Transformer

Rong-Lu Hong[1], Jin-Bin Im[1], Sang-Jun Park[1], En-Lian Zhang[1], Nomgon Ochirsuren[1], Young-Gun Beak[1], Shin-Hyun Kang[1], Seong-Won Chung[1], Cheol-Su Lee[2], Seunghyeon Wang[1] and Ju-Hyung Kim[1]

[1] Department of Architectural Engineering, Hanyang University, Seoul, Republic of Korea
[2] UND Co., Ltd., Suseong-gu, Daegu, Republic of Korea
abraham0814@hanyang.ac.kr, jbim0202@hanyang.ac.kr, sangjpark@hanyang.ac.kr, eunryeon@hanyang.ac.kr, nomgon.ochirsuren53@gmail.com, byggun222@hanyang.ac.kr, shinhuny@naver.com, jsw7019@naver.com, bruce@unde.co.kr, wsh1019@hanyang.ac.kr , kcr97jhk@hanyang.ac.kr

Abstract

In petrochemical plant construction, pipe installation critically the affects overall progress and costs, making accurate on-site pipe quantification essential. However, the wide range of pipe diameters, high-density layouts, and complex site conditions poses significant challenges for automated counting methods. To address these issues, this study constructs the "PetroPipe" dataset, comprising 2,380 images (340 originals and 2,040 augmented) with pipes ranging from 0.75 to 60 inches in diameter under various resolutions. Multiple data augmentation strategies and a multi-stage training process enhance robustness under dynamic construction conditions. Building on this dataset, a pipe counting method is proposed that integrates instance segmentation with a counting head and a transformer encoder. Based on the Mask R-CNN framework, the model applies a discrete-bin classification approach to achieve accurate global quantity predictions, while maintaining high-quality instance-level detection and segmentation. The experimental results demonstrate that the developed model outperforms the standard Mask R-CNN model on the test set, increasing the counting accuracy from 65.55% without a counting head to 85.50%. When evaluated on an unlabelled 18-second video captured in a real-world scenario, it attained an accuracy of approximately 98.1%, demonstrating its capacity to adapt to practical industrial conditions. These results indicate that the proposed approach effectively accommodates varying pipe diameters, dense layouts, and complex backgrounds. Consequently, it provides a scalable, efficient pipe counting, progress monitoring, and resource optimization in petrochemical construction environments.

Keywords –

Petrochemical plant pipe; Instance Segmentation; Counting; Transformer

1 Introduction

During construction of petrochemical plants, pipe installation is a crucial element of the overall construction schedule. It is a prominent component of workload and cost distribution [1]. Statistical evidence suggests that pipe installation often dictates the overall project timeline, with materials and labor accounting for approximately 71% of the total pipeline construction costs [2]. Given the large volumes and varied specifications of piping materials, this key stage significantly influences project outcomes. To ensure smooth progress, project managers must accurately and efficiently record and verify the on-site quantities, types, specifications, and manufacturers of these materials [3].

However, the widely used material management approaches, including barcodes, radio frequency identification (RFID) tags, and manual counting, frequently face significant challenges in practice. On the one hand, manual verification and data entry are time-consuming and labour-intensive, and they are easily affected by human errors and environmental interferences. In addition, RFID and barcode systems require special hardware and pre-applied labels, conditions that cannot always be satisfiedin complex and rapidly changing construction environments [4]. Real-time tracking and frequent updates of material information are often required. Under these conditions, traditional methods fail to provide the speed and flexibility required for modern detail-oriented project management [3].

In this context, rapid advances in computer vision and image processing technologies have provided new opportunities for digitalization and intelligent management of construction sites. Previous studies have successfully applied deep learning-based visual analysis methods to tasks such as crack detection, structural damage assessment, and construction quality control [5,

6], indicating that using image data for automated construction material counting is both feasible and competitive. Compared with conventional labelling methods, image processing requires no additional hardware markers and can extract information in real time from video streams, thereby significantly improving operational efficiency and data accuracy [5]. Nevertheless, the direct application of image processing techniques to pipe counting in petrochemical plants remains challenging. Piping materials are frequently stacked in high densities, differ in size and shape, and are often partially obscured [3]. Consequently, traditional detection methods that rely solely on bounding boxes struggle to delineate pipe boundaries accurately in dense scenes, leading to counting errors [7]. Moreover, the complexity of the environment and variability in data distributions highlight the urgent need to enhance model robustness and generalization.

To address these challenges, instance segmentation provides a more direct technical pathway for improving counting accuracy and detection reliability. Among these, Mask R-CNN, as a classic and widely used deep-learning method, has been extensively explored and validated in counting tasks for plants, vehicles, and animals. However, it exhibits insufficient accuracy and limited generalization when handling complex scenarios characterized by high target density, significant scale variations, and severe occlusions [8–10]. By contrast, transformers have demonstrated initial success in capturing global context, integrating multi-scale information, and enhancing feature representations in other counting tasks [11, 12]. Incorporating a transformer into the instance segmentation and counting framework may improve the adaptability and generalization of the model in petrochemical plant settings characterized by diverse pipe diameters, high densities, and significant scale variations. This study addresses the current gaps in the literature related to these aspects.

This study extends the classical Mask R-CNN instance segmentation model by integrating a Transformer-based counting head into the existing framework. This integration enables precise multi-object instance segmentation while directly estimating the overall number of targets. Using a multi-task joint training strategy, the model achieves coordinated optimization of detection, segmentation, and counting at the feature level, thereby improving its robustness and generalization in complex construction scenarios. Moreover, the proposed approach adapted for other types of densely distributed construction materials and can be seamlessly integrated into on-site digital information management systems, ultimately providing more efficient, accurate, and intelligent support for project scheduling, resource allocation, and cost optimization.

2 Related Work

Over the past decade, traditional computer vision techniques based on image processing have progressed in counting construction materials. Thammasorn et al. (2013) proposed a template matching method for counting rectangular and circular steel pipes, achieving high accuracy when the camera angle is relatively fixed and the main targets occupy most of the image [13]. Ghazali et al. (2017) combined the Hough transform with post-processing and morphological operations to detect and count circular and rectangular steel materials, thereby demonstrating initial practical outcomes [14]. Liu et al. (2019) employed the Prewitt operator to extract oriented gradient features and used support vector machine (SVM) to locate and count bundled rebars, achieving an average accuracy of approximately91% in their experiments [15].

However, traditional image processing and feature engineering methods are sensitive to environmental conditions. Factors such as the camera angle, illumination changes, steel material size, cross-sectional shape, and background clutter, can cause significant performance fluctuations [3]. In actual construction sites, uneven lighting, weather variations, and increasingly complex background textures further compromise the stability and generalizability of such methods [5]. This highlights the difficulty of designing sufficiently robust features that can handle the diversity and uncertainty of on-site environments using traditional approaches.

In recent years, deep learning methods, including convolutional neural networks (CNNs), have gained wide acceptance and validation in the visual analysis of construction engineering [16]. Compared with traditional image processing techniques, these approaches can handle complex scenes more effectively while improving both accuracy and robustness, thereby offering strong technical support for construction informatization and intelligent decision-making [5, 17]. Motivated by these advances, researchers have started exploring deep learning techniques for automated counting of construction materials. For example, Fan et al. (2019) reframed the task of counting rebars as an image classification problem by classifying local patches within an image [17]. However, this approach requires the examination of each patch individually, which prolongs the inference time. In addition, its reliance on parameters such as target radius restricts its applicability in dynamic and ever-changing construction environments. In efforts to improve practical performance, Wang et al. (2023) integrated UAV imagery with Faster R-CNN and trained 384 model variants on both original and augmented datasets collected under real construction conditions [5]. Their best model achieved an accuracy of 94.16% at AP50. Similarly, Li et al. (2022) developed a precise counting model using YOLOv3 and released a large-

scale steel pipe dataset covering diverse on-site conditions, thereby providing abundant resources for research [3].

To achieve deeper feature representation and stronger global context capture, Bai et al. (2023) introduced the CounTr model, which uses a pure Transformer architecture with multi-head self-attention. By using a hierarchical self-attention decoder to fuse multi-scale and global information, CounTr delivered significantly better performance than traditional CNN-based approaches on both the UCF_CC_50 high-density crowd dataset and the FDST diversified dataset [11]. Feng et al. (2023) presented a pig counting and localization method based on density maps. Their approach employs a lightweight hybrid network that integrates selective convolution and a Vision Transformer to generate density maps and then uses an improved K-means algorithm to locate individual targets on these maps. The experimental results showed a low counting error (MAE=0.726), along with precision and recall rates of 88.22% and 86.02%, respectively, making it suitable for large-scale and crowded scenarios [12]. However, in the context of pipe counting, the current research remains largely confined to object detection and classification, with insufficient in-depth exploration of the integration of instance segmentation and Transformer-based techniques. Instance segmentation provides independent, pixel-level masks and precise boundary information for each pipe, thereby effectively enhancing the counting accuracy and stability in complex backgrounds [7]. It also provides a solid foundation for subsequent pipe layout planning and optimization. Incorporating Transformers into this framework is expected to strengthen the capture of the global context and integration of multi-scale features. These improvements facilitate better adaptation and generalization to the challenging conditions present in petrochemical plants, including a wide range of pipe diameters, high densities, and significant scale variations.

In summary, applying a counting method based on instance segmentation and Transformers for pipe counting in petrochemical plants is both a logical and potentially valuable. This approach not only compensates for the shortcomings of existing research under complex industrial conditions, but can also provide more comprehensive and robust technical support for achieving more accurate counting and spatial layout optimization in future industrial applications.

3 Material and Methods

3.1 Data collection

The effective training of deep learning models often relies on large and diverse image datasets. However, research on counting construction materials, particularly pipes, has been limited by the lack of large-scale data resources.

To address this shortcoming, this study constructed a comprehensive dataset designed for a petrochemical plant environment, referred to as the “PetroPipe” dataset. This dataset is publicly available on Kaggle and comprises 340 images with 14,855 pipe instances [18]. All images were captured with an iPhone 14 Pro under various conditions and from multiple angles, increasing the data diversity and improving the adaptability of the model. As shown in Figure 1, the pipes cover diameters ranging from 0.75 60 inches, ensuring that the trained models will be more robust in real-world scenarios. The dataset includes images of three different resolutions (3061×3024, 3024×4032, and 4032×3024), reflecting the variety of image sizes encountered in industrial environments. The inclusion of multiple resolutions helps the model maintain strong generalization performance under varying imaging conditions.

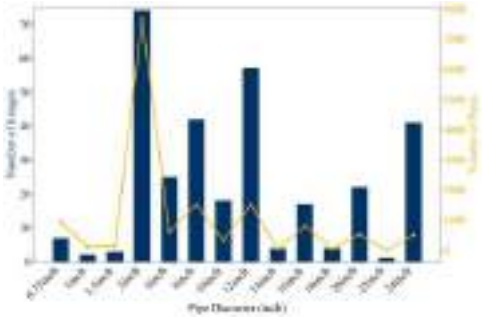


Figure 1. Distribution in the PetroPipe dataset

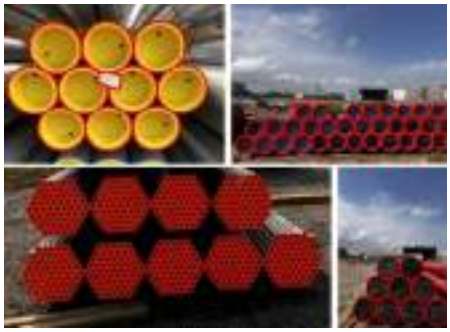


Figure 2. Example of polygon-based annotations

For data annotation, polygon-based annotations were created for each pipe instance using the Labelme tool. These annotations were then converted into the COCO format to ensure direct compatibility with the Mask R-CNN model [19]. Figure 2 shows an example of an annotated image. The performance of the model was evaluated under near-real-time monitoring conditions using an 18-second video containing multiple pipe instances

3.2 Data augmentation

Vrious data augmentation operations were applied to each original image to enhance the robustness of the model in complex real-world scenarios. As shown in Figure 3, these include blur, brightness, color, contrast, horizontal flip, and sharpness adjustments. Such strategies effectively increase the diversity of the dataset, enabling the model to adapt better to different lighting conditions, camera angles, and image quality variations [20]. A total of, 2,040 augmented images were generated, and together with the original 340 images, yielded a dataset of 2,380 images. The dataset was then split into training, validation, and test sets in a ratio of 6:2:2 to ensure reliable training and evaluation.

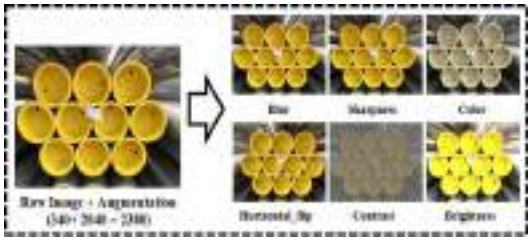


Figure 3. Examples of the data augmentation operations

3.3 Model Architecture and Counting Head Design

The model architecture and experimental workflow are shown in Figure 5. To achieve rapid and accurate global object counting, this study integrates a counting head into the standard Mask R-CNN framework. This integration enables the model to produce high-quality instance-level segmentation while estimating the total number of targets in the entire image from a global perspective. The base structure uses a ResNet-50 architecture combined with a feature pyramid network (FPN) as the backbone. And a region proposal network (RPN) generates initial proposals. Each proposal undergoes region of interest(ROI) align and box head feature extraction, resulting in a 1024dimensional feature representations. By leveraging these representations, the box regression and classification branches provide precise object detection and localization, whereas the mask branch delivers instance-level semantic segmentation [8]. However, the conventional Mask R-CNN architecture does not directly count the total number of objects across an entire image.

As shown in Figures 4 and 5, the proposed counting head is introduced at the output of the box head to address this limitation. Instead of simply averaging the ROI features, a global token is extracted from a lower-level FPN feature map (P2 layer) using global average pooling. The global token is concatenated with all the ROI features (ROI Tokens) to form a sequence comprising one global token and multiple ROI tokens. The sequence is then fed into a Transformer encoder, where multi-head self-attention and feed-forward layers achieve deep fusion of global and instance-level features. In the Transformer output's, the first token corresponds to the global token, and can be regarded as a fused global feature representation.

A linear classifier (count_classifier) subsequently maps the fused global features to scores over a set of predefined discrete intervals representing plausible target counts (e.g. dividing the range of 0–50 into several bins). Instead of predicting a continuous number directly, the model selects the bin with the highest score to infer the total object count. This classification-based strategy stabilizes the training process and enhances interpretability. Moreover, the counting head imposes a minimal additional computational overhead, offering an efficient and innovative solution that integrates both local and global features for object detection and counting tasks.

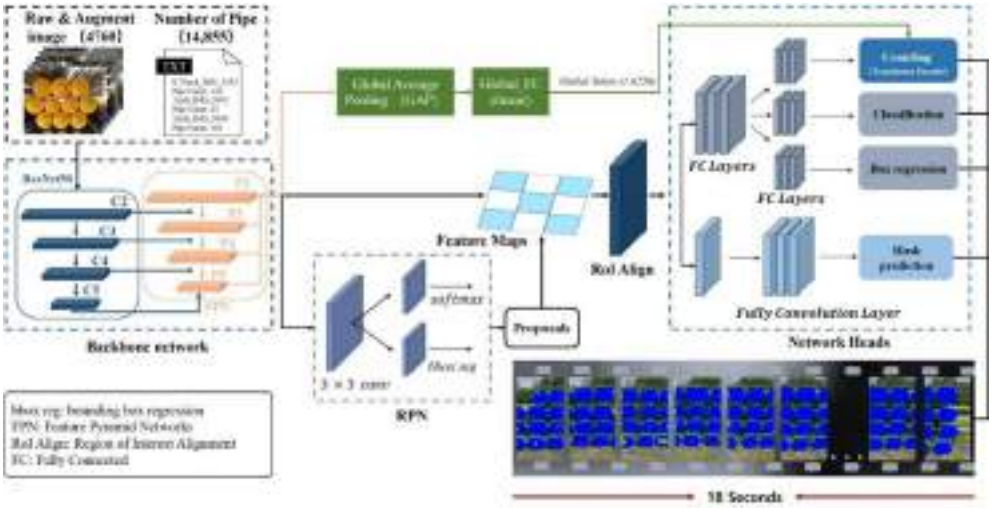


Figure 5. Model architecture and experimental workflow

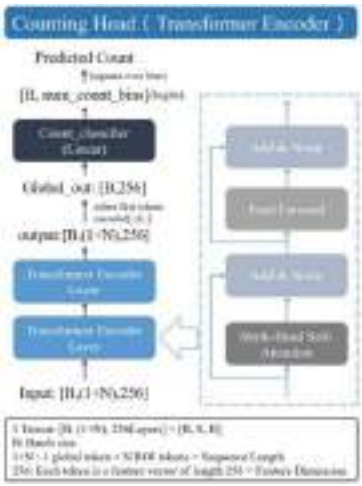


Figure 4. Internal structure of the counting head

3.4 Experimental Settings

This study employed a multi-stage multi-task optimization strategy when integrating the counting head, ensuring steady improvements in the detection and segmentation performance. Initially, the backbone weights were obtained from a ResNet-50+FPN model pre-trained on the COCO dataset. To maintain stable convergence and avoid abrupt parameter shifts, the backbone parameters were frozen during the early training stage (the first two epochs), and only the ROI heads and counting head parameters were updated. At this point, the loss weight for the counting branch was set relatively low (0.1) to prevent the counting task from dominating the training process too soon.

Once the detection, segmentation, and counting tasks achieved a preliminary balance, as indicated by completion of the freezing phase, the backbone parameters were unfrozen, allowing the entire network to be fine-tuned. This step enabled the model to adapt better to the characteristics of the target data and enhance its overall performance. Thus, to overcome the two major challenges in multitask training, namely, the interference of the counting branch with the overall network and the imbalance of loss weights, we ensured in the early stage, through a lower counting loss weight and dynamically adjusted learning rates, that the counting head does not cause excessive gradient impact when jointly training with the detection and segmentation tasks. In the later stage, when the backbone was unfrozen and fine-tuned jointly, the counting loss weight was moderately increased and a staged learning rate schedule was adopted, enabling the counting head to acquire richer semantic information without undermining the previously learned detection and segmentation capabilities. During training, the AdamW optimizer and a StepLR scheduler were employed to gradually reduce the learning rate in phases, ensuring a smooth transition from rapid initial convergence to more refined adjustments. Owing to this multi-stage approach, the model preserved the detection and segmentation improvements after incorporating the counting head. In addition, it achieved a significant increase in the global counting precision, demonstrating effective multi-task optimization. Details of the experimental environment and parameter settings are listed in Table 1.

Table 1. Details of the experimental environment and parameter settings

Parameter	Description
Operating System	Windows 11 Pro
Hardware	GPU: RTX 3060 (12GB), CPU: AMD Ryzen 5 5600 RAM: 32GB
Development Environment	PyTorch 2.5.1, Python 3.9, CUDA 12.4
Optimizer	AdamW (lr=1e-5, weight_decay=1e-4)
LR Scheduler	StepLR: halving lr every 5 epochs; total 100 epochs
Initial Weights	MaskRCNN_ResNet50_FPN_ Weight. COCO_V1
Training Strategy	Batch=2; First 2 epochs: freeze backbone (train ROI & counting head)

3.5 Evaluation Metrics

To objectively evaluate the performance of the instance segmentation model proposed in this paper, we define the evaluation metrics at two scales: bounding box and mask. Specifically, the metrics average precision (AP) and mean intersection over union (mIoU) were used to assess the model's overall detection ability and its fine localization performance. We calculated $\boldsymbol{AP_{50}}$ and $\boldsymbol{AP_{75}}$ at IoU thresholds of 0.5 and 0.75, respectively. The AP metric is defined as

$$\boldsymbol{AP} = \int_{0}^{1} \boldsymbol{p}(\boldsymbol{r})\boldsymbol{dr} \quad (1)$$

where $\boldsymbol{p}(\boldsymbol{r})$ is the interpolated precision value under the precision–recall curve. The AP computed at the bounding box scale is used to measure the overall localization ability of the model, whereas the AP at the mask scale further reflects the fine segmentation performance of the model at pixel level [5,6]. The AP metric has been widely applied in the field of instance segmentation because it effectively evaluates the overall detection and fine localization capabilities of the model, and intuitively and comprehensively reflects the recognition performance of the model under varying

degrees of strictness [6,8]. In addition, we computed mIoU at both the bounding box and mask scales using the general formula defined as follows:

$$\boldsymbol{mIoU} = \frac{1}{N}\sum_{i=1}^{N}\frac{\boldsymbol{R}_{pred}^{i} \cap \boldsymbol{R}_{gt}^{i}}{\boldsymbol{R}_{pred}^{i} \cup \boldsymbol{R}_{gt}^{i}} \quad (2)$$

where $\boldsymbol{R}_{pred}^{i}$ and $\boldsymbol{R}_{gt}^{i}$ are the predicted and ground-truth regions for the ith target, respectively. This formula is applicable to both the bounding box and mask scales. At the bounding box scale, the mIoU emphasizes the overall localization ability of the model, whereas at the mask scale, it more precisely reflects the model's ability to delineate target boundaries at pixel level [5,8]. Therefore, computing the mIoU at both scales can comprehensively demonstrate the performance advantages of the model at different granularity levels. To evaluate the object counting performance, we compared the number of instances predicted by the model with the counting module to those predicted by the native detection output without parameter adjustments to assess the contribution of the counting module to counting accuracy and its impact on overall model performance. We designed a multitask joint loss function that integrates detection (classification and bounding box regression), mask segmentation, and object counting.

$$\boldsymbol{L_{Total}} = \boldsymbol{L_{cls.}} + \boldsymbol{L_{bbox\,reg}} + \boldsymbol{L_{mask}} + \boldsymbol{\lambda L_{count}} \quad (3)$$

where λ is used to control the influence of the counting loss term, thereby achieving collaborative optimization among the different task losses. Compared with other potential metrics, such as the Dice coefficient or pixel accuracy, the AP and mIoU metrics offer more comprehensive, intuitive, and stable advantages, enabling an effective and detailed evaluation of the precision and generalization performance of the instance segmentation model.

4 Results

4.1 Results of Training and Validation

During model training, both the training and validation losses were monitored, as shown in Figure 6. Although the initial plan was to train for 100 epochs, the model adapted quickly. Initially, the training loss was 1.11, and the validation loss was 0.71. By the 5th epoch, the training loss decreased to 0.23, and the validation loss was 0.22, indicating effective feature learning. After the 15th epoch, both training and validation losses stabilized at 0.14. However, given that the transformer model is strongly dependent on global information and positional features, it generally requires a longer training period to fully integrate the features and achieve stable convergence. Therefore, in this experiment, the training process was extended to 100 epochs to ensure that the model achieved stable, low-loss convergence and exhibited strong generalization capabilities under the chosen optimization strategy.

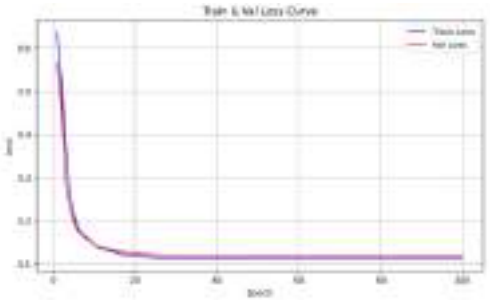


Figure 6. Training and validation losses

4.2 Model Evaluationusing Test Images

Comparative experiments were conducted to evaluate the performance of the standard Mask R-CNN (ResNet-50) and the proposed model under identical datasets and configuration settings. As shown in Table 2, the test results indicate that our model slightly outperformed the standard Mask R-CNN when evaluated at both the bounding box and mask scales. Although the improvements are modest, these results demonstrate that our model exhibits greater stability in terms of capturing object details and segmentation accuracy.

Table 2. Comparison of model performance

Model	Mask R-CNN		Our Model	
Label	BBox	Mask	BBox	Mask
AP 50	97.71%	97.26%	98.07%	97.45%
AP 75	97.50%	96.98%	97.89%	97.20%
mIOU	85.56%	76.47%	87.32%	76.54%
Counting Accuracy	65.55%		85.50%	

For the counting task, Mask R-CNN employs a shared confidence scoring mechanism for detection boxes and segmentation masks; thus, the counting accuracy is computed as a unified metric rather than separately for different scales. Furthermore, no confidence threshold was set in this study to ensure that all predicted instances were counted. The experimental results showed that the counting accuracy of the standard model was 65.55%. By contrast, Our model incorporates an independent counting module, which integrates a transformer encoder to consolidate global information and instance features, and employs a discrete bin classification strategy to directly output the instance count. Owing to this improvement, the counting accuracy of the new model increased to 85.50%, indicating that the introduced counting head provides more stable global statistical information for the model, thereby overcoming the potential instability of relying solely on detection outputs.

However, as shown in Figures 7(c) and 7(d), the segmentation and counting accuracies of our model remained limited in scenarios with extremely dense and small-diameter pipes. Such conditions involve blurred boundaries, severe overlap, and low feature distinctiveness, thereby increasing the difficulty of detection and counting. By contrast, as shown in Figures 7(a) and 7(b), our model achieved nearly 100% counting accuracy in scenarios with fewer pipes. These findings indicate that although the proposed approach is effective in less congested environments, further improvements are required to address highly crowded and complex conditions.

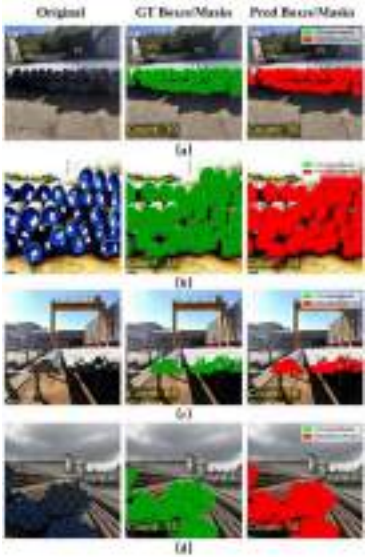


Figure 7. Our model's pipe segmentation and counting examples under different densities

4.3 Video Testing

To further assess the practicality and adaptability of the improved model, inference was conducted on an unlabeled, real-world video (608×1080, 18s, 30 FPS) containing 460 pipes characterized by a high density and small object size. Operating in the evaluation mode with float32 precision, the model processed each frame without additional training. An IoU-based matching strategy was employed to avoid multiple counts of the same pipe, either by updating existing Identifications (IDs) or assigning new IDs as required.

By applying frame-by-frame inference, the model detected 469 pipes, achieving approximately 98.1% counting accuracy compared with the actual total of 460. This result demonstrates that even without annotated training data, the counting head approach can significantly enhance the counting precision in real-world scenarios. Although the average per-frame inference time was about 0.4176 seconds (approximately 2.39 FPS) and thus did not meet strict real-time requirements, the method maintained a relatively high degree of accuracy under dense conditions.

The visualization results (Figure 8(a) show the initial frame and Figure 8(d) shows the final frame) indicate stable detection and counting performance, even in complex, tightly arranged scenes. These findings provide a valuable reference for future research and practical applications of unannotated video monitoring.

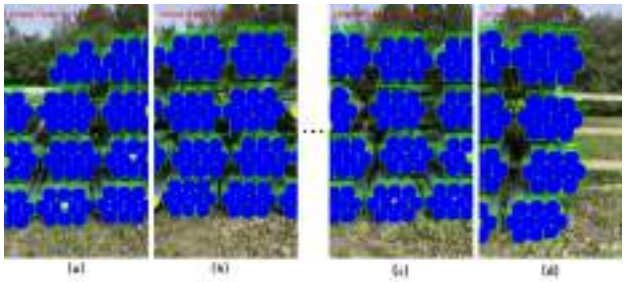


Figure 8. Sample of sequential frames

5 Conclusions

This study focused on pipe counting issues in petrochemical plants. It proposes a counting method that integrates instance segmentation with a Transformer, and through a multi-task joint training strategy, achieves outstanding performance in complex, densely populated, and highly variable construction site environments. The experimental results showed that the model achieving an AP50 of 98.07% (BBox) / 97.45% (mask), an AP75 of 97.89% (BBox) / 97.20% (mask), and an mIoU of 87.32% (BBox) / 76.54% (mask), and inceased the counting accuracy from 65.50% (without the counting head) to 85.50%. When applied to real-world, unlabeled video data, the model demonstrates strong generalization capability and achieves a counting accuracy of about 98.1%, confirming its practical value in real scenarios. Nevertheless, certain limitations remain in practical applications. Although the proposed model demonstrated robust generalization ability and counting accuracy in practical application scenarios, there is still room for improvement in the detection performance of densely packed, small-diameter targets and inference speed. Future research, in addition to further expanding the dataset scale and diversity, should consider optimizing the internal structure of transformers by incorporating multiscale attention mechanisms or local-global feature fusion strategies to enhance the recognition of dense and small-sized targets. The current backbone network can be replaced with a more efficient and lightweight CNN or transformer architectures (e.g., EfficientNet, Swin-Tiny, and MobileViT) and integrated with model compression and pruning techniques to further reduce computational costs and inference time while preserving accuracy, thus better addressing the practical deployment requirements

in industrial settings. This study provides a technological foundation for pipe counting and information management in petrochemical facilities, establishing a solid basis for further improvements in speed, data expansion, and framework optimization, thereby contributing to the intelligent and automated advancement of this field.

Acknowledgement

This research was funded by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (Ministry of Science and ICT) (grant number RS202400356697).

References

[1] G. Qin, H. Ding, X. Liu, H. Zhang, and G. Ma, "Construction management of petrochemical process pipeline installation," *Smart Construction Research*, 2018

[2] Z. Rui, P. A. Metz, D. B. Reynolds, G. Chen, and X. Zhou. Historical pipeline construction cost analysis. *International Journal of Oil, Gas and Coal Technology*, 4(3):244–263, 2011.

[3] Y. Li and J. Chen. Computer vision–based counting model for dense steel pipe on construction sites. *Journal of Construction Engineering and Management*, 148(1):04021178, 2022.

[4] K. K. Duan and S. Y. Cao. Emerging RFID technology in structural engineering–A review. *In Structures*, 28: 2404–2414, 2020.

[5] S. Wang, M. Kim, H. Hae, M. Cao, and J. Kim, The Development of a Rebar-Counting Model for Reinforced Concrete Columns: Using an Unmanned Aerial Vehicle and Deep-Learning Approach, *Journal of Construction Engineering and Management*, 149(11), 2023.

[6] Z. Liu, J. K. Yeoh, X. Gu, Q. Dong, Y. Chen, W. Wu, and D. Wang. Automatic pixel-level detection of vertical cracks in asphalt pavement based on GPR investigation and improved mask R-CNN. *Automation in Construction*, 146:104689, 2023.

[7] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask r-cnn. *In Proceedings of the IEEE International Conference on Computer Vision*, pages 2961–2969, Venice, Italy, 2017.

[8] Y. Shao, X. Guan, G. Xuan, H. Liu, X. Li, F. Gu, Z. Hu, Detection of Straw Coverage under Conservation Tillage Based on an Improved Mask Regional Convolutional Neural Network (Mask R-CNN). *Agronomy*, 14:1409, 2024.

[9] B. Xu, W. Wang, G. Falzon, P. Kwan, L. Guo, G. Chen, A. Tait, D. Schneider, Automated cattle counting using Mask R-CNN in quadcopter vision system, *Computers and Electronics in Agriculture*, 171:105300, 2020.

[10] H. Tahir, M. Shahbaz Khan, M. Owais Tariq, Performance Analysis and Comparison of Faster R-CNN, Mask R-CNN and ResNet50 for the Detection and Counting of Vehicles, *2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, pages 587–594, Greater Noida, India, 2021.

[11] H. Bai, H. He, Z. Peng, T. Dai, and S. H. G. Chan. Countr: An end-to-end transformer approach for crowd counting and density estimation. *In European Conference on Computer Vision*, pages 207–222, Cham Switzerland, 2022.

[12] W. Feng, K. Wang, and S. Zhou. An efficient neural network for pig counting and localization by density map estimation. 11: 81079–81091, *IEEE Access*, 2023.

[13] P. Thammasorn, S. Boonchu, and A. Kawewong. Real-time method for counting unseen stacked objects in mobile. *In 2013 IEEE International Conference on Image Processing*, pages 4103–4107, Melbourne, VIC, Australia, 2013.

[14] M. F. Ghazali, L. K. Wong, and J. See. Automatic detection and counting of circular and rectangular steel bars. *In 9th International Conference on Robotic, Vision, Signal Processing and Power Applications: Empowering Research and Innovation*, pages 199–207, Singapore, 2017.

[15] C. Liu, L. Zhu, and X. Zhang. Bundled round bars counting based on iteratively trained SVM. *In Intelligent Computing Theories and Application: 15th International Conference*, pages 3–6, Nanchang, China, 2019.

[16] S. Wang and J. Han, Automated detection of exterior cladding material in urban area from street view images using deep learning, *Journal of Building Engineering*, 96: 110466, 2024

[17] Z. Fan, J. Lu, B. Qiu, T. Jiang, K. An, A. N. Josephraj, and C. Wei. Automated steel bar counting and center localization with convolutional neural networks, *arXiv preprint*, 1906.00891, 2019.

[18] R.-L. Hong, S.-W. Chung, Y. Beak, PetroPipe, (n.d.).https://doi.org/10.34740/KAGGLE/DSV/10939352, Accessed: 03/10/2025.

[19] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman. LabelMe: a database and web-based tool for image annotation. *International Journal of Computer Vision*, 77:157–173, 2008.

[20] S. Wang, Evaluation of impact of image augmentation techniques on two tasks: Window detection and window states detection, *Results in Engineering*, 24: 103571, 2024.

Optimizing Structure from Motion Parameters for Volume Estimation of Construction and Demolition Waste

Ashwani Jaiswal [1] **Nikhil Bugalia** [1] **and Quang Phuc Ha** [2]

[1]Department of Civil Engineering, IIT Madras, India
[2]Faculty of Engineering Information and Technology, UTS, Australia

ce22d045@smail.iitm.ac.in, nbugalia@civil.iitm.ac.in, quang.ha@uts.edu.au

Abstract -

Reliable quantification of localized construction and demolition waste (CDW) stockpiles is crucial for promoting sustainable waste management and recycling practices. While prior Structure-from-Motion (SfM) research predominantly focuses on methodological parameters, comparatively limited research has addressed the role of environmental and capturing factors in small-scale reconstructions. This study investigates the accuracy of SfM-based 3D reconstructions for volumetric estimation of localized CDW stockpiles, benchmarking against a consumer-grade handheld LiDAR device. A comprehensive workflow was developed to achieve volume estimation from video sequences, encompassing data acquisition, pre-processing, processing, and post-processing steps. Unlike previous studies that typically prioritize high-resolution images and overlook camera positioning and lighting conditions, this study experimentally assesses these critical parameters. Results indicated minimal influence of lighting conditions, with optimal camera positioning identified as 1.5 times the stockpile height, a frame extraction rate of 1 FPS, and 720p resolution providing the most accurate reconstructions. Analysis of 30 stockpile samples demonstrated a strong correlation (R^2 = 0.929) between SfM and LiDAR volumes, demonstrating that SfM achieves volume accuracy comparable to LiDAR data while offering a faster (3.8 times) and cost-effective solution for the given experimental setup.

Keywords -

Construction and Demolition Waste; Structure from Motion (SfM); Waste Management; Point Cloud; Optimization.

1 Introduction

The global generation of over 10 billion tons of construction and demolition waste (CDW) annually presents a critical environmental and economic challenge, particularly when not effectively managed [1]. Accurate quantification of the localized CDW stockpile, which is often deposited illegally along the street, isolated land areas, and water bodies (an example shown in Figure 1), is essential for enabling efficient and improved waste management and reverse supply chain (RSC) [2]. Inefficiencies in waste collection operations, such as unknown quantity, location, and composition on the upstream side of the RSC, bring logistical complexities and a decline in the quality and quantity of recycled materials [3]. The unknown volumes of localized CDW stockpiles present a critical bottleneck among these challenges. Recent research has increasingly emphasized optimizing the RSC through technological interventions and digitization [4]. However, the existing research has primarily focused on minimizing waste generation, and the literature on technical interventions for other critical processes like waste collection and transportation is relatively scarce.



Figure 1. Examples of localized CDW.
(Source: Authors)

Accurately capturing a three-dimensional (3D) scene is vital to reliably quantifying the volume of a stockpile [5]. Point clouds have emerged as a widely adopted data format for representing 3D environments, providing real-world details, including spatial coordinates and color attributes. Recent advancements have leveraged diverse data acquisition devices, such as Terrestrial Laser Scanners, Unmanned Aerial Vehicles, Simultaneous Localization and Mapping, and RGB-D sensors [6]. However, these acquisition devices are often accompanied by operational challenges, extended processing, and high costs, limiting their applicability for small-scale applications like localized CDW stockpiles

and underscoring the urgent need for portable and scalable solutions.

Recent advancements in consumer-grade handheld LiDAR-equipped devices, such as the iPhone, have established their reliability and versatility as practical solutions for small-scale applications [7]. In addition to handheld LiDAR, Structure from Motion (SfM) is another significant technology that can be employed for small-scale applications [5]. SfM is a pivotal computer vision technique widely employed to reconstruct (3D) structures from two-dimensional (2D) image sequences and is frequently used in scientific research and engineering [8]. A wide range of SfM tools is available. In particular, the interface of Meshroom can be easily customized as it covers very minute details from a research point of view [9]. Generally, SfM is effective for 3D reconstruction. However, it is shaped by various influencing factors, like methodological, environmental, and capture parameters [10]. Among these, methodological parameters include factors related to the process of a tool, which stands out as the most refined, given the extensive research and advancements achieved over recent decades [8].

While significant progress has been made in improving SfM methodology, less attention has been given to two other key aspects: environmental and capture-related parameters. Environmental factors, like lighting and shadows, affect the scene's conditions. In contrast, capture-related factors, including the number of images, their quality, and camera setup, directly impact the accuracy of the SfM-generated point cloud [11]. A notable study focused on point clouds in tunnel environments, specifically examining the influence of lighting conditions and the number of images [10]. While it highlighted essential aspects, the study primarily addressed tunnel-specific scenarios and lacked consideration of broader applications, such as small-scale reconstructions. The existing body of literature recommends including more images and a higher percentage of overlap, which is necessary for effective 3D reconstruction [8]. However, these recommendations often lack specificity and fail to consider critical variables such as scene complexity and camera model used. In addition, image quality and resolution are rarely addressed in detail, particularly for small-scale, such as those involving localized CDW [11]. While high-quality images are generally assumed to be integral to the reconstruction process, the specific influence of these factors remains insufficiently explored. This highlights the pressing need for more precise, context-specific guidelines to inform 3D reconstruction practices across various applications.

To address the research gap, this study aims to address the following research question: *"How do environmental and capture parameters impact the quality and accuracy of SfM-based 3D point clouds for estimating localized CDW stockpile volumes from video sequences?"* By exploring this question, the study aims to extend an understanding of the influence of parameters on 3D reconstruction while offering practical understandings for optimizing volume estimation workflow. Accordingly, the study's primary objective is to investigate environmental and capture parameters affecting SfM-based point cloud accuracy for volume estimation of localized CDW stockpile compared with handheld LiDAR-based point cloud.

The structure of the paper is as follows. Section 2 outlines the proposed methodology. Section 3 presents the experimental design focusing on parameter influences. Section 4 presents the experimental results, discussion, and conclusions in Section 5.

2 Methodology

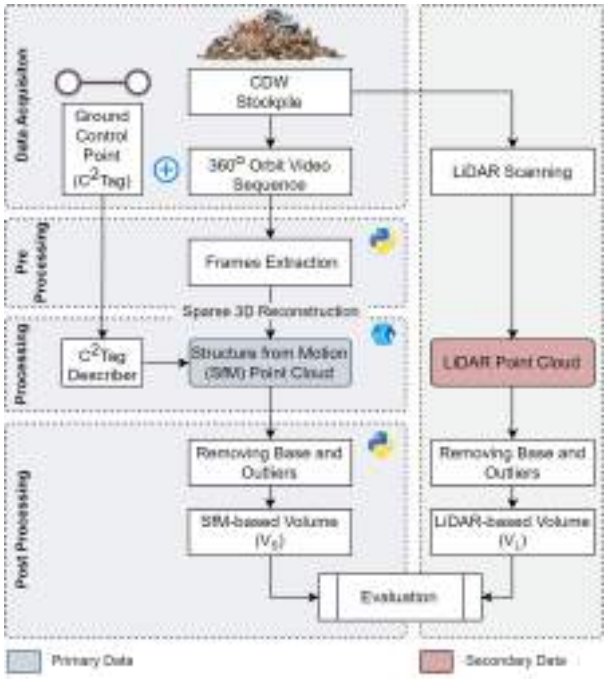


Figure 2. The overall workflow for CDW volume estimation using smartphone video sequence.

Figure 2 illustrates the proposed workflow for comparing the volume of CDW stockpiles from the SfM-based point cloud to the LiDAR-based point cloud. The process begins with a video sequence of the stockpile, captured using a smartphone camera as input. This video is transformed into a precise volume estimate through a systematic step sequence. Initially, individual frames are extracted from the video to facilitate sparse 3D reconstruction. Subsequently, post-processing techniques are employed to eliminate base and outliers, as well as

the final volume estimation of the stockpile. A detailed breakdown of each step is provided in the sections that follow.

2.1 Data Acquisition

Data acquisition in this study involves the formation of CDW stockpiles within a laboratory setup. The laboratory was illuminated using a three-point lighting system, providing sufficient brightness to ensure optimal experimental conditions. The primary smartphone camera for capturing visual data was an iPhone 15 Pro with a 48 MP rear camera. Video sequences were recorded in 4K (2160p) resolution at 60 frames per second. They typically lasted around 30 seconds, covering a complete orbit around the stockpile while maintaining walking speed, as shown in Figure 3a. The stockpile was positioned at the center of the setup, and a 360-degree orbital video sequence was captured to ensure comprehensive coverage, keeping the stockpile consistently framed at the center throughout. To maintain consistency, lighting conditions were standardized at 1000 lux for all experiments except those explicitly testing shadow and lighting parameters.

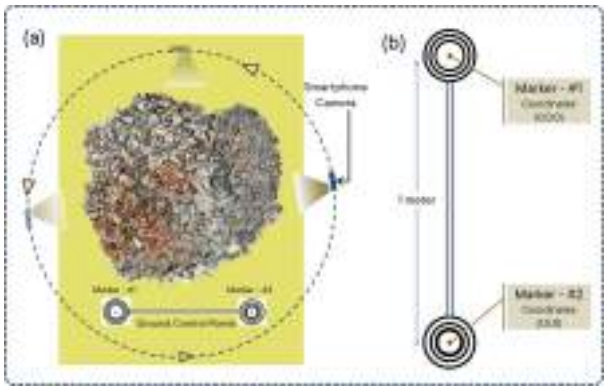


Figure 3. (a) Laboratory setup for data acquisition (b) C^{2}Tag calibration stick developed for this study.

In addition to primary data collection through video sequences captured with the iPhone camera, secondary data were obtained by scanning the stockpiles using the integrated LiDAR sensor of the iPhone 15 Pro. According to the literature, the iPhone's LiDAR sensor has been proven to rapidly produce precise and detailed 3D models, with its applications increasingly documented across various scientific research fields [7]. This secondary dataset was used as a validation reference, ensuring the robustness and accuracy of the primary data derived from video sequences. Monocular SfM fundamentally faces challenges with scale uncertainty, as it reconstructs scenes up to an arbitrary scale factor due to the absence of direct depth information from a single camera view [8]. This limitation necessitates external references, such as Ground Control Points (GCPs), to ensure geometric accuracy and enhance the georeferencing of 3D models. To address the challenge of scale accuracy during 3D reconstruction in this study, C^{2}Tag markers were employed as GCP, facilitating precise calibration and reconstruction [12]. A calibration stick was customized and utilized for accurate scale reference during scanning. The stick, measuring 1 meter in length, featured C^{2}Tag markers affixed to each end, as illustrated in Figure 3b. All the video sequences were obtained by placing the calibration stick.

2.2 Pre-Processing

The pre-processing step employed Python's OpenCV library to extract images from video sequences [13]. A custom Python script was developed using the CaptureVideo function of OpenCV to perform this operation. The script was designed to be alterable, allowing for modifying parameters such as frame extraction rate and resolution to tailor the output based on specific experiment requirements.

2.3 Processing

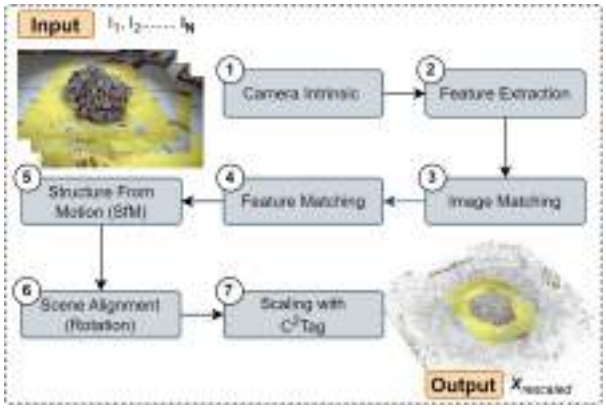


Figure 4. 3D reconstruction pipeline in Meshroom, including scaling with scene alignment and C^{2}Tag Node.

For 3D reconstruction from the extracted images, this study utilized AliceVision Meshroom 2023.3.0, an open-source 3D reconstruction pipeline known for its robust processing [9]. The reconstruction was performed in an HP Workstation Z2, with an Intel i9 processor, 64 GB RAM, and an Nvidia A4000 GPU. The detailed pipeline adopted for transforming images into a point cloud for this study is depicted in Figure 4. Each node in the pipeline is numerically labeled from 1 to 7. The pipeline initiates with N input images. Initially, in Node 1, the camera intrinsic matrix (K) is generated, which

is used to interpret image geometry accurately. In Node 2, features are extracted from each image using multiple feature detection describers like DSP-SIFT, AKAZE, and C^{2}Tag. Node 3 focuses on pairwise image matching to establish relationships between images. Node 4, feature matching uses Euclidean distance to align features across the matched images accurately. A detailed explanation of Node 5-7 is mentioned below.

Node 5, which implements Structure from Motion [9], first estimates the camera's orientation and position (R_t) in 3D space, commonly called the extrinsic matrix, by leveraging epipolar geometry. Matched features from Node 4 are fused into tracks, where each track corresponds to a point in space visible from multiple camera perspectives through triangulation and bundle adjustment. This process allows for the projection and reconstruction of the 3D scene. The relationship between 3D points (X) and their corresponding 2D pixel (x) is mathematically expressed in Equation (1). Z represents the depth of the point in the camera coordinate, which is determined using triangulation.

$$X = Z \cdot (K \cdot R_t)^{-1} \cdot x \tag{1}$$

Upon examining several point clouds generated through the pipeline mentioned above, using images captured by the iPhone 15 Pro camera, it was observed that its orientation was misaligned (see Figure 5a). A manual transformation is applied to resolve the misalignment by introducing Node 6, named Scene Alignment. Specifically, the point cloud is rotated 90 degrees clockwise around the x-axis by multiplying the entire point cloud X with the rotation matrix R_x:

$$R_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}$$

Thus, the aligned point cloud X_{aligned} is (see Equation (2)).

$$X_{\text{aligned}} = R_x \cdot X \tag{2}$$

As discussed earlier, SfM suffers from inherent scale ambiguity. To address this, a C^{2}Tag calibration stick was placed during the data acquisition process, and subsequently, Node 7, named Scaling with C^{2}Tag, was introduced in the pipeline to scale the aligned point cloud (see Figure 5b). To achieve this, a scaling factor S is computed as (Equation 3):

$$S = \frac{d_{\text{real}}}{\left| X_{\text{aligned}(1)} - X_{\text{aligned}(2)} \right|} \tag{3}$$

Where, $X_{\text{aligned}(1)}$ and $X_{\text{aligned}(2)}$ are points in the point cloud corresponding to the same markers 1 and 2 in the real world, respectively. d_{real} is the calibration stick length. The scaling factor S is then applied to the aligned point cloud X_{aligned}, resulting in the rescaled point cloud (see Equation (4)).

$$X_{\text{rescaled}} = S \cdot X_{\text{aligned}} \tag{4}$$

This entire Meshroom pipeline ensures the 3D reconstructed scene is correctly oriented and accurately scaled, facilitating further analysis and visualization.

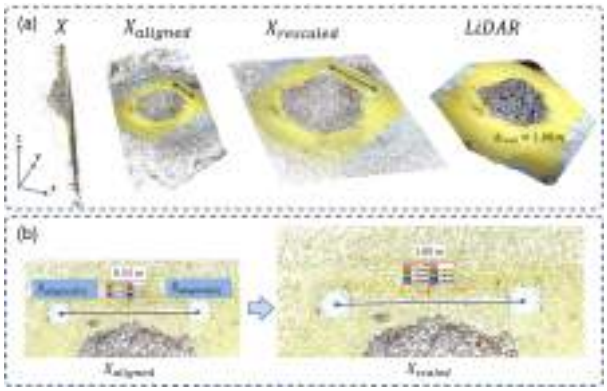


Figure 5. (a) Transformation of the point cloud from Node 5 to Node 7, (b) visual illustration of scaling with C^{2}Tag.

2.4 Post-Processing

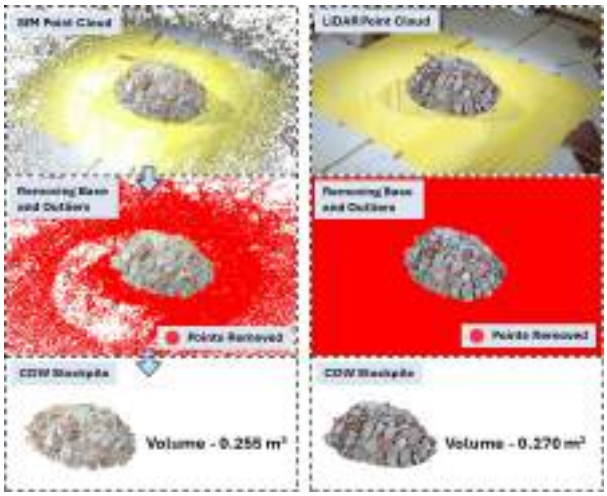


Figure 6. Visual representation of the post-processing step to estimate the SfM and LiDAR point cloud volume.

To ensure precise volume calculation of the CDW stockpile, a yellow sheet was placed beneath to establish a reference base plane (see Figure 3a). Following the

point cloud generation, the subsequent step involved refining the point cloud by eliminating the base plane and extraneous outliers (see Figure 6). Specifically, all points corresponding to the yellow sheet were identified and assigned as the base plane, ensuring that the calculated volume exclusively represents the stockpile material above this plane. The volume estimation was then conducted using the Delaunay Triangulation.

3 Experimental Design

A series of systematic experiments were conducted to evaluate the impact of environmental and capture parameters on 3D reconstruction and volume estimation. These experiments were carefully designed to identify optimal conditions for achieving efficient and reliable outcomes. Among capture parameters, camera height and distance were selected due to their direct impact on the field of view [14]. Shadows and lighting were examined, as they significantly affect feature detection [8]. Frame rate extraction and resolution were also investigated to balance temporal resolution and computational efficiency, ensuring capturing sufficient details without introducing unnecessary computational overhead [10]. The investigations were conducted in the following order: (1) camera height and distance, (2) shadows and lighting, (3) frame rate extraction, and (4) frame resolution. To ensure comprehensive analysis, the experiments were conducted on three separate CDW stockpiles in a controlled environment, with results reported based on the average values across these trials.

3.1 Camera Height and Distance

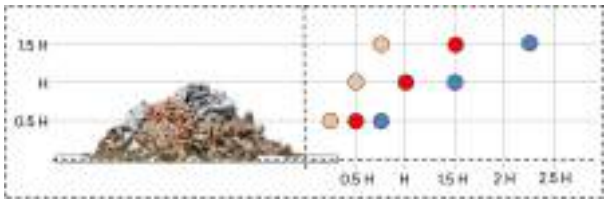


Figure 7. Visual representation of camera height and distance experiment matrix.

To assess the influence of camera height and distance of capture on the 3D reconstruction of a stockpile, an experiment was conducted with the height of the stockpile (H) as a reference parameter. Three camera heights were selected relative to the stockpile: 0.5H, H, and 1.5H to account for varying localized CDW stockpile sizes and ensure diverse camera positions. For instance, 0.5H represents scenarios where the stockpile height is twice the camera height. Similarly, the distance from the stockpile's edge was determined using three distance-to-height (D/H) ratios: 0.5, 1, and 1.5, ensuring that the camera distance scaled proportionally to the stockpile height. This setup resulted in 9 unique camera positions, as shown in Figure 7.

3.2 Shadow and Lighting

This experiment evaluated the effect of shadow and lighting with a design intended to replicate field environmental lighting scenarios where shadows are often unavoidable. This experiment aimed to determine whether variation in lighting conditions during repeated scans of the same stockpile affects reconstruction accuracy and volume accuracy. Figure 8 illustrates the three lighting conditions tested to evaluate their impact on 3D reconstruction:

Case 1: 10 lux lighting (complete darkness).

Case 2: 1000 lux with a single-point light source, which created a shadow on the opposite side of the stockpile.

Case 3: 1000 lux with a three-point light setup, ensuring even illumination and minimizing shadows.



Figure 8. (Case 1) Experimental setup for 0 lux lighting, (Case 2) 1000 lux with a single-point light source, and (Case 3) 1000 lux with a three-point light source.

3.3 Frame Extraction Rate

In this experiment, the number of frames extracted (e.g., FE03 for three frames extracted per second) was varied to identify the optimal requirement for reconstruction and volume estimation, independent of the video length. For this experiment, the custom Python script developed for the pre-processing step was used to perform the frame extraction from video sequences. While increasing the number of extracted frames typically enhances reconstruction precision, it also prolongs the processing time, potentially hindering the requirement to perform quick volume estimation of localized CDW. To balance these factors, nine different frame extraction rates were tested, including FE30, FE10, FE06, FE04, FE03, FE2, FE01, and FE0.5, to determine the optimal frame extraction value.

3.4 Frame Resolution

The literature indicates that frame resolution is typically expected to be used at the raw quality captured. This experiment investigated the effect of reduction in frame resolution on reconstruction to determine the optimal value. The resolution of frames extracted from the video sequence was progressively reduced from the original raw image quality to 2160p, 1080p, 720p, 480p, and 360p using the custom Python script described in the pre-processing step.

3.5 Evaluation Criteria

This study employs two primary evaluation criteria to identify the optimal values. The first criterion focuses on the accuracy of the volume estimation when compared to secondary data, calculated using Equation (5):

$$\text{Volume Accuracy (\%)} = \left(1 - \frac{|V_S - V_L|}{V_L}\right) \times 100 \quad (5)$$

where V_S and V_L are the SfM-based and LiDAR-based volume estimates obtained after the post-processing step. The second criterion assesses the points generated per unit time, representing the total number of points generated during the entire 3D reconstruction process divided by the reconstruction time. A higher value indicates an optimized reconstruction with a shorter processing time, reflecting the optimal value, provided that the volume accuracy remains high for the same condition.

4 Results and Discussion

4.1 Results of Camera Height and Distance

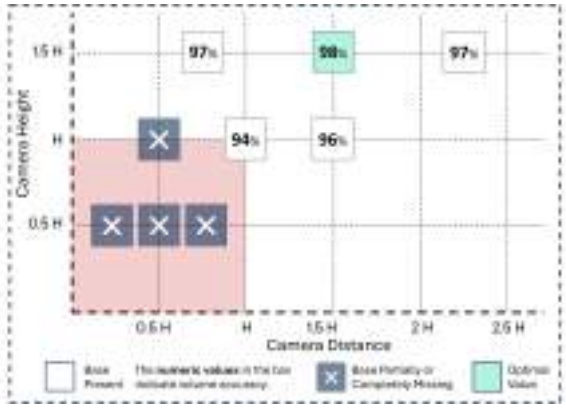


Figure 9. Result matrix of camera height and distance experiment.

The findings from this experiment reveal that the camera positions below the height and distance threshold of H resulted in incomplete reconstruction, with the reference base partially or entirely missing, hampering the post-processing step. (see Figure 9). In contrast, camera positions above the height and distance threshold of H enabled complete reconstruction, allowing for the effective removal of the base and outliers, which ultimately led to successful volume estimation. Among the camera positions, a height and distance of 1.5H proved optimal, achieving the highest volume accuracy of 98%. The points generated per unit of time remained consistent at approximately 330 for camera positions above the H threshold. However, the points generated per unit of time were inconsistent for positions below the threshold, where the reconstruction was incomplete. This suggests that for localized CDW volume estimation, a threshold of camera height and distance above H can be effectively utilized.

4.2 Results of Shadow and Lighting

This experiment was conducted using the optimal camera height and distance, fixed at 1.5H. As shown in Figure 10a, the volume accuracy remained consistently close to 97% across all tested conditions. However, a noteworthy observation is that the point generation rate per unit time in Case 1 was significantly lower than in the other two cases, with the total number of generated points being approximately half that of Case 3. This variation is attributed to reduced feature matching in Case 1, highlighting the key role of lighting in reconstruction. Despite these variations, no significant changes in volume accuracy were observed, indicating that lighting has minimal impact on localized CDW volume estimation.

4.3 Results of Frame Extraction Rate

Figure 10b illustrates that accuracy remained consistent despite a decreased frame extraction rate from FE30 to FE01. However, a marked drop in volume accuracy was observed for FE0.5. Simultaneously, the number of points generated per unit of time steadily increased until reaching FE01, beyond which it sharply decreased. These results suggest that FE01 represents the optimal parameter for maximizing both reconstruction efficiency and volume accuracy, as it is associated with the lowest reconstruction time of 93 seconds and a 98% volume accuracy.

4.4 Results of Frame Resolution

This experiment was carried out using the optimal frame extraction rate of FE01 while varying the frame resolution. The results shown in Figure 10c indicate that reducing the frame resolution below 720p caused a significant decline in volume accuracy. However, this reduction in resolution also decreased reconstruction time from 93 seconds to 56 seconds. Therefore, 720p was identified as the optimal

resolution, balancing volume accuracy and reconstruction efficiency. In contrast, the iPhone LiDAR exhibited much slower performance, requiring an average of 350 seconds per scan.

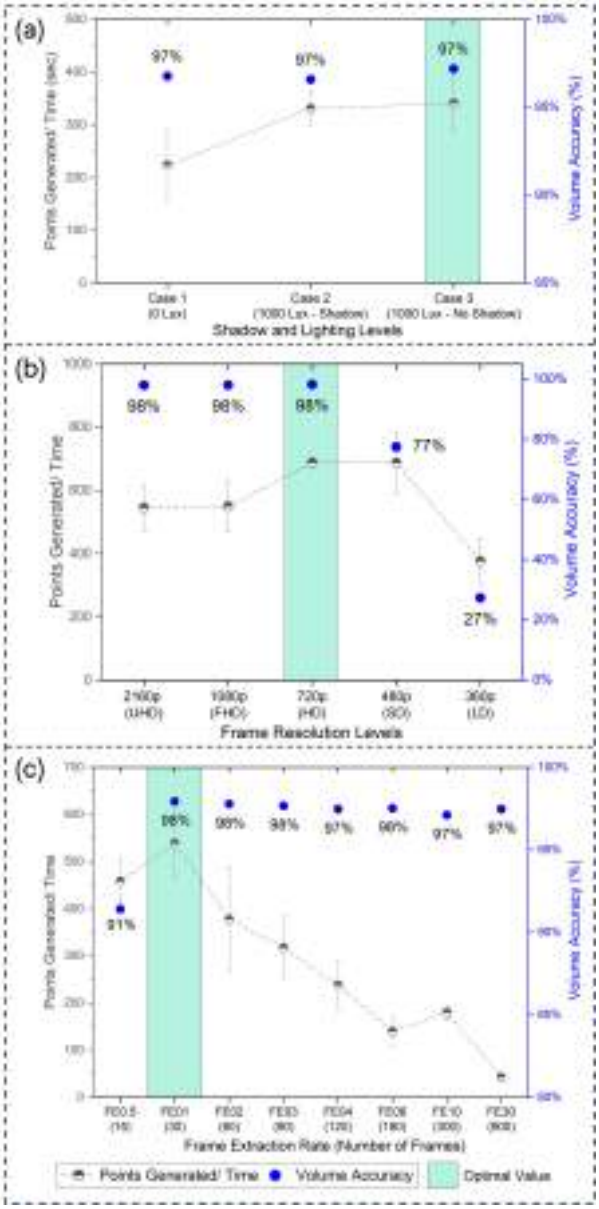


Figure 10. (a) Plot of shadow and lighting experiment, (b) frame extraction rate experiment, and (c) frame resolution experiment.

4.5 Accuracy Assessment

The optimal results derived from the experiments were subsequently employed to evaluate the SfM-based CDW volume estimation workflow. Using the optimized conditions, video sequences of 30 distinct CDW stockpiles were captured to evaluate the performance of the workflow. Figure 11 illustrates a scatter plot comparing volume estimates derived from the SfM-based method (V_S) to those obtained using LiDAR (V_L), yielding a high correlation with an R^2 value of 0.929. This high degree of correlation highlights the proposed SfM-based workflow's suitability for localized CDW volume estimation. Furthermore, the SfM-based workflow shows exceptional efficiency, achieving an average end-to-end time of approximately 97 seconds from data acquisition to post-processing. This time represents a performance that is 3.8 times faster than the 373 seconds required by the LiDAR-based volume estimation workflow, highlighting its operational efficiency. However, the SfM-based end-to-end time includes the processing time, which depends on the system specification used for reconstruction and is subject to change. It also highlights the practicality of the SfM-based workflow for on-site applications.

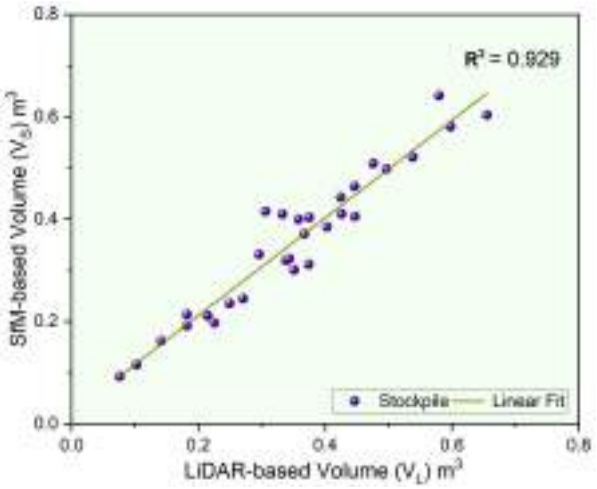


Figure 11. Scatter plot of SfM-based CDW stockpile volume comparison with the LiDAR volume.

4.6 Limitations and future research directions

The present study highlights the potential of SfM for quantifying localized CDW under controlled environments, specifically evaluating the influence of environmental and capture parameters. However, practical application in real-world field environments is limited by several factors, such as site layout, terrain complexity, material composition, spatial variability, and climate conditions. Future research should systematically address these factors to enhance the proposed workflow effectiveness and applicability in the real world. Another limitation of this study is its reliance on the iPhone as the representative smartphone camera, excluding alternative smartphone models. Preliminary investigation indicated that the SfM point clouds remains largely consistent across

different devices under identical capturing conditions, with the only significant change observed in terms of scene alignment. Consequently, Node 6 of reconstruction pipeline require adjustment when using another smartphone camera.

5 Conclusion

Quantifying localized CDW stockpiles is critical in enhancing waste collection and recycling practices. With the rapid advancements in SfM-based 3D reconstruction technologies, it is essential to understand how environmental and capture parameters influence small-scale reconstructions, particularly in the context of localized CDW quantification. This study systematically investigates and identifies, through rigorous experimentation, the optimal environmental and capture conditions required for efficient reconstruction and accurate quantification of CDW stockpiles. Key experimental findings suggest that environmental parameters, such as shadows and lighting, have an insignificant impact on accuracy but moderately affect feature detection during reconstruction, which aligns with the previous studies. However, no significant changes in volume accuracy were observed. In contrast, capture-related parameters, including camera position, frame extraction rate, and frame resolution, emerge as primary determinants of SfM-based 3D reconstruction performance. Notably, optimal point cloud generation and volume accuracy were achieved when the camera height and distance were 1.5 times the stockpile height. A frame extraction rate of FE01 and a resolution of 720p were identified as optimal settings for efficient reconstruction and accurate volume estimation, suggesting that high-fidelity images can be avoided to reduce reconstruction time. Using the optimal values derived from experiments, this study analyzed 30 CDW stockpile samples, demonstrating a high correlation (R^2 = 0.929) between the volume estimates derived from SfM and those obtained from LiDAR. These results showcase the optimization of key parameters, enabling efficient 3D reconstruction and accurate volume estimation of localized CDW stockpiles, making it particularly suitable for small-scale applications.

References

[1] K. Bhattacharjee, S. Chaudhary, A. Vishnoi, D. A. Patel, and N. Bugalia. Characterization of health and safety hazards of deconstruction activities. *American Journal of Industrial Medicine*. doi:10.1002/ajim.23652.

[2] Z. Wu, T. W. Ann, L. Shen, and G. Liu. Quantifying construction and demolition waste: An analytical review. *Waste management*, 34(9):1683–1692, 2014. doi:10.1016/j.wasman.2014.05.010.

[3] R. Brandão, D. J. Edwards, M. R. Hosseini, A. C. Silva Melo, and A. N. Macêdo. Reverse supply chain conceptual model for construction and demolition waste. *Waste Management & Research*, 39(11):1341–1355, 2021. doi:10.1177/0734242X21998730.

[4] A. Jaiswal, P. Roy, N. Bugalia, K. Varghese, and Q. P. Ha. Framework to identify directions for future construction and demolition waste management technologies. In *E3S Web of Conferences*. EDP Sciences, 2024. doi:10.1051/e3sconf/202449604005.

[5] R. Wróżyński, K. Pyszny, M. Sojka, C. Przyby la, and S. Murat-B lażejewska. Ground volume assessment using'structure from motion'photogrammetry with a smartphone and a compact camera. *Open geosciences*, 9(1):281–294, 2017. doi:10.1515/geo-2017-0023.

[6] A. Alsayed and M. R. A. Nabawy. Stockpile volume estimation in open and confined environments: A review. *Drones*, 7(8):537, 2023. doi:10.3390/drones7080537.

[7] G. Luetzenburg, A. Kroon, K. K. Kjeldsen, K. D. Splinter, and A. A. Bjørk. High-resolution topographic surveying and change detection with the iphone lidar. *Nature Protocols*, pages 1–22, 2024. doi:10.1038/s41596-024-01024-9.

[8] O. Özyeşil, V. Voroninski, R. Basri, and A. Singer. A survey of structure from motion*. *Acta Numerica*, 26: 305–364, 2017. doi:10.1017/S096249291700006X.

[9] C. Griwodz, S. Gasparini, L. Calvet, P. Gurdjos, F. Castan, B. Maujean, G. De Lillo, and Y. Lanthony. Alicevision meshroom: An open-source 3d reconstruction pipeline. In *Proceedings of the 12th ACM multimedia systems conference*, pages 241–247, 2021. doi:10.1145/3458305.3478443.

[10] R. García-Luna, S. Senent, R. Jurado-Piña, and R. Jimenez. Structure from motion photogrammetry to characterize underground rock masses: Experiences from two real tunnels. *Tunnelling and underground space technology*, 83:262–273, 2019. doi:10.1016/j.tust.2018.09.026.

[11] I. Nikolov and C. Madsen. Benchmarking close-range structure from motion 3d reconstruction software under varying capturing conditions. In *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection Proceedings, Part I 6*, pages 15–26. Springer, 2016. doi:10.1007/978-3-319-48496-9_2.

[12] L. Calvet, P. Gurdjos, and V. Charvillat. Camera tracking using concentric circle markers: Paradigms and algorithms. In *2012 19th IEEE International Conference on Image Processing*, pages 1361–1364. IEEE, 2012. doi:10.1109/ICIP.2012.6467121.

[13] G Bradski. The opencv library. *Dr. Dobb's Journal of Software Tools*, 2000.

[14] A. Bisson-Larrivée and J.-B. LeMoine. Photogrammetry and the impact of camera placement and angular intervals between images on model reconstruction. *Digital Applications in Archaeology and Cultural Heritage*, 26: e00224, 2022. doi:10.1016/j.daach.2022.e00224.

Transformer-based Multi-resolution Fast 3D Reconstruction for Structural Damage Detection

Hui Zuo[1], Tao Sun[2], Hao Xie[1], Xiao Ma[3], Nima Shirzad-Ghaleroudkhani[1] and Qipei Mei[1]

[1]Department of Civil and Environmental Engineering, University of Alberta, Edmonton, AB T6G 1H9, Canada
[2]Department of Civil Engineering, McGill University, Montréal, QC H3A 0C3 , Canada
[3]Department of Biomedical Engineering, University of Alberta, Edmonton, AB T6G 2T4, Canada
polo.huizuo@ualberta.ca, tao.sun@mail.mcgill.ca, hxie9@ualberta.ca, xm11@ualberta.ca, shirzadg@ualberta.ca, qipei.mei@ualberta.ca

Abstract –

This study proposes a multi-resolution fast 3D reconstruction framework that integrates transformer-based damage detection with rapid 3D modeling to enhance bridge surface defect identification and spatial localization. The framework consists of three phases: (I) 3D reconstruction using Structure from Motion (SfM) to generate a structural model with sparse point cloud, (II) damage segmentation via a customized Swin UNETR model for precise defect detection, and (III) multi-resolution dense reconstruction that prioritizes high-resolution modeling of detected defects while reducing the resolution of non-critical areas to improve efficiency. Experimental validation on the High Level Bridge in Edmonton, Canada, demonstrated the framework's capability to accurately map surface defects onto a 3D model, providing an intuitive and detailed localization for structural assessment. This approach offers significant potential for efficient and accurate bridge inspection, supporting data-driven maintenance strategies.

Keywords –

Transformer; Point Cloud; 3D Reconstruction; Structure from Motion; Structural Damage Detection

1 Introduction

Bridges are critical components of civil infrastructure but are subjected to various adverse loads throughout their service life, accelerating structural deterioration [1]. The accumulation of these loads induces repeated stress and strain, leading to the formation of surface cracks that gradually compromise structural integrity over time [2]. Thus, regular inspections and timely repair based on damage assessment are essential to ensure bridge safety [3]. Recent advances in computer vision have greatly improved defect detection methods, enabling efficient analysis of large-scale image datasets for rapid damage assessment [4-5]. However, the widespread and varied nature of bridge damage presents challenges for existing techniques in simultaneously detecting and localizing surface defects [6]. Moreover, mapping surface damage onto a 3D model has proven beneficial for effective bridge maintenance, offering critical insights that support accurate decision-making by bridge owners [7]. Therefore, achieving precise detection and localization of bridge surface damage within a 3D model is of significant importance [8].

Currently, vision-based damage detection methods can be mainly divided into image classification, object detection, and semantic segmentation [9]. With advancements in deep learning, semantic segmentation has become a leading approach [10]. Various convolutional neural networks (CNNs) have been applied to damage segmentation, including U-Net [11], ResNet [12], DenseNet [13], Mask R-CNN [14], DeepLabV3+ [15], and hybrid models [16]. However, due to the limited receptive fields of CNN-based models, capturing global features in complex real-world scenarios remains challenging [17]. One promising solution is the use of transformer. Since Dosovitskiy et al. [18] introduced the encoder-based Vision Transformer (ViT), transformer has gained attraction in computer vision. Its embedding and self-attention mechanisms have shown superior accuracy compared to complex CNN models for image segmentation tasks [19]. Swin UNETR [20], which integrates the Swin Transformer [21] with an enhanced ViT, has demonstrated outstanding accuracy and efficiency across benchmark datasets [22]. Moreover, combining Unmanned Aerial Vehicle (UAV) with advanced algorithms offers significant potential for rapid inspection of high-risk areas, such as bridge piers and abutments [23].

To obtain the 3D spatial information of bridge damage, point cloud data is commonly used to generate a 3D bridge model. With the development in localization technology, stereo vision-based local positioning methods have been applied, addressing the reliance of

UAVs on the Global Positioning System (GPS) [24]. And Structure from Motion (SfM) is widely regarded as one of the common techniques for 3D reconstruction from a series of sparse images. This method enables the creation of a complete 3D model, integrating local detection results and reflecting the spatial position of objects [25]. Although semantic segmentation can classify damage at the pixel level, it faces challenges in consistently identifying and matching the same damage across sequential image frames, thus hindering the integration of damage segmentation with spatial localization [6]. Additionally, generating 3D bridge models with SfM presents several challenges. First, precise image capture and detailed trajectory planning are often required to avoid motion blur during data collection to ensure successful reconstruction. This time-consuming process significantly reduces efficiency [26]. Second, processing large volumes of images with SfM greatly extends computation time, further limiting reconstruction efficiency [27].

In response, this study integrates the high accuracy of image-based damage detection with the spatial localization capabilities of 3D reconstruction. A new method is proposed that combines transformer-based damage detection with rapid 3D reconstruction, enabling seamless and efficient mapping of surface damage on bridges onto the 3D model. This approach offers decision-makers an intuitive and comprehensive localization of structural assessment.

2 Methodology

The proposed multi-resolution 3D reconstruction framework comprises three phases: (I) 3D reconstruction, (II) transformer-based damage segmentation, and (III) multi-resolution dense reconstruction. Phase I generates a 3D structural model with data captured by UAV using SfM. Phase II employs a transformer model to detect and segment surface defects from feature images. Phase III performs multi-resolution dense reconstruction, producing a dense point cloud where defects are rendered in high resolution while the overall structural resolution is reduced, enabling fast and precise damage localization. The overall scheme is shown in Figure 1.

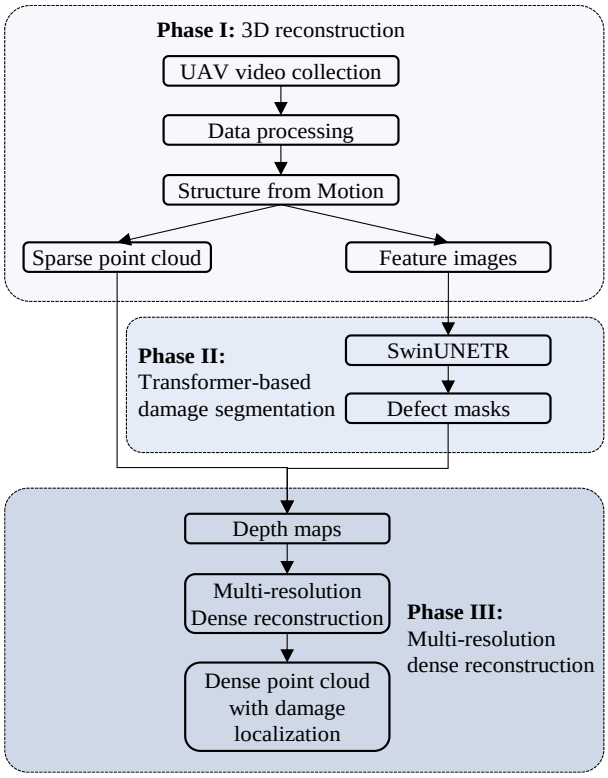


Figure 1. Proposed Framework

2.1 3D reconstruction

In this phase, the incremental SfM [28] is used to process RGB images collected from UAV inspections, with each image associated with a local coordinate system (LCS) defined by the camera's field of view and orientation. The goal is to convert these LCS-based observations into a global coordinate system (GCS) to enable accurate 3D reconstruction. This process requires the extraction of intrinsic and extrinsic camera parameters and involves feature matching and incremental reconstruction.

To accurately represent the reconstructed model in a global coordinate system, it is essential to convert the image data from the LCS to the GCS [29]. The intrinsic camera matrix $\mathbf{K}$ (see Equation (1)) and the extrinsic rotation and translation matrix $\mathbf{R}|\mathbf{t}$ (see Equation (2)) are necessary for transforming coordinates from the LCS to the GCS. The intrinsic matrix $\mathbf{K}$ captures the camera's internal parameters, including the focal length (f_x, f_y), camera lens distortion s and the principal point coordinates (c_x, c_y).

$$\mathbf{K} = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

While the extrinsic matrix $\mathbf{R}\,|\,\mathbf{t}$ defines the camera's rotation and translation in the global frame, including the camera position transformation matrix $r_{i,j}$, and the translation vector $\mathbf{t} = \left[t_x, t_y, t_z\right]^T$.

$$\mathbf{R}\,|\,\mathbf{t} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2}$$

Finally, the relationship in Equation (3) can be used to transform a 3D point in the LCS (x_L, y_L, z_L) into the GCS (X_G, Y_G, Z_G) through the extrinsic parameters:

$$\begin{bmatrix} x_L \\ y_L \\ z_L \\ 1 \end{bmatrix} = \left[\mathbf{R}\,|\,\mathbf{t}\right] \begin{bmatrix} X_G \\ Y_G \\ Z_G \\ 1 \end{bmatrix} \tag{3}$$

where the rotation matrix **R** is further decomposed into rotations around the X, Y, and Z axes as shown below.

$$\mathbf{R} = \begin{bmatrix} \cos\theta_z & -\sin\theta_z & 0 \\ \sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} \cos\theta_y & 0 & \sin\theta_y \\ 0 & 1 & 0 \\ -\sin\theta_y & 0 & \cos\theta_y \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & -\sin\theta_x \\ 0 & -\sin\theta_x & \cos\theta_x \end{bmatrix} \tag{4}$$

2.2 Transformer-based damage segmentation

In this phase, the Swin UNETR, designed for 3D tumor segmentation, is customized for 2D structural damage segmentation. The model processes 2D structural images with two channels by dividing them into non-overlapping patches using a patch partition layer. These patches are transformed into windows suitable for self-attention operations. The encoded features from the Swin transformer are passed through skip connections at various resolutions to a CNN-based decoder. The final output is a segmentation map with two channels, identifying damaged areas in the structure. The architecture of the customized Swin UNETR is shown in Figure 2.

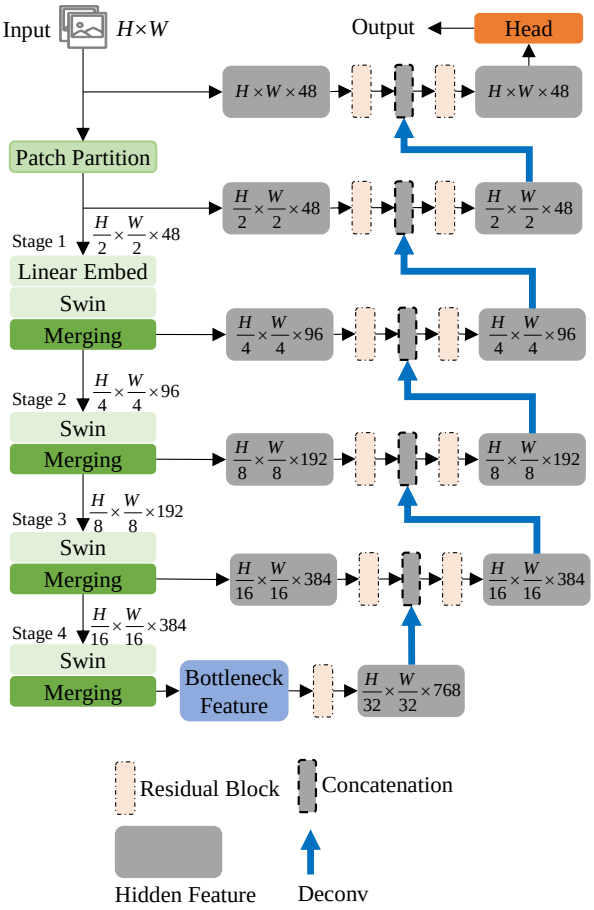


Figure 2. Architecture of the customized Swin UNETR

For the encoding part, the model input $\chi \in \mathbb{R}^{H\times W\times S}$ represents a token with a patch resolution of (H', W') and a dimension of $H'\times W'\times S$. Initially, the model uses a patch partitioning layer to segment the input image into fixed-size, non-overlapping patches with dimensions $\frac{H}{H'}\times\frac{W}{W'}$. These patches are then embedded into a feature space of dimension *C*. Self-attention is applied to these patches within non-overlapping windows to efficiently model relationships between tokens. Specifically, windows of size $M\times M$ partition the token into $\frac{H'}{M}\times\frac{W'}{M}$ regions at a given layer l in the transformer encoder. In the following layer, window positions are shifted by $\left(\frac{M}{2},\frac{M}{2}\right)$ pixels to capture cross-region interactions. The encoder computes layer outputs using the following equations:

$$\begin{aligned} \tilde{z}^{l} &= W\text{-}MSA\left(LN\left(z^{l-1}\right)\right) + z^{l-1} \\ z^{l} &= MLP\left(LN\left(\tilde{z}^{l}\right)\right) + \tilde{z}^{l} \\ \tilde{z}^{l+1} &= SW\text{-}MSA\left(LN\left(z^{l}\right)\right) + z^{l} \\ z^{l+1} &= MLP\left(LN\left(\tilde{z}^{l+1}\right)\right) + \tilde{z}^{l+1} \end{aligned} \tag{5}$$

where W-MSA and SW-MSA represent standard and shifted window-based multi-head self-attention modules. LN and MLP denote layer normalization and multi-layer perceptrons. $\tilde{z}^{l}$ and $\tilde{z}^{l+1}$ indicate the output of W-MSA and SW-MSA, respectively.

For the 2D task, the encoder begins with a patch size of 2×2 and a feature dimension of 2×2×4=16. The embedding dimension C is set to 48. The encoder consists of four stages, each with two transformer blocks, totalling eight layers. A linear embedding layer reduces the spatial resolution by half at each stage, resulting in feature maps of size $\frac{H}{2}\times\frac{W}{2}$, $\frac{H}{4}\times\frac{W}{4}$, $\frac{H}{8}\times\frac{W}{8}$, $\frac{H}{16}\times\frac{W}{16}$.

The decoder adopts a U-shaped architecture where feature representations from the encoder are integrated through skip connections at corresponding resolutions. In each stage i, $i \in \{0,1,2,3,4\}$, the feature maps are reshaped to $\frac{H}{2^i}\times\frac{W}{2^i}$ and processed through residual blocks composed of two 3×3 convolutional layers, followed by instance normalization. To progressively restore spatial resolution, deconvolutional layers double the size of feature maps, and these upsampled features are concatenated with the outputs from the encoder. This process continues until the original resolution is recovered. Finally, a 1×1 convolutional layer and a sigmoid activation function are used to generate the segmentation mask, highlighting regions of structural damage.

The Focal Loss [30] was employed as the loss function, which can effectively address the class imbalance commonly present in damage detection tasks, where damaged regions occupy a much smaller area compared to the background. As displayed in Equation (6), it modifies the standard cross-entropy loss by adding a modulating factor that down-weights easy examples.

$$FL\left(p_t\right) = -\alpha_t\left(1-p_t\right)^{\gamma}\log\left(p_t\right) \tag{6}$$

where p_t is the predicted probability for the true class, and

$$p_t = \begin{cases} p, if\ \ y=1 \\ 1-p, if\ \ y=0 \end{cases} \tag{7}$$

where p is the predicted probability and y is the ground truth label. α_t is a balancing factor for class imbalance, γ is the focusing parameter that adjusts the rate at which easy examples are down-weighted.

For performance evaluation, several metrics are used to assess the segmentation quality, including Intersection over Union (IoU), Dice Coefficient, and Pixel-wise Accuracy:

$$IoU = \frac{TP}{TP+FP+FN} \tag{8}$$

$$Dice = \frac{2TP}{2TP+FP+FN} \tag{9}$$

$$Accuracy = \frac{TP}{TP+FN} \tag{10}$$

where *TP*, *FP* and *FN* stand for True Positive, False Positive, False Negative, respectively.

By leveraging the Swin transformer's ability to model multi-scale contextual information and long-range dependencies, the network can effectively capture complex patterns of structural defects.

2.3 Multi-resolution Dense Reconstruction

SfM provides essential camera parameters (intrinsic and extrinsic), sparse 3D points, and their 2D image correspondences. However, this sparse representation lacks the detail needed for comprehensive modeling. To overcome this, Multi-View Stereo (MVS) leverages the camera's internal and external parameters to perform stereo-matching and identify points in space that have photometric consistency [31], thus transforming sparse point clouds into dense models. In this phase, Open Multi-View Stereo (OpenMVS) [32], an open-source library, is adapted for dense 3D reconstruction.

OpenMVS primarily employs a depth map fusion-based MVS approach, which consists of several stages. Initially, for each input image, the most relevant neighboring images are selected to form stereo pairs. Depth maps are then estimated for each image by identifying photometrically consistent points across these pairs. Once individual depth maps are computed, they are fused into a unified and dense point cloud that captures fine surface details.

To optimize the dense 3D reconstruction process and prioritize defect areas over the whole structure, a significant modification is to ignore specified regions in the input corresponding images during the dense reconstruction stage. In this workflow, the structural components of the object are reconstructed at a reduced resolution, significantly accelerating the dense point cloud generation for the overall structure. Meanwhile, defect regions identified by Swin UNETR are preserved and processed at full resolution. By this way, the fine details of critical defect areas are captured with high fidelity, while the less critical structural areas are downsampled to minimize computational load. As a result, the overall dense reconstruction process becomes

faster and more efficient compared to standard methods that process the entire structure uniformly.

3 Experiments and Results



Figure 3. Data Collection

To evaluate the proposed multi-resolution 3D reconstruction framework, experiments were conducted on the High Level Bridge in Edmonton, Canada (see Figure 3). Opened in 1913, the bridge accommodates both Light Rail Transit (LRT) and vehicle traffic with concrete piers, making it a suitable structure for testing defect detection and 3D reconstruction methods.

A DJI Mini 3 Pro drone was employed to collect visual data of the first concrete pier following the approach bridge. The UAV was operated to fly around the pier, capturing continuous 4K video footage to ensure comprehensive coverage of the structure's surface. After the data collection, frames were extracted from the recorded video at a rate of one image per second.

The extracted images were processed using SfM to obtain a sparse 3D point cloud and corresponding camera poses. The SfM pipeline effectively reconstructed the pier's general geometry by identifying and matching feature points across images, followed by incremental camera pose estimation and 3D point triangulation. The resulting sparse point cloud and the associated camera positions are visualized in Figure 4.

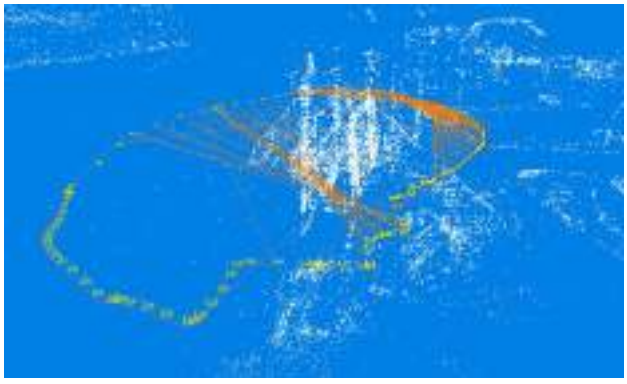


Figure 4. Sparse point cloud with camera pose

To achieve optimal performance in segmenting surface defects on the concrete pier, the Swin UNETR model was trained using domain-specific data. The public dataset Crack5769 [33] was selected for training. This dataset consists of 5,769 pixel-wise labeled concrete crack images, each with a resolution of 256×256 pixels, making it well-suited for the concrete pier context. The dataset was divided into training, validation, and testing sets in a ratio of 8:1:1 to ensure balanced evaluation and prevent overfitting. The model was trained on an Ubuntu 22.04 system equipped with an NVIDIA RTX A6000 GPU. The training was configured with batch size of 16, feature size of 48, learning rate of 0.0001, Adam optimizer and a total of 300 epochs.

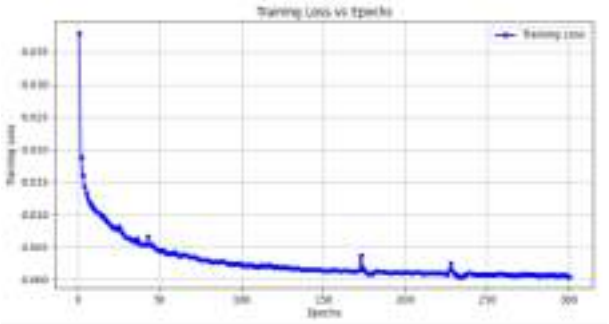


Figure 5. Training loss

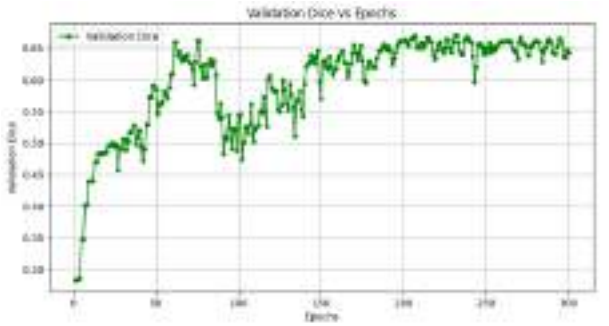


Figure 6. Validation Dice Coefficient

Throughout the training process, the model's performance was monitored by tracking both the training loss and validation Dice coefficient across epochs. Figure 5 illustrates the training loss over 300 epochs, showing a gradual and consistent decrease, which indicates effective model convergence and learning stability. Figure 6 presents the validation Dice coefficient across epochs, demonstrating steady improvement and stabilization in segmentation performance as the training progressed. These plots confirm that the model was learning meaningful features for crack detection without overfitting. Finally, the trained model achieved an average IoU of 0.54, an average Dice coefficient of 0.67, and an average pixel-wise accuracy of 0.69 on the testing set. These results suggest that the Swin UNETR model

effectively learned to detect and segment cracks on concrete surfaces.

After achieving the optimal transformer model, the undistorted images from SfM were then passed through the customized Swin UNETR for damage segmentation. The segmentation results were subsequently integrated into the multi-resolution dense reconstruction process.

Using sparse point cloud generated by SfM and the segmentation masks from Swin UNETR, the OpenMVS pipeline was employed to perform multi-resolution dense reconstruction. The process began with depth map estimation, where depth information was calculated for each image to guide point cloud densification, as illustrated in Figure 7.

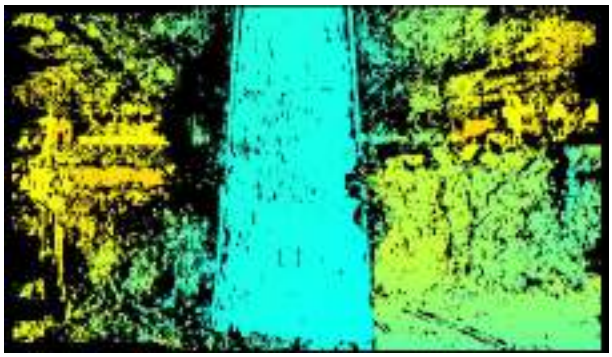


Figure 7. Depth map

The final output consisted of a dense 3D point cloud with varying levels of detail. Structural areas such as flat pier surfaces were reconstructed with 2 times reduced point density, whereas cracks maintained high-density point clouds, as shown in Figure 8.

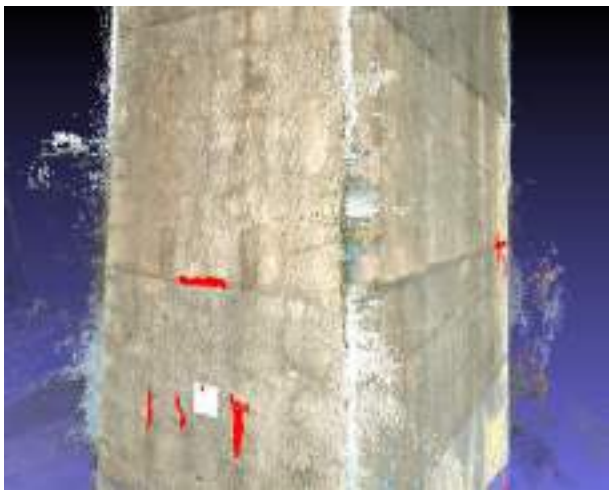


Figure 8. Dense point cloud

To evaluate the efficiency of the proposed method, a runtime comparison was conducted between the uniform-resolution reconstruction and the proposed multi-resolution approach. The results, summarized in Table 1, demonstrate a significant reduction in processing time while maintaining high detail in defect areas. Specifically, the multi-resolution approach achieved substantial reductions in processing time across all reconstruction stages, particularly in depth-maps estimation and point cloud densification. The multi-resolution approach reduced the total reconstruction time by around 64% and decreased memory consumption, while maintaining defect quality. This balance between efficiency and accuracy makes the proposed framework well-suited for large-scale structural health monitoring applications where both accuracy and speed are critical.

Table 1. Runtime Comparison

Action	Normal	Multi-Resolution	Improvement
Estimate Depth-maps	25 m 10 s 221 ms	8 m 3 s 478 ms	↓ 17.1 m
Geometric Consistency	6 m 22 s 885 ms	2 m 28s 319 ms	↓ 3.9 m
Dense fused Depth-maps	1 m 11 s 192 ms	51 s 758 ms	↓ 0.3 m
Densify Point Cloud	39 m 23 s 54 ms	14 m 19 s 721 ms	↓ 25.1 m

Overall, the experimental results validate the effectiveness of the proposed framework in achieving efficient and precise structural damage reconstruction.

4 Conclusion

This study introduces a multi-resolution fast 3D reconstruction framework that integrates UAV-based data collection, transformer-based damage segmentation, and multi-resolution dense reconstruction for efficient and precise structural damage detection. The use of SfM effectively captures the global geometry of the structure, while the customized Swin UNETR model accurately segments surface defects. By prioritizing high-resolution reconstruction in damaged areas and reducing detail in non-critical regions, the proposed framework significantly improves computational efficiency without compromising defect detection quality. Experimental results on the High Level Bridge demonstrated a 64% reduction in processing time and effective defect localization, validating the framework's applicability for structural health monitoring.

However, the current approach has limitations. The model's performance is constrained by the quality and diversity of the training dataset, potentially impacting defect detection under varying lighting and environmental conditions. Additionally, the focus on crack detection limits its generalizability to other defect types, such as spalling, mold or corrosion, etc.

Real-world implementation also presents challenges such as environmental factors affecting UAV inspections, the trade-off between computational efficiency and accuracy, and the need for robust generalization across different defect types and structural conditions. Addressing these limitations, future developments will explore adaptive flight planning, cloud-based computing for real-time processing, and multi-task learning strategies to extend defect detection capabilities. Expanding the framework to identify various defect types, integrating real-time processing and optimizing the system for large-scale infrastructure inspections will further enhance the framework's scalability and practical deployment for long-term structural health monitoring.

Acknowledgement

The authors gratefully acknowledge the support from the NSERC-Alberta Innovates Advance Grants under grant number ALLRP 576826 - 22.

References

[1] L. Sun, Z. Shang, Y. Xia, S. Bhowmick, and S. Nagarajaiah, "Review of Bridge Structural Health Monitoring Aided by Big Data and Artificial Intelligence: From Condition Assessment to Damage Detection," *Journal of Structural Engineering*, vol. 146, no. 5, p. 04020073, May 2020, doi: 10.1061/(ASCE)ST.1943-541X.0002535.

[2] R. Gomasa, V. Talakokula, S. K. R. Jyosyula, and T. Bansal, "Non-destructive Damage Identification of Blended Concrete Systems Using Embedded Piezo Sensors," in *Civil Structural Health Monitoring*, W. Abdullah, M. T. Chaudhary, H. Kamal, J. Parol, and A. Almutairi, Eds., Cham: Springer Nature Switzerland, 2024, pp. 52–61. doi: 10.1007/978-3-031-62253-3_5.

[3] S. Jiang, Y. Zhang, F. Wang, and Y. Xu, "Three-dimensional reconstruction and damage localization of bridge undersides based on close-range photography using UAV," *Measurement Science and Technology*, vol. 36, no. 1, p. 015423, Nov. 2024, doi: 10.1088/1361-6501/ad90fb.

[4] Q. Mei and M. Gül, "Multi-level feature fusion in densely connected deep-learning architecture and depth-first search for crack segmentation on images collected with smartphones," *Structural Health Monitoring*, vol. 19, no. 6, pp. 1726–1744, Nov. 2020, doi: 10.1177/1475921719896813.

[5] Q. Mei, M. Gül, and M. R. Azim, "Densely connected deep neural network considering connectivity of pixels for automatic crack detection," *Automation in Construction*, vol. 110, p. 103018, Feb. 2020, doi: 10.1016/j.autcon.2019.103018.

[6] Y. Ni, J. Mao, H. Wang, Z. Xi, and Z. Chen, "Surface Damage Detection and Localization for Bridge Visual Inspection Based on Deep Learning and 3D Reconstruction," *Structural Control and Health Monitoring*, vol. 2024, no. 1, p. 9988793, Jan. 2024, doi: 10.1155/2024/9988793.

[7] X. Gao and P. Pishdad-Bozorgi, "BIM-enabled facilities operation and maintenance: A review," *Advanced Engineering Informatics*, vol. 39, pp. 227–247, Jan. 2019, doi: 10.1016/j.aei.2019.01.005.

[8] K. Hattori, K. Oki, A. Sugita, T. Sugiyama, and P. Chun, "Deep learning-based corrosion inspection of long-span bridges with BIM integration," *Heliyon*, vol. 10, no. 15, Aug. 2024, doi: 10.1016/j.heliyon.2024.e35308.

[9] X. Pan and T. Y. Yang, "Postdisaster image-based damage detection and repair cost estimation of reinforced concrete buildings using dual convolutional neural networks," *Computer-Aided Civil and Infrastructure Engineering*, vol. 35, no. 5, pp. 495–510, 2020, doi: 10.1111/mice.12549.

[10] C. Xiang, J. Guo, R. Cao, and L. Deng, "A crack-segmentation algorithm fusing transformers and convolutional neural networks for complex detection scenario," *Automation in Construction*, vol. 152, p. 104894, Aug. 2023, doi: 10.1016/j.autcon.2023.104894.

[11] G. Pan, Y. Zheng, S. Guo, and Y. Lv, "Automatic sewer pipe defect semantic segmentation based on improved U-Net," *Automation in Construction*, vol. 119, p. 103383, Nov. 2020, doi: 10.1016/j.autcon.2020.103383.

[12] Z. Fan, H. Lin, C. Li, J. Su, S. Bruno, and G. Loprencipe, "Use of Parallel ResNet for High-Performance Pavement Crack Detection and Measurement," *Sustainability*, vol. 14, no. 3, Art. no. 3, Jan. 2022, doi: 10.3390/su14031825.

[13] İ. Akgül, "Mobile-DenseNet: Detection of building concrete surface cracks using a new fusion technique based on deep learning," *Heliyon*, vol. 9, no. 10, Oct. 2023, doi: 10.1016/j.heliyon.2023.e21097.

[14] L. Deng, H. Zuo, W. Wang, C. Xiang, and H. Chu, "Internal Defect Detection of Structures Based on Infrared Thermography and Deep Learning," *KSCE Journal of Civil Engineering*, vol. 27, no. 3, pp. 1136–1149, Mar. 2023, doi: 10.1007/s12205-023-0391-7.

[15] Z. Pan *et al.*, "High-precision segmentation and quantification of tunnel lining crack using an improved DeepLabV3+," *Underground Space*, Dec. 2024, doi: 10.1016/j.undsp.2024.10.002.

[16] H. Chu, W. Wang, and L. Deng, "Tiny-Crack-Net: A multiscale feature fusion network with attention mechanisms for segmentation of tiny cracks," *Computer-Aided Civil and Infrastructure Engineering*, vol. 37, no. 14, pp. 1914–1931, 2022, doi: 10.1111/mice.12881.

[17] Y. Jin, D. Han, and H. Ko, "TrSeg: Transformer for semantic segmentation," *Pattern Recognition Letters*, vol. 148, pp. 29–35, Aug. 2021, doi: 10.1016/j.patrec.2021.04.024.

[18] A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," Jun. 03, 2021, *arXiv*: arXiv:2010.11929. doi: 10.48550/arXiv.2010.11929.

[19] X. Shen *et al.*, "Self-attentional microvessel segmentation via squeeze-excitation transformer Unet," *Computerized Medical Imaging and Graphics*, vol. 97, p. 102055, Apr. 2022, doi: 10.1016/j.compmedimag.2022.102055.

[20] A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H. R. Roth, and D. Xu, "Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, A. Crimi and S. Bakas, Eds., Cham: Springer International Publishing, 2022, pp. 272–284. doi: 10.1007/978-3-031-08999-2_22.

[21] Z. Liu *et al.*, "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows," Aug. 17, 2021, *arXiv*: arXiv:2103.14030. doi: 10.48550/arXiv.2103.14030.

[22] R. Kakavand, P. Tahghighi, R. Ahmadi, W. B. Edwards, and A. Komeili, "Swin UNETR Segmentation with Automated Geometry Filtering for Biomechanical Modeling of Knee Joint Cartilage," *Annals of Biomedical Engineering*, Jan. 2025, doi: 10.1007/s10439-024-03675-x.

[23] Z. Yu, Y. Shen, and C. Shen, "A real-time detection approach for bridge cracks based on YOLOv4-FPM," *Automation in Construction*, vol. 122, p. 103514, Feb. 2021, doi: 10.1016/j.autcon.2020.103514.

[24] S. Jiang, Y. Cheng, and J. Zhang, "Vision-guided unmanned aerial system for rapid multiple-type damage detection and localization," *Structural Health Monitoring*, vol. 22, no. 1, pp. 319–337, Jan. 2023, doi: 10.1177/14759217221084878.

[25] L. Deng, T. Sun, L. Yang, and R. Cao, "Binocular video-based 3D reconstruction and length quantification of cracks in concrete structures," *Automation in Construction*, vol. 148, p. 104743, Apr. 2023, doi: 10.1016/j.autcon.2023.104743.

[26] C.-Q. Feng, B.-L. Li, Y.-F. Liu, F. Zhang, Y. Yue, and J.-S. Fan, "Crack assessment using multi-sensor fusion simultaneous localization and mapping (SLAM) and image super-resolution for bridge inspection," *Automation in Construction*, vol. 155, p. 105047, Nov. 2023, doi: 10.1016/j.autcon.2023.105047.

[27] J. L. Carrivick, M. W. Smith, and D. J. Quincey, *Structure from Motion in the Geosciences*. John Wiley & Sons, 2016.

[28] P. Moulon, P. Monasse, and R. Marlet, "Adaptive Structure from Motion with a Contrario Model Estimation," in *Computer Vision – ACCV 2012*, K. M. Lee, Y. Matsushita, J. M. Rehg, and Z. Hu, Eds., Berlin, Heidelberg: Springer, 2013, pp. 257–270. doi: 10.1007/978-3-642-37447-0_20.

[29] G. Kim and Y. Cha, "3D Pixelwise damage mapping using a deep attention based modified Nerfacto," *Automation in Construction*, vol. 168, p. 105878, Dec. 2024, doi: 10.1016/j.autcon.2024.105878.

[30] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection," Feb. 07, 2018, *arXiv*: arXiv:1708.02002. doi: 10.48550/arXiv.1708.02002.

[31] H. Ming, Q. Li, H. Xia, and P. Li, "Refined 3D Modeling of Complex Models Based on Stereo Vision," *IEEE Access*, vol. 12, pp. 93487–93501, 2024, doi: 10.1109/ACCESS.2024.3424293.

[32] D. Cernea, "OpenMVS: Multi-View Stereo Reconstruction Library," 2020. [Online]. Available: https://cdcseacave.github.io/openMVS

[33] X. W. Ye, T. Jin, Z. X. Li, S. Y. Ma, Y. Ding, and Y. H. Ou, "Structural Crack Detection from Benchmark Data Sets Using Pruned Fully Convolutional Networks," *Journal of Structural Engineering*, vol. 147, no. 11, p. 04721008, Nov. 2021, doi: 10.1061/(ASCE)ST.1943-541X.0003140.

Crack Detection and Segmentation for Bridges Using State-of-the-Art Deep Learning Methods: Single-Stage vs. Two-Stage Detectors

Ahmed Assad[1], Mohamad Bo Arki[1], Miray Sweid[1] and Amin Hammad[2]

[1]Austrsalian University, Kuwait
[2] Concordia Institute for Information Systems, Concordia University Canada
a.assad@au.edu.kw, amin.hammad@concordia.ca

Abstract

Concrete bridges are critical infrastructure assets that require regular inspection to ensure public safety and efficient functionality. Traditional methods for crack detection are often manual, time-consuming, and prone to subjectivity, which can compromise inspection accuracy and efficiency. Deep learning techniques have been increasingly utilized in bridge crack detection. However, very few efforts have developed instance segmentation models that can be used to detect and quantify cracks in bridges. Different than other computer vision methods, instance segmentation provides pixel-level classification while also differentiating between individual cracks, facilitating crack measurement and proper maintenance decision-making. This study presents a framework for real-time crack detection and segmentation in bridges. The study assesses the performance of advanced deep learning segmentation models including both single-stage detector (YOLOv9 and YOLOv11) and two-stage architectures (Mask-RCNN) coupled with backbone feature extractors (ResNet). A comprehensive dataset of annotated bridge crack images was utilized to train and test the developed models. YOLOv11 emerged as the most efficient model, achieving the highest precision for bounding box detection (96.0%) and segmentation (91.2%) while maintaining the fastest inference time (19 ms per image) and the shortest training duration (127.49 minutes). Mask R-CNN, while excelling in bounding box accuracy (95.59%), demonstrated slower training and inference times, highlighting its limitations in real-time applications. The performance was further evaluated through precision-recall curves, showing strong precision across a wide range of recall values. This research emphasizes the potential of deep learning to transform bridge condition assessment by offering scalable and accurate crack detection tools.

Keywords –

Bridge condition assessment; Crack detection; Instance segmentation; YOLO; Mask R-CNN.

1 Introduction

Bridges are vital infrastructure systems for economic development and public welfare. Nonetheless, their ubiquitous nature, accelerated deteriorating, and increased public demand in the light of tight maintenance budgets pose a momentous challenge [1]. It is essential to develop innovative and reliable tools to support optimization of maintenance and repair decisions for bridges. Holistic asset management framework is needed to bridge the gap between the demand for renewal measures and the inadequate allocated resources [2].

Infrastructure asset management is a systematic approach for monitoring and maintaining physical infrastructure assets, including roads, bridges, during their entire lifecycle [3]. This approach entails evaluating, enhancing, and ultimately substituting these assets to ensure their safety, functionality, and cost-efficiency over time. The objective of infrastructure asset management is to optimize the value and lifespan of assets utilizing data-informed decisions that balance performance and cost [4]. It encompasses multiple phases that often begin with routine inspections of infrastructure assets followed by forecasting their deterioration behavior over time under various environmental and usage conditions. Subsequently, optimal maintenance plans, repairs schedules, and rehabilitation programs are determined to maximize the assets service life [3].

This research intends to investigate the first step of bridge asset management, condition assessment. In this step, bridges are inspected to detect any existing defects and determine their severity. Condition assessment models provide the foundation for capital renewal programs and restoration efforts, particularly following extensive hazards [5]. Traditional visual inspection methods though simple can be misleading due to the inherent subjectivity arising from being greatly dependent on the expertise and skills of inspectors [6]. In

addition, these methods are laborious, tedious, and hazardous in some situations [7]. As a result, newer techniques for bridge assessment, such as Ground Penetrating Radar (GPR), thermography, and image processing, have been explored. Nonetheless, their widespread adoption has been limited due to high equipment costs and the need for expert interpretation.

Recently, deep learning methods have emerged as a cost-effective alternative for assessing bridge conditions. However, most of the developed models focused on detecting cracks with little emphasis on estimating their areas/severities. This research aims to evaluate various deep learning models for instance segmentation of bridge cracks by comparing their accuracy and real-time detection capabilities . This is a hot research area due to the difficulty in attaining both high precision and recall crack detection. In infrastructure condition assessment, it is crucial to identify cracks with high precision to prevent false alarms, while simultaneously achieving high recall scores to ensure that most cracks are detected. To achieve this, innovative crack identification models based on instance segmentation will be developed. Instance segmentation is the best suited computer vision task for bridge crack detection as it identifies and segments each crack separately, enabling detailed analysis (e.g., length, width, and direction) of each crack. These details are essential for severity assessment and individual crack propagation. By isolating individual cracks, instance segmentation generates precise data for scheduling maintenance, hence ensuring structural safety.

2 Literature Review

Due to the evident drawbacks of visual inspections, both researchers and professionals opt to explore alternative methods for bridge condition assessment. For example, Alsharqawi et al. [8] developed a condition assessment model of bridge decks utilizing Ground Penetrating Radar (GPR). Employing GPR to identify subsurface anomalies enhanced detection accuracy. However, GPR is an expensive apparatus that requires skilled inspectors for data acquisition and analysis. In a different approach, Omar et al. [9] utilized infrared thermography to assess defects of concrete bridges. Although the model achieved accurate results, it was sensitive to weather conditions and surface moisture, which could significantly compromise the quality and interpretability of thermal data.

Other researchers exploited image processing methods such as edge detection and thresholding to detect defects on bridges [10,11]. However, these methods are highly sensitive to environmental conditions and image quality. They may fail in real-world scenarios characterized by varying illumination conditions [12]. In addition, the requirement to manually design feature extractors makes generalizing the obtained results rather challenging [13].

Recently, deep learning techniques have been increasingly utilized in bridge crack detection achieving remarkable performance driven by rapid advancements in computer vision [14]. Unlike image processing methods, deep learning detection techniques may directly learn from image data, facilitating self-adaptation and automated cracks feature extraction with minimal or no human intervention [14,15]. Deep convolutional neural networks (DCNNs) were successfully utilized for various crack detection tasks. These tasks are typically divided into three categories: crack classification, which detects cracks at the image level; crack object detection, which identifies cracks at the object level; and crack segmentation, which pinpoints cracks at the pixel level [16]. Most of the developed models involved either crack classification or carack object detection. Various DCNN architectures have been used for these purposes such as AlexNet [14] VGG-net [17] or ResNet [18]. For example, Zhang et al. [19] developed a detection model to locate bridge surface defects like cracks and spalling. Teng et al. [20] compared the performance of several CNNs to detect cracks on concrete bridge decks. However, these efforts lack the potential accurate estimates of the length and width of cracks, thereby rendering them less suitable as crack measurement algorithms [21].

DCNN-based crack segmentation methods can offer high-resolution detailed pixel-level crack classification [14]. Thess methods assign a class label (crack or background) to each pixel in an image. Several DCN architectures were used for crack segmentation such as RCNN-ResNet50, Faster RCNN-ResNet101, RetinaNet-ResNet50, and RetinaNet-ResNet101 [22]. Nonetheless, most of the developed segmentation models are semantic-based ones. These models attempt to capture the overall structure and semantics of the image without differentiating between individual cracks. Instance segmentation, another computer vision task, goes beyond semantic segmentation by not only assigning class labels to pixels, but also differentiating between individual instances of the same class. Very few efforts provided instance segmentation models to detect bridge cracks [23]. In addition, no attempts were reported to evaluate the performance of different DCNN architectures in instance segmentation.

The scarcity of instance segmentation models that were developed to detect bridge cracks can be attributed to several challenges such weak crack features, noisy images, and imbalanced data [24]. This study tries to address this gap by developing a novel instance segmentation model to identify and quantify cracks in bridges. In addition, this paper provides a comparative study between the performance of single-stage and two-stage detectors in cracks instance segmentation. This is

achieved by comparing the accuracy and speed of instance segmentation models developed utilizing cutting edge DCNN algorithms including (You-only-look-once) YOLOv9, YOLOv11, and Mask R-CNN. To the best of the authors' knowledge, this is the first investigation of the capabilities of these algorithms in instance segmentation tasks.

3 Methodology

Figure 1 illustrates the methodology employed in this research. After conducting literature review, crack image data were gathered from publicly available online sources. In addition, a drone was also used to collect some crack images and footages from bridges in Kuwait. Data was then preprocessed before being used to create the instance segmentation models utilizing YOLOv9, YOLOv11, and MaskR-CNN algorithms. These algorithms represent the state-of-the-art single-stage and two-stage detectors. Multiple metrics were then exploited to assess and compare the perspective performance of those algorithms in terms of detection speed and accuracy.

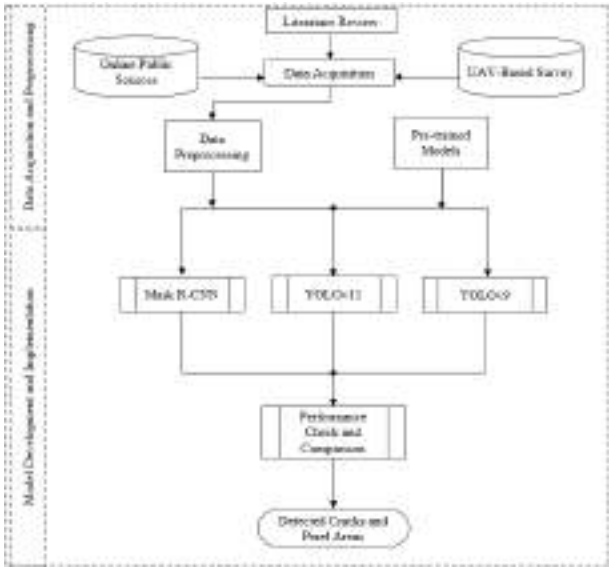


Figure 1. Research methodology

3.1 Overview of deep learning segmentation models

Deep learning models for object detection can be broadly classified into two main categories: one-stage and two-stage detectors [25]. Two-stage detectors, such as Mask R-CNN, initially produce regions of interest (ROIs) through a region proposal network [26]. The regions are then categorized and refined in the second stage to ensure accurate object localization and classification. This method is known for its accuracy resulting from the meticulous enhancement of identified objects. Unlike traditional CNNs, one-stage detectors such as YOLO process an entire image in a single pass through the network, compromising some accuracy in exchange for substantial speed enhancements [25]. One-stage models provide detections without independent region proposals and refinement stages, enabling faster operation and making them ideal for real-time applications [25].

Mask R-CNN is a deep learning model for object detection and segmentation, characterized by its precision and efficiency. It can accurately detect and each object within an image, making it exceptionally effective for complex image processing tasks. Mask R-CNN was developed by researchers at Facebook AI Research in 2017 to enhance the Faster R-CNN object detection model [26]. This was achieved by incorporating a parallel branch for predicting object masks alongside the existing bounding box detection branch [26]. Mask R-CNN's design comprises three primary components: a backbone network, a region proposal network (RPN), and two concurrent branches for bounding box identification and mask prediction. The backbone is typically a convolutional neural network that collects information from the input images to be utilized by both branches. In this analysis, we used ResNet-50 FPN as the backbone for Mask R-CNN due to its well-established balance between accuracy and computational efficiency. The ResNet-50 architecture, coupled with the feature pyramid network (FPN), enhances multi-scale feature extraction, making it highly suitable for detecting and segmenting cracks of varying sizes. The RPN produces a collection of region proposals that could encompass objects, utilizing the feature maps created by the backbone network. The bounding box branch predicts the class label and bounding box coordinates for each region proposal, whereas the mask branch generates a binary mask for each object instance within the bounding box [26]. The utilization of Mask R-CNN necessitates considerable computational resources for both training and inference [27], posing a constraint in real-time crack segmentation applications where such resources are scarce.

YOLOv9 and YOLOv11 are the latest versions of the YOLO algorithms that can be used to perform instance segmentation tasks [28-30]. The architecture of YOLO is composed of three parts: backbone, neck, and head. The backbone is the deep learning architecture that acts as a feature extractor. The neck combines the features acquired from the various layers of the backbone. The head predicts the classes and bounding box regions. In YOLOv9 the concept of programmable gradient information (PGI) was proposed. PGI provides complete input information to calculate the objective function so that reliable gradient information is obtained to update

network weights. This is achieved by introducing a new auxiliary head to the architecture that links the input data to the target output to pass the additional information. In addition, YOLOv9 incorporates reversible functions within its architecture to prevent information loss during data transmission through deep networks. A generalized efficient layer aggregation network (GELAN) was also adopted in YOLOv9 to enhance feature extraction by facilitating efficient layer aggregation, allowing the model to capture information across different layers while reducing the computational complexity [28-29].

In YOLOv11, the backbone is made of numerous convolution layers that extract distinct features at various resolution levels. These layers are connected through multiple C3K2 blocks to the neck components. C3K2 blocks replace the C2f block in the earlier versions (like YOLOv8) and employ two smaller convolutions instead of one large convolution, reducing processing time. The neck consists of spatial pyramid pooling fast (SPPF) that generates a fixed feature representation of objects regardless of their sizes without compromising spatial information. There is also a newly introduced cross-stage partial with spatial attention (C2PSA) block which facilitates learning global relationships between pixels at different positions while focusing more effectively on important regions. The neck component is connected to the head layer via three C3K2 blocks with each exclusively used to predict certain object sizes (small, medium, and large). Introducing the novel C3K2 and C2PSA blocks enhances the performance for more complex tasks with much fewer parameters [24]. While YOLOv9 offers two variants, YOLOv11 offers five variants with varying depth, width, and max channel parameters to accommodate different needs of computational speed and accuracy [30]. In this analysis, we used YOLOv9c and YOLOv11l models due to their comparable sizes in terms of the number of parameters.

3.2 Data Acquisition and Pre-processing

A comprehensive dataset of bridge crack images was carefully collected reflecting diverse real-world conditions. Images with no cracks or redundant images were removed. We initially gathered 2,650 raw images from various web sources, of which 2,408 were kept after the initial preprocessing and quality checks. Additionally, we collected 660 high-resolution images via UAV survey, giving a combined set of 3,068 images. Cracks were then annotated utilizing Roboflow, an online platform for image data management [31]. Once annotated, the dataset was divided into three subsets: training, validation, and testing with percentages of 70%, 20%, and 10%, respectively. The training dataset was used to train and calibrate the model. The validation set was used for performance enhancement while avoiding overfitting. The testing dataset was used for the final evaluation to provide an unbiased measure of the model's accuracy.

The training dataset was enriched by some augmentation techniques to simulate variations encountered in real-world scenarios, hence, improving the robustness and generalizability of the model. In this step, images were rotated randomly within a range of -15° to +15° to account for variations in object orientation. The hue values of the images were altered randomly within a range of -9° to +9°, introducing color variations to mimic different lighting conditions and reduce sensitivity to color changes. Brightness levels of the images were adjusted randomly between -15% and +15% to account for diverse illumination conditions. Finally, random noise, up to 0.1% of the total pixels, was introduced to the images to simulate distortions that may occur in real-world data acquisition. Data augmentation techniques expanded the training dataset to 6,443 images.

3.3 Model Training and Validation

All models were trained on an Nvidia Tesla T4 (16 GB memory) GPU. The stochastic gradient descent (SGDM) optimizer was used in all training experiments with a momentum set to 0.9. An initial learning rate of 0.01 was selected with a cosine learning rate decay. Cosine learning rate decay facilitates faster initial training, improved convergence, and better generalization by avoiding abrupt changes that can destabilize training. The batch size was set to 16 for all models. All models were trained for at least 50 epochs, or until their training loss converged. For fair comparison, the number of iterations for Mask R-CNN was calculated by multiplying the number of epochs with the ratio of the number of training samples to the batch size (e.g., number of iterations = epochs × (training samples/batch size)). A patience parameter was added to stop the training if the training loss plateaued with no improvements for at least 10 epochs. To avoid overfitting, batch normalization was used in YOLO models while L2 regularization was applied for Mask R-CNN.

In addition, transfer learning was used to speed up the learning process. This was achieved by exploiting pre-trained model weights as starting points for training the cracks instance segmentation models. Configurations and associated pre-trained weights for Mask R-CNN and YOLO models were retrieved from the Detectron 2 and perspective GitHub repositories, respectively [29-30,32].

3.4 Performance Evaluation

The aim of this step is to evaluate the efficacy of the developed models in bridge cracks detections and segmentation. The accuracy was assessed utilizing the popular object detection evaluation metric, mean average precision (mAP). Average precision (AP) is defined as

the average detection precision under various recalls, computed independently for each class. *AP* is calculated as the numerical integration of the precision-curve:

$$AP = \int_0^1 p(r)dr \quad (1)$$

Where p is the precision and r is the recall. The mean of the average precision among all considered classes, mAP, provides a holistic assessment of the detector's performance. To obtain the precision recall curve, classification predictions and the intersection-over-union (*IOU*) between the predicted bounding boxes (B_p) and ground truth ones (B_g) shall be calculated for each image:

$$IOU = area(Bp \cap Bg) / area(Bp \cup Bg) \quad (2)$$

True positive, false positive and false negative predictions are defined based on the *IOU* and some predefined threshold value. In this context, true positive represent the correct prediction where (IOU > threshold + correct classification). Similarly, false positives represent that the incorrect predictions with (IOU < threshold). On the other hand, false negative denotes that undetected, objects. Thus, precision (p) is defined as the proportion of correctly predicted positive instances out of all predicted positives while recall (r) is the proportion of correctly predicted positive instances out of all actual positives. These metrics are given by the equations below:

$$p = TP/(TP + FP) \quad (3)$$
$$r = TP/(TP + FN) \quad (4)$$

Where TP is true positive predictions, FP is false positive predictions, and FN is false negative predictions. In this study, *AP* is calculated at *IOU* threshold of 0.50, yielding mAP0.50.

4 Results and discussion

Figure 2 depicts the loss curves for the training and validation phases using the YOLOv11 model. The training loss curve exhibits a consistent and gradual decline, demonstrating effective learning of the model. Similarly, the validation loss curve shows a steady decrease, indicating the model's ability to generalize on unseen data. The convergence of the training loss to a low value (less than 0.4) reflects the model's successful acquisition of complex features required for the bridge crack detection task. While similar trends were observed for YOLOv9 and Mask R-CNN detectors, these curves are not included here due to space constraints.

In segmentation tasks, the bounding box regression measures how accurately the model is predicting the bounding box coordinates. On the other hand, segmentation prediction measures how well the model is predicting the pixel-level mask or outline of the detected crack.

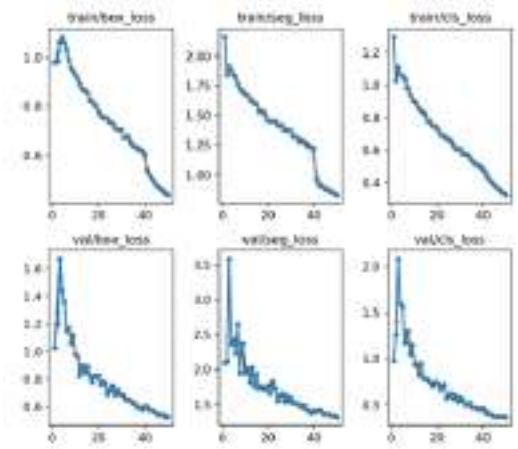


Figure 2. Training vs. validation loss for YOLOv11

Figure 3 summarizes the achieved mean average precision (mAP0.50) of both box and mask predictions for each detector. It can be observed that all three models achieved high performance in bounding box detection, with mAP0.50 exceeding 94%. Mask R-CNN achieved the highest score (95.59%), closely followed by YOLOv9 (95.0%) and YOLOv11 (94.9%). However, YOLOv9 and YOLOv11 achieved higher segmentation accuracies with mAP0.50 of 87.8% and 87.1%, respectively, reflecting strong performance in pixel-level accuracy. Mask R-CNN lags at 75.5% possibly due to limitations in handling finer details compared to the anchor-free architecture of YOLO models. In addition, YOLO models were able to significantly reduce the training time compared to the Mask R-CNN detector. For example, YOLOv11 reduced training time by approximately 82.6% from 12.22 hours for Mask R-CNN model to only 2.12 hours.

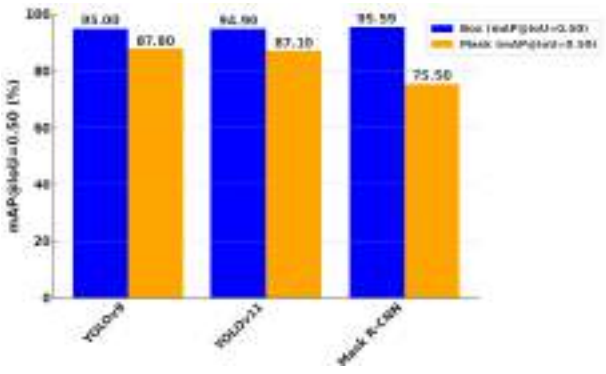


Figure 3. mAP of the investigated models

Due to their superior detection speed compared to Mask R-CNN, Table 1 depicts a detailed performance assessment of YOLOv9 and YOLOv11 models. Based on Table 1, YOLOv11 presents a better trade-off between precision, inference speed, and training time compared to

YOLOv9. YOLOv11 achieves a higher precision in bounding box detection (96.0%) and almost similar segmentation precision (91.2%), while maintaining faster inference speed of 19 ms/image and significantly shorter training time of 127.49 minutes. Regarding the recall metric, YOLOv9 achieved slightly higher bounding box recall (95.5%) compared to YOLOv11 (93.8%). Similar results were obtained for segmentation recall metrics as can be shown in Table 1 suggesting that YOLOv9 is better at identifying more crack objects in an image. Despite this marginal reduction in recall for both bounding box detection and segmentation, YOLOv11 overall performance and efficiency make it highly suitable for real-time applications. The smaller model size of YOLOv11 (27.6M parameters) further contributes to its speed and practical applicability. YOLOv9 on the other hand has competitive recall and mAP0.50 metrics, suggesting it may detect more objects when longer inference times are acceptable where exhaustive crack identification is required.

Table 1 Detailed performance assessment

Metric	YOLOv9	YOLOv11
Bounding Box		
Precision (p)	93.3%	96.0%
Recall (r)	95.0%	93.8%
mAP@IoU=0.50	95.0%	94.9%
Segmentation		
Precision (p)	91.5%	91.2%
Recall (r)	91.0%	89.0%
mAP@IoU=0.50	87.8%	87.1%
Training Time	270.91mnts	127.49mnts
Inference Time	34.1ms/image	19ms/image
Number of Parameters (M)	27.9	27.6

Table **2** presents performance comparison of the developed models with previous studies in the literature. Our YOLOv9 and YOLOv11 models outperform several other models in terms of accuracy while still maintaining competitive inference speeds. YOLOv11 offers the optimal trade-off between accuracy, speed, and computational efficiency.

Table 2 Performance comparison with prior studies

Model	mAP0.50 (%)	Inference time (ms/image)	Params (M)
YOLOv9	87.8%	34	27.9
YOLOv11	87.1%	19	27.6
YOLOv7 [22]	74.8%	22	36.9
RetinaNet [22]	72.8%	65	56.0
YOLOMF [33]	77.3%	12	11.5

The Precision-Recall curve in Figure 4 demonstrates YOLOv11 performance in crack instance segmentation task. With a mean average precision (mAP) of 0.871 at IoU=0.50, the model achieves a high level of precision across a wide range of recall values. The near-plateau at high precision values indicates the model's ability to maintain accurate predictions even as recall increases, indicating robust detection capabilities. The steep drop at the tail of the curve suggests a natural trade-off between precision and recall as the threshold for detection decreases. Overall, the curve highlights effectiveness in achieving a strong balance between precision and recall for crack detection and segmentation tasks.

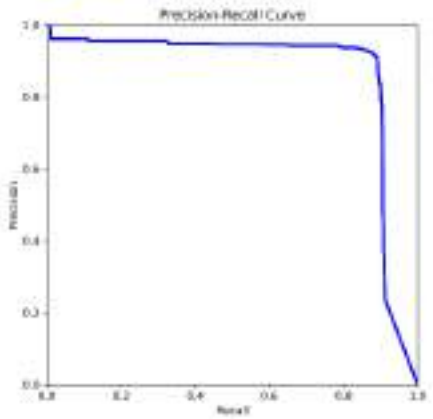


Figure 4. Precision-Recall curve for crack detection using YOLOv11

Figure 5 displays segmentation results for crack instance segmentation utilizing the YOLOv11 model. The blue overlays highlight the detected crack regions, with confidence scores labeled for each prediction. The figure demonstrates the model's capability to capture fine details and complex geometries of cracks, even in challenging cases with varying shapes and sizes. The high-confidence predictions (e.g., Crack 1.0) indicate robust performance. To calculate the pixel area of each identified crack, each segmented mask generated by the model was converted to a grayscale region. The boundaries of the segmented regions were identified, and the pixel area was calculated using OpenCV's contour methods [34]. This provides the total number of pixels enclosed within each mask, representing the crack's segmented area in the image. The calculated pixel areas can then be converted into real-world units if the spatial resolution of the image is known.

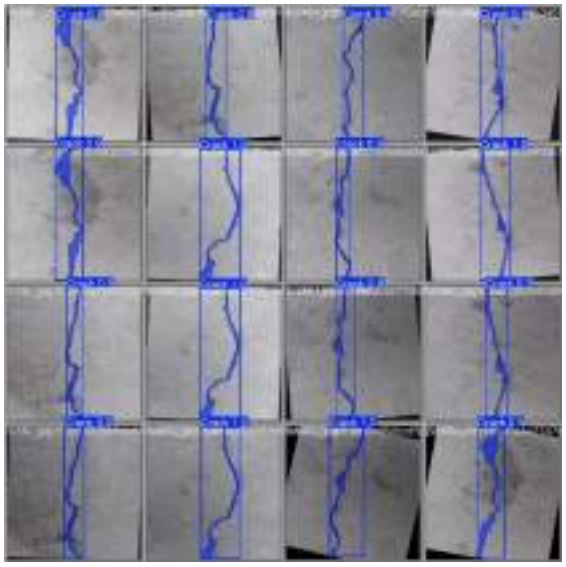


Figure 5. Sample of crack detection and segmentation using YOLOv11

5 Conclusion

Despite the extensive advancements of deep learning in crack detection, the development of instance segmentation models capable of assessing crack severity remains significantly underexplored. This study provides a comprehensive evaluation of single-stage and two-stage detectors for automated crack detection and segmentation in bridges. The findings highlight that YOLOv9 and YOLOv11 excel in speed and efficiency, while Mask R-CNN with ResNet-50 FPN demonstrates competitive accuracy. Notably, YOLO models achieved almost same accuracy with significant time savings, reducing training time by approximately 82% and inference time by 76%. This makes YOLO models highly practical for real-time crack detection applications. Leveraging novel architectures, data augmentation techniques, and transfer learning, YOLOv11 balances precision (91.5%) and recall (91.0%), presenting an effective candidate for large-scale infrastructure inspections. The results emphasize the importance of selecting segmentation models based on specific application requirements, such as prioritizing speed for real-time detection or accuracy for detailed analysis.

The developed segmentation models can be readily integrated with existing bridge management systems. UAVs can be used to capture high-resolution images, which are processed to extract and quantify crack features. The results can be shown on mobile applications for real-time assessments complementing traditional visual inspections. This integration combines automated analysis and expert judgment to enhance bridge evaluations and enable more informed maintenance and repair decisions.

While the model performed well in controlled conditions, environmental factors may impact its accuracy during real-world inspections. Variations in lighting, differences in crack textures, and weathering can challenge segmentation. Although our dataset was enhanced using augmentation techniques to simulate real-world scenarios, these measures may not capture all field complexities. Future work should thoroughly explore these effects and further enhance the model by focusing on quantifying the actual areas of cracks.

References

[1] A. Assad and A. Bouferguene. Machine learning-based deterioration modeling of highway bridges considering climatic conditions. *Interdisciplinary Symposium on Smart & Sustainable Infrastructures*, pages 1039–1051, Vancouver, Canada, September 2023.

[2] A. Assad and A. Bouferguene. Enhancing bridge management: A data-driven approach for accurate forecasting of concrete bridge condition. *Journal of Engineering Research*, 2307–1877, 202

[3] L. Dromey, K. Ruane, J. J. Murphy, B. O'Rourke, and S. Lacey. A bridge-rehabilitation strategy based on the analysis of a bridge-inspection data set. *Infrastructure Asset Management*, 7(1):25–35, 2020.

[4] N. S. Grigg. *Infrastructure finance: the business of infrastructure for a sustainable future*, volume 536. John Wiley & Sons, 2010.

[5] N. Faris, T. Zayed, A. Fares, S. Abdelkhalek, and E. M. Abdelkader. Automated rebar recognition and corrosion assessment of concrete bridge decks using ground penetrating radar. *Automation in Construction*, 166:105631, 2024.

[6] E. Abdelkader, O. Moselhi, M. Marzouk, and T. Zayed. Hybrid Elman neural network and an invasive weed optimization method for bridge defect recognition. *Transportation Research Record*, 2675(3):167–199, 2021.

[7] I. H. Kim, H. Jeon, S. C. Baek, W. H. Hong, and H. J. Jung. Application of crack identification techniques for an aging concrete bridge inspection using an unmanned aerial vehicle. *Sensors*, 18(6):1881, 2018.

[8] M. Alsharqawi, T. Zayed, and S. Abu Dabous. Integrated condition-based rating model for sustainable bridge management. *Journal of Performance of Constructed Facilities*, 34(5):04020091, 2020.

[9] T. Omar, M. L. Nehdi, and T. Zayed. Infrared thermography model for automated detection of delamination in RC bridge decks. *Construction and*

Building Materials, 168:313–327, 2018

[10] Y. Pan and L. Zhang. Roles of artificial intelligence in construction engineering and management: A critical review and future trends. *Automation in Construction*, 122:103517, 2021.

[11] R. Fan, M. J. Bocus, Y. Zhu, J. Jiao, L. Wang, F. Ma, S. Cheng, and M. Liu. Road crack detection using deep convolutional neural network and adaptive thresholding. *Automation in Construction*, 2019.

[12] J. Zhong, J. Zhu, J. Huyan, T. Ma, and W. Zhang. Multi-scale feature fusion network for pixel-level pavement distress detection. *Automation in Construction*, 141:104436, 2022.

[13] H. Jiang, X. Hu, K. Li, J. Zhang, J. Gong, and M. Zhang. PGA-SiamNet: Pyramid feature-based attention-guided siamese network for remote sensing orthoimagery building change detection. *Remote Sensing*, 12:484, 2020.

[14] R. Ali, J. H. Chuah, M. S. A. Talip, N. Mokhtar, and M. A. Shoaib. Structural crack detection using deep convolutional neural networks. *Automation in Construction*, 133:103989, 2022.

[15] Z. Tong, J. Gao, and H. Zhang. Innovative method for recognizing subgrade defects based on a convolutional neural network. *Construction and Building Materials*, 169:69–82, 2018.

[16] S. Zhou and W. Song. Concrete roadway crack segmentation using encoder-decoder networks with range images. *Automation in Construction*, 120:103403, 2020.

[17] S. Kulkarni, S. Singh, D. Balakrishnan, S. Sharma, S. Devunuri, and S. C. R. Korlapati. CrackSeg9k: A collection and benchmark for crack segmentation datasets and frameworks. In *European Conference on Computer Vision*, pages 179–195, Cham, Switzerland, October 2022. Springer Nature Switzerland.

[18] J. Wang, X. He, S. Faming, G. Lu, H. Cong, and Q. Jiang. A real-time bridge crack detection method based on an improved inception-resnet-v2 structure. *IEEE Access*, 9:93209–93223, 202.

[19] C. Zhang, C. Chang, and M. Jamshidi. Concrete bridge surface damage detection using a single-stage detector. *Computer-Aided Civil and Infrastructure Engineering*, 35:389–409, 2020.

[20] S. Teng, Z. Liu, G. Chen, and L. Cheng. Concrete crack detection based on well-known feature extractor model and the YOLO_v2 network. *Applied Sciences*, 11:813, 2021.

[21] L. Yu, S. He, X. Liu, S. Jiang, and S. Xiang. Intelligent crack detection and quantification in the concrete bridge: A deep learning-assisted image processing approach. *Advances in Civil Engineering*, 2022:1813821, 2022.

[22] T. Tran, S. Nguyen, H. Lee, and V. Tran. Advanced crack detection and segmentation on bridge decks using deep learning. *Construction and Building Materials*, 400:132839, 2023.

[23] J. Li, C. Yuan, and X. Wang. Real-time instance-level detection of asphalt pavement distress combining space-to-depth (SPD) YOLO and omni-scale network (OSNet). *Automation in Construction*, 155:105062, 2023.

[24] N. Nguyen, S. Perry, D. Bone, H. Le, and T. Nguyen. Two-stage convolutional neural network for road crack detection and segmentation. *Expert Systems with Applications*, 186:115718, 2021.

[25] M. Carranza-García, J. Torres-Mateo, P. Lara-Benítez, and J. García-Gutiérrez. On the performance of one-stage and two-stage object detectors in autonomous vehicles using camera data. *Remote Sensing*, 13(1):89, 2020.

[26] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2961–2969, 2017.

[27] L. Zhang, J. Wu, Y. Fan, H. Gao, and Y. Shao. An efficient building extraction method from high spatial resolution remote sensing images based on improved Mask R-CNN. *Sensors*, 20(5):1465, 2020.

[28] C. Y. Wang, I. H. Yeh, and H. Y. Mark Liao. YOLOv9: Learning what you want to learn using programmable gradient information. In *European Conference on Computer Vision*, pages 1–21. Springer, Cham, 2025.

[29] Ultralytics, YOLOv9. On-line: https://docs.ultralytics.com/models/yolov9/, Accessed: 23/12/2024.

[30] G. Jocher and J. Qiu. Ultralytics YOLO11, version 11.0.0, 2024. On-line: *https://github.com/ultralytics/ultralytics*, Accessed: 23/12/2024.

[31] B. Dwyer, J. Nelson, T. Hansen. Roboflow (Version 1.0) [Software]. On-line: https://roboflow.com, Accessed: 23/12/2024.

[32] Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, and R. Girshick. Detectron2, 2019. On-line: *https://github.com/facebookresearch/detectron2*, Accessed: 23/12/2024.

[33] T. Jiang, L. Li, B. Samali, Y. Yu, K. Huang, W. Yan, & L. Wang. Lightweight object detection network for multi-damage recognition of concrete bridges in complex environments. *Computer-Aided Civil and Infrastructure Engineering*, *39*(23), 3646-3665, 2024.

[34] OpenCV Development Team. Open-Source Computer Vision Library. On-line: *https://opencv.org/*, Accessed: 23/12/2024.

Enhancing Building Semantics Preservation in AI Model Training with Large Language Model Encodings

Suhyung Jang[1], Ghang Lee[1,2], Jaekun Lee[1], and Hyunjun Lee[1]

[1]Building Informatics Group, Department of Architecture and Architectural Engineering, Yonsei University, Republic of Korea
[2]Institute for Advanced Studies, Technical University of Munich, Germany
rgb000@yonsei.ac.kr, glee@yonsei.ac.kr, jake102518@yonsei.ac.kr, hjlee97@yonsei.ac.kr

Abstract

Accurate representation of building semantics—encompassing both generic object types and specific subtypes—is essential for effective AI model training in the architecture, engineering, construction, and operation (AECO) industry. Conventional encoding methods (e.g., one-hot) often fail to convey the nuanced relationships among closely related subtypes, limiting AI's semantic comprehension. To address this limitation, this study proposes a novel training approach that employs large language model (LLM) embeddings (e.g., OpenAI GPT and Meta LLaMA) as encodings to preserve finer distinctions in building semantics. We evaluated the proposed method by training GraphSAGE models to classify 42 building object subtypes across five high-rise residential building information models (BIMs). Various embedding dimensions were tested, including original high-dimensional LLM embeddings (1,536, 3,072, or 4,096) and 1,024-dimensional compacted embeddings generated via the Matryoshka representation model. Experimental results demonstrated that LLM encodings outperformed the conventional one-hot baseline, with the "llama-3 (compacted)" embedding achieving a weighted average F1-score of 0.8766, compared to 0.8475 for one-hot encoding. The results underscore the promise of leveraging LLM-based encodings to enhance AI's ability to interpret complex, domain-specific building semantics. As the capabilities of LLMs and dimensionality reduction techniques continue to evolve, this approach holds considerable potential for broad application in semantic elaboration tasks throughout the AECO industry.

Keywords –

Building semantics; Large language model (LLM) embedding; Building information model (BIM); Graph neural network (GNN); Artificial intelligence (AI)

1 Introduction

The effective application of artificial intelligence (AI) technologies within the architecture, engineering, construction, and operation (AECO) industry hinges on the accurate representation of construction project information in machine-comprehendible formats. Extensive research has focused on utilizing diverse high-level data formats [1]—such as site photos for real-time site and worker condition monitoring [2,3], point clouds for tracking as-built construction progress [4–7], building information model (BIM) graphs for representing building topologies [8–10] and textual descriptions of project intents [11–13]—to represent information relevant to deliver the contextual information of the building projects. Representing and interpreting these digitally represented project information associated with planning, designing, constructing, and operating building projects (i.e., concepts that are significant for managing the projects), referred to as "building semantics" in this paper, are not only essential for AI systems to support reliable and informed decisions in construction projects, but also for their training.

To enable AI models to comprehend the high-level data format through supervised learning, this high-level information is labeled (partially or fully) and entered into AI models utilizing encoding methods. While these encodings play a critical role in differentiating the classes during training, prior studies often overlook the selection of encoding method, defaulting to conventional methods such as one-hot or label encodings. Although the advent of large language models (LLMs) and their generated embeddings has demonstrated the ability to capture domain-specific contextual nuances within the AECO fields [14,15], their utilization has primarily been confined to retrieving semantically similar information. The potential of leveraging these semantically rich LLM embeddings to provide task-relevant contextual nuances during AI model training remains largely unexplored.

This study proposes a novel AI model training

method that employs LLM embeddings as encodings (i.e., "LLM encoding") to preserve finer distinctions between building semantics. To evaluate performance differences compared to conventional encoding methods, we conducted an experiment focusing on a building object subtype classification task using GraphSAGE models [16]. For 42 building object subtypes and 5 BIM models utilized in the practice by a major contractor in the Republic of Korea, GraphSAGE models were trained for the node classification task. The experiment was designed to contrast one-hot encoding with LLM encodings generated by various models, including OpenAI's 'text-embedding-3-small' and 'text-embedding-3-large' [17], as well as Meta's 'llama3' [18]. Additionally, we investigated diverse embedding dimensions, compacted using the Matryoshka representation model, to assess whether compacted LLM embedding still preserves semantic nuances during the training.

The remainder of this paper is organized as follows: Section 2 introduces the background and motivation for this study, including a review of previous research on semantic enrichment. Section 3 explains how LLM embeddings can be used as encodings to train AI models. Section 4 describes the experimental design used to validate the proposed method. Section 5 presents the experimental results and Section 6 discusses academic and practical implications, current limitations, and future studies. Finally, Section 7 concludes the paper.

2 Background

With the continually increasing efforts to employ AI in the AECO industry, there has been growing interest in training AI to understand the terms and concepts used in our field. In other words, there is a recognized need to teach computers the terminology and concepts we use to plan, design, construct, and operate construction projects. This study collectively refers to the information used in managing building projects as "building semantics."

The understanding of these building semantics by AI is largely achieved via supervised learning. In supervised learning, the objects of study—namely, the object, the concept (or class) in the target domain, and the encoding method used to distinguish and convey them—correspond to the components of the triangle of reference (i.e., the semantic triangle) [19], which are *referent*, *thought* (*reference*), and *symbol*. For example, in AI training aimed at classifying BIM object types, the wall objects in a BIM model (the *referent* in Figure 1) belong to the "Wall" category (*reference*). This might be represented to a computer as [1, 0, 0, 0, 0, 0], i.e., the first of six object types (*symbol*).

Previous research has shown that building semantics—that is, the diverse concepts or classes that exist in real or digital project representations—encompasses a wide range of meanings. For instance, to verify whether a construction worker on site is wearing the proper personal protective equipment (PPE), such as straps, hardhats, harnesses, and hooks, AI is used to recognize different types of PPE and check correct usage [2,3]. For studies focused on construction progress monitoring [4–7], AI that recognizes objects on an as-built construction site and tracks progress compared to the as-planned schedules. Other studies aim to predict object types or their subtypes in BIM models [8–10]. In short, there has been extensive research on enabling AI to interpret a broad range of building semantics.

Furthermore, the building objects used as examples of these concepts and classes, as well as the subjects of analysis, have been represented in ways that are suitable for AI comprehension. For instance, when monitoring the correct use of PPE, images of workers on site can be used; for as-built progress monitoring, point clouds can be adopted; and for object relationships in a BIM model, graph data can be employed. These representations help convey real or digitally depicted states to AI in forms appropriate for training.

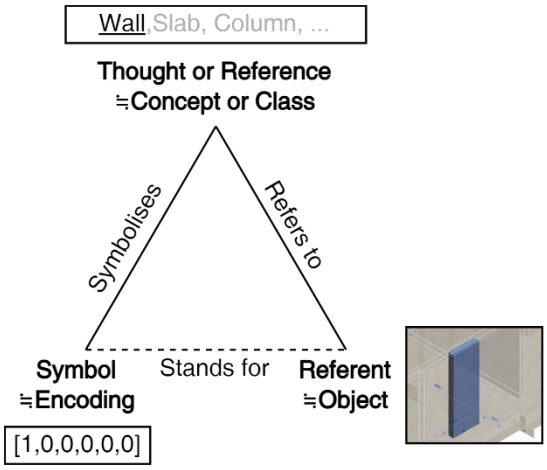


Figure 1. The triangle of reference (i.e., semantic triangle) in AI training for building semantics.

All these efforts show how to convey objects (the *referent*) to AI models, enabling them to interpret concepts (the *reference*) to understand building semantics. However, the types and effects of symbol—that is, the encodings used to help AI distinguish among concepts and classes—are comparatively underexplored. Although state-of-the-art large pretrained models can directly interpret concepts that are provided as images or text (i.e., can directly interpret referents), such models demand excessively high computational costs and datasets, making it difficult to train or fine-tune them for

new purposes. Hence, it is more common to develop smaller models via supervised learning. In this context, conventional encoding methods that do not capture the relative similarities among classes and concepts (e.g., labels or one-hot encoding) have long been used without much question. For instance, one-hot encoding treats all categories as equidistant from one another, as illustrated in Figure 2A, where generic building object types ("Wall," "Column," "Foundation," "Beam," "Stair," and "Slab"), while ignores inherent similarities and differences between the categories.

This study proposes using semantically rich LLM embeddings (i.e., LLM encoding) as the encoding method for supervised training, thereby preserving building semantics. Unlike one-hot encoding, LLM embedding uses semantic knowledge learned from large training corpora to generate highly contextualized vector representations that can capture the relative similarities and differences among building semantics. Figures 2A and 2B illustrate how LLM embeddings represent semantic distances between generic object types (i.e., "Wall" and "Slab") and their subtypes using t-SNE visualization for OpenAI's 'text-embedding-3-large' (3,720 dimensions) and Meta's 'llama-3' (4,096 dimensions). In both Figures 2A and 2B, the embeddings form clearly distinct clusters of subtypes—one each for the generic types "Wall" and "Slab" (with a horizontal distinction in Figure 2B and a diagonal distinction in Figure 2C). In addition, both LLM embeddings recognize that parapet-related wall subtypes are more closely grouped together, compared to other wall subtypes (highlighted with gray backgrounds in Figures 2A and 2B). Comparing the x-axis and y-axis scales for OpenAI's GPT-based embedding in Figure 2B and Meta′s LLaMA-based embedding in Figure 2C, it is evident that the two embeddings use different scales to distinguish each subtype (for example, $x \in [1,2]$, $y \in [-1.2,0]$ for 'text-embedding-3-large' in Figure 2B, and $x \in [1.2,2.6]$, $y \in [8,10]$ for 'text-embedding-3-large' in Figure 2C). Additionally, judging from the proportions of the ellipses around the parapet subtype clusters, we can see that the two LLM embeddings perceive the building semantics in subtly different ways.

Although it is reasonable to suspect that LLM embeddings might better symbolize the relative building semantics compared to one-hot encoding, this hypothesis remains underexplored. Building upon the premise that incorporating such building semantics into an AI model's encoding could enable more effective supervised learning, this study was conducted to propose a method that makes this approach feasible in practice.

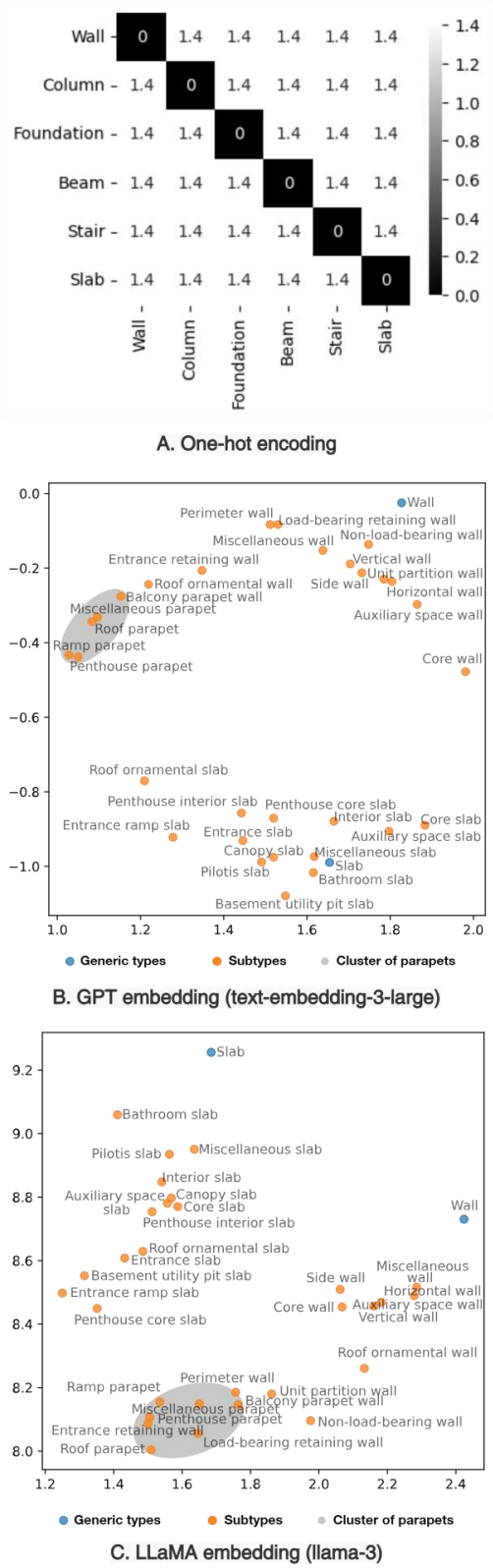


Figure 2. Building semantics represented using diverse encoding types.

3 LLM Encoding and Matryoshka Representation Model

3.1 LLM Encoding

Typically, encodings are used to enable computers to distinguish between given classes. In neural network training, the model evaluates whether the output—derived from label encoding or one-hot encoding—matches the target class via a sigmoid function. If the output aligns with the target, the loss decreases, and training proceeds to minimize that loss. However, to utilize LLM embeddings as encodings (i.e., LLM encoding) in neural network training, the loss calculation method needs modification. LLM embeddings reside in high-dimensional spaces, capturing various semantic nuances. Applying a sigmoid function, which retains only the highest-value cell, can dilute these semantic features and hinder proper learning.

Instead of using a sigmoid at the final layer, LLM encoding sets the final layer's dimension to match that of the target LLM embedding. The loss is then computed based on the quantitative difference between the neural network's output embedding e_p and the target label's LLM embedding e_t. One way to achieve this is by using cosine embedding loss, defined as equation (1):

$$L(\mathrm{e}_p, \mathrm{e}_t) = 1 - \frac{\mathrm{e}_p \cdot \mathrm{e}_t}{\|\mathrm{e}_p\| \|\mathrm{e}_t\|} \quad (1)$$

where $\cdot$ denotes the dot product and $\|\cdot\|$ denotes the Euclidean norm. This approach preserves semantic relationships in the embedding space, unlike one-hot encodings or sigmoid-based methods that force a single "correct" cell to dominate.

This training method is applicable to general tasks where each class carries semantic meaning. Once trained, the model can return the class closest to the predicted output by comparing the predicted embedding e_p against each candidate embedding e_t and selecting the one with the smallest loss.

3.2 Matryoshka Representation Model

LLMs generate embeddings in high-dimensional spaces, such as OpenAI's text-embedding-3-small (1,536 dimensions), text-embedding-3-large (3,072 dimensions), or Meta's LLaMA-3 (4,096 dimensions), to encode rich semantic information. However, leveraging these embeddings for real-world applications often requires dimensionality reduction to address computational inefficiencies. The Matryoshka representation model addresses this challenge by projecting high-dimensional embeddings into lower-dimensional spaces while preserving semantic features [20]. For implementation, the 'tomaarsen/mpnet-base-nli-matryoshka' model available on Hugging Face, was utilized.

4 Experiments

The experiment aims to evaluate whether utilizing LLM encoding improves the performance of AI models in enriching building semantics. Specifically, we assessed the AI model's performance on a subtype classification task of semantic elaboration, as depicted in Figure 3.

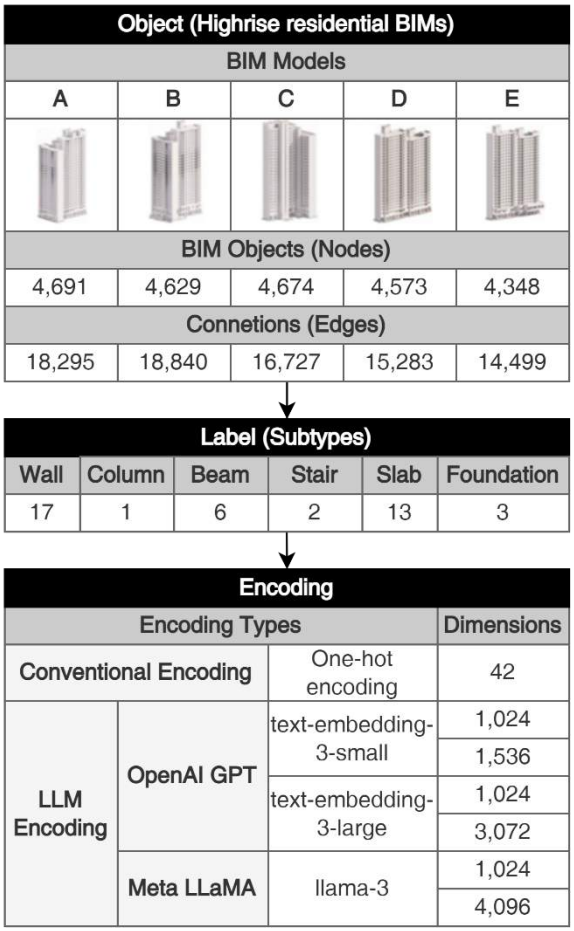


Figure 3. Experimental design.

For the experiment, five BIM models of high-rise residential projects, denoted A, B, C, D, and E, each containing over four thousand generic BIM objects, were utilized. Across these projects, we identified 42 subtypes managed by one of the major contractors in Republic of Korea (Table 1), ranging from "core wall" and "bathroom slab" to "transfer column" and "penthouse core slab from our prior study [10]. These projects were cross-validated by using each as a test dataset while the others served as training datasets. On these datasets, we trained GraphSAGE [21] models with three layers, each having a layer dimension of 1,024. Within this setup, we

tested various encodings with varying dimension sizes using the Matryoshka Representation Model. The encodings included: (1) one-hot encoding, a conventional method; (2) 'text-embedding-3-small' with a 1,024-dimension compacted representation; (3) 'text-embedding-3-small' with its original 1,536 dimensions; (4) 'text-embedding-3-large' with a 1,024-dimension compacted representation; (5) 'text-embedding-3-large' with its original 3,072 dimensions; (6) 'llama-3' with a 1,024-dimension compacted representation; and (7) 'llama-3' with its original 4,096 dimensions.

Table 1. Subtypes utilized in the experiment.

Code	Subtype	Code	Subtype
0	Core wall	21	Wall girder
1	Horizontal wall	22	Interior lintel
2	Vertical wall	23	Non-loadbearing lintel
3	Perimeter wall	24	Entrance stairs
4	Side wall	25	Interior stairs
5	Entrance retaining wall	26	Core slab
6	Loadbearing retaining wall	27	Entrance slab
7	Miscellaneous wall	28	Entrance ramp slab
8	Unit partition wall	29	Piloti slab
9	Non-loadbearing wall	30	Basement utility pit slab
10	Auxiliary space wall	31	Interior slab
11	Roof ornamental wall	32	Bathroom slab
12	Roof parapet	33	Miscellaneous slab
13	Penthouse parapet	34	Auxiliary space slab
14	Ramp parapet	35	Penthouse interior slab
15	Balcony parapet wall	36	Penthouse core slab
16	Miscellaneous parapet	37	Roof ornamental slab
17	Transfer column	38	Canopy slab
18	Core beam	39	Entrance strip foundation
19	Transfer beam	40	Mat foundation
20	Miscellaneous beam	41	Haunch

To compare the impact of different encoding strategies on the model's ability to classify building object subtypes, the average F1-score was primarily utilized. In addition, statistical tests were conducted to determine the significance of the performance differences between encoding methods. The normality of F1-score differences was tested using the Shapiro-Wilk test with a significance level of $\alpha = 0.05$. When the normality assumption was satisfied ($p > 0.05$), the paired t-test was employed to evaluate the statistical significance of the mean differences between paired observations. In cases where the normality assumption was violated ($p \leq 0.05$), the Wilcoxon signed-rank test was utilized as a non-parametric alternative, with statistical significance set at $\alpha = 0.05$.

5 Results

Table 2 presents the average F1-scores for classifying each subtype using both one-hot and LLM-based encodings. The encodings examined include text-embedding-3-small, text-embedding-3-large, and LLaMA-3, each tested in both compacted (1,024-dimensional) and original formats (1,536, 3,072, or 4,096 dimensions). Overall, subtypes with larger training and test sets demonstrated higher predictive performance.

Table 2. Weighted average F1-scores for different encoding types.

Encoding type	Dimensions	Weighted average F1-score
One-hot encoding	42	0.8475
text-embedding-3-small	1,024	0.8705
	1,536 (original)	0.8498
text-embedding-3-large	1,024	0.8655
	3,072 (original)	0.8529
llama-3	1,024	0.8766
	4,096 (original)	0.8714

With the original LLM embeddings, the weighted average F1-score increased sequentially from one-hot encoding (0.8475) to text-embedding-3-small (0.8498), text-embedding-3-large (0.8529), and finally LLaMA-3 (0.8714). However, when using the compacted 1,024-dimensional embeddings, the ranking of text-embedding-3-large and text-embedding-3-small reversed compared to their original counterparts, resulting in a slight overall increase in weighted average F1-scores to 0.8705, 0.8655, and 0.8766, respectively. These findings suggest that GraphSAGE, which outputs a 1,024-dimensional representation, may not adequately capture the semantic richness of higher-dimensional LLM embeddings (e.g., 1,536, 3,072, or 4,096 dimensions).

5.1 One-hot and LLM Encodings

In this subsection, we analyzed the significance of performance differences between one-hot and LLM encodings, including both original and compacted formats. The Shapiro-Wilk normality test (n=42) indicated that the normality assumption was violated in the F1-score differences for five out of six comparisons (p-values < 0.05), necessitating the use of the non-parametric Wilcoxon signed-rank test (Table 3). Among LLM embeddings, 'llama-3 (compacted)' exhibited a distinctly low p-value (0.000007), indicating a more skewed fold-by-fold performance distribution that may arise from uneven semantic feature retention after the largest dimensional compression (4,096 to 1,024).

Table 3. p-values of the Shapiro-Wilk test for identifying the normality of the difference in performance between one-hot and LLM encodings.

Shapiro-Wilk test	One-hot encoding
text-embedding-3-small (compacted)	0.065730
text-embedding-3-small (original)	0.006618
text-embedding-3-large(compacted)	0.013222
text-embedding-3-large (original)	0.011914
llama-3 (compacted)	0.000007
llama-3 (original)	0.049937

The results of the paired t-test and Wilcoxon signed-rank test (n=42) are presented in Table 4. Although the performance of the model slightly improved when LLM encodings were utilized, statistical analysis revealed that no significant differences were found, except for 'text-embedding-3-large (compacted)' (p=0.006596). This suggests that compressing text-embedding-3-large to 1,024 dimensions appears to preserve key semantic cues while removing noise, thus yielding more consistent performance gains and a statistically significant improvement over one-hot encoding.

Table 4. p-values of paired t-test and Wilcoxon signed-rank test for identifying the significance of the difference in performance between one-hot and LLM encodings.

Paired t-test and Wilcoxon signed-rank test	One-hot encoding
text-embedding-3-small (compacted)	0.285281 (z=1.06852)
text-embedding-3-small (original)	0.873084 (t = -0.16074)
text-embedding-3-large(compacted)	0.006569 (z=2.71792)
text-embedding-3-large (original)	0.163854 (z=1.39222)
llama-3 (compacted)	0.714148 (z= 0.36629)
llama-3 (original)	0.219283 (z= 1.22843)

5.2 Between LLM Encodings

In this subsection, we analyzed the statistical significance between different LLM encodings. Similar to the one-hot comparisons, the Shapiro-Wilk test (n=42) identified violations of normality assumptions in all cases, leading us to employ the Wilcoxon signed-rank test (n=42) for all comparisons. For original LLM embeddings, no significant differences were observed despite slight performance increases (Table 5).

Table 5. p-values of Wilcoxon signed-rank test for identifying the significance of the difference in performance between original LLM encodings.

Wilcoxon signed-rank test	text-embedding-3-large (original)	llama-3 (original)
text-embedding-3-small (original)	0.068390 (z = 1.82243)	0.529727 (z = 0.62842)
text-embedding-3-large (original)	-	0.751786 (z = 0.31628)

Conversely, for compacted LLM encodings, significant differences were observed (Table 6). This demonstrates that compacted LLM embeddings effectively leverage the semantic richness provided by large datasets and advanced architectures. For example, 'llama-3 (compacted)' benefits from its extensive training corpus of over 1.4 trillion tokens and 65 billion model parameters, compared to the smaller 'text-embedding-3-large (compacted)'.

Table 6. p-values of Wilcoxon signed-rank test for identifying the significance of the difference in performance between compacted LLM encodings.

Wilcoxon signed-rank test	text-embedding-3-large (compacted)	llama-3 (compacted)
text-embedding-3-small (compacted)	0.010346 (z = 2.56403)	0.011765 (z = 2.51910)
text-embedding-3-large (compacted)	-	0.000583 (z = 3.43962)

These results suggest that as LLM technology continues to evolve, the utility of LLM encodings for AI-based training will further improve. A notable observation is the consistent outperformance of compacted LLM encodings compared to their original counterparts, highlighting the effectiveness of

dimensionality reduction techniques like the Matryoshka representation model in preserving essential semantic information. Additionally, the findings imply that increasing the size of AI models may be necessary to fully harness the potential of semantically rich LLM encodings. Future research should explore the optimal balance between model size and computational efficiency to maximize performance in semantic elaboration tasks.

6 Discussion

6.1 Academic and Practical Implications

This study underscores the importance of integrating semantically rich encodings into AI models to enhance the understanding of building semantics. From an academic perspective, the approach broadens current research by demonstrating that LLM embeddings—when compacted appropriately—can outperform conventional one-hot methods, thereby offering a more nuanced way to represent complex object subtypes. Practitioners, on the other hand, gain a practical encoding method that leverages widely available pretrained LLMs for improved decision support. By preserving finer distinctions among building semantics, AI models can more accurately comprehend real-world project information, ultimately helping industry professionals make more context-aware informed decisions supported by AI models.

6.2 Limitations

Despite its promising outcomes, this study has several limitations that call for cautious interpretation. First, the experiments limitedly covered a subtype classification task involving 42 subtypes across five high-rise residential BIM models. These may not comprehensively reflect the full range of building types or the myriad semantic tasks encountered in real-world practice. Additional work is needed to validate the proposed method on more diverse building typologies (e.g., hospitals, commercial complexes, infrastructure) and under different project conditions (e.g., varying project phases, levels of detail). Second, while the focus here was on subtype classification using a graph neural network, building semantics encompasses a broad range of tasks—such as construction scheduling, cost estimation, and design intent capture—that might benefit differently from LLM encodings. Lastly, the reliance on existing LLMs means that domain-specific terms not extensively represented in their training corpora may still suffer from suboptimal embeddings, highlighting the need for either fine-tuning or training domain-specific LLMs on specialized AECO datasets.

6.3 Future Directions

Building on the insights gained in this study, several avenues for future work emerge. First, exploring alternative loss functions—beyond cosine similarity—may further refine model training for specific semantic tasks. Second, investigating different dimensionality-reduction techniques could help preserve even more nuanced contextual information while mitigating the computational overhead of high-dimensional embeddings. Third, training or fine-tuning LLMs on construction-centric corpora, such as project specifications or technical standards, may yield embeddings that better capture domain-specific language and nuances. Finally, a more holistic approach that combines text-based encodings with additional data modalities—such as 2D/3D images, point clouds, or sensor data—could offer a deeper, multimodal understanding of building semantics. By expanding the scope of tasks and data types, future research can further validate the robustness and applicability of LLM encodings in increasingly complex AECO environments.

7 Conclusion

This paper tackled the challenge of preserving and enriching building semantics in AI models by introducing a training method that employs LLM embeddings as encodings. Through extensive experiments on BIM object subtype classification, we demonstrated that compacted LLM encodings, such as "llama-3 (compacted)," outperform conventional one-hot encoding, achieving higher weighted average F1-scores (0.8766 over 0.8575) and showing stronger potential to represent subtle semantic distinctions. Although statistical significance was not uniformly observed for all embeddings, the results highlight the growing promise of leveraging advanced language models to capture the domain-specific context vital to the AECO industry.

By integrating LLM-based encodings into class labeling for supervised AI model training, this study contributes a viable solution for domain practitioners and researchers seeking to improve model accuracy and semantic fidelity. As LLM technology continues to advance—and as construction datasets become richer—opportunities to refine these methods will likely multiply. Ultimately, the framework presented here offers a stepping stone toward more intelligent, context-aware AI systems capable of handling the increasing complexity and specialization inherent in modern building projects.

Acknowledgements

This work is supported in 2025 by the Korea Agency for Infrastructure Technology Advancement (KAIA)

grant funded by the Ministry of Land, Infrastructure and Transport (Grant RS-2021-KA163269 and RS-2024-00407028) and the Hans Fischer Senior Fellowship program at the Technical University of Munich - Institute for Advanced Studies (TUM-IAS) in Germany.

References

[1] S. Jang, H. Roh, G. Lee, Generative AI in architectural design: Application, data, and evaluation methods, Automation in Construction 174 (2025) 106174. https://doi.org/10.1016/j.autcon.2025.106174.

[2] D. Gil, G. Lee, Zero-shot monitoring of construction workers' personal protective equipment based on image captioning, *Automation in Construction*, 164:105470, 2024.

[3] W.-C. Chern, J. Hyeon, T.V. Nguyen, V.K. Asari, H. Kim, Context-aware safety assessment system for far-field monitoring, *Automation in Construction* 149:104779, 2023.

[4] J. Ma, J. Jung, F. Leite, Deep Learning–Based Scan-to-BIM Automation and Object Scope Expansion Using a Low-Cost 3D Scan Data, *Journal of Computing in Civil Engineering*, 38:04024040, 2024.

[5] J. Kim, J. Kim, N. Koo, H. Kim, Automating scaffold safety inspections using semantic analysis of 3D point clouds, *Automation in Construction*, 166:105603, 2024.

[6] R. Sacks, L. Ma, R. Yosef, A. Borrmann, S. Daum, U. Kattel, Semantic Enrichment for Building Information Modeling: Procedure for Compiling Inference Rules and Operators for Complex Geometry, *Journal of Computing in Civil Engineering*, 31:04017062, 2017.

[7] J.W. Ma, T. Czerniawski, F. Leite, Semantic segmentation of point clouds of building interiors with deep learning: Augmenting training datasets with synthetic BIM-based point clouds, *Automation in Construction*, 113:103144, 2020.

[8] Z. Wang, R. Sacks, B. Ouyang, H. Ying, A. Borrmann, A Framework for Generic Semantic Enrichment of BIM Models, *Journal of Computing in Civil Engineering*, 38:04023038, 2024.

[9] G. Austern, T. Bloch, Y. Abulafia, Incorporating Context into BIM-Derived Data—Leveraging Graph Neural Networks for Building Element Classification, *Buildings*, 14:527, 2024.

[10] S. Jang, G. Lee, M. Park, J. Lee, S. Suh, B. Koo, Semantic elaboration of low-LOD BIMs: Inferring functional requirements using graph neural networks, *Advanced Engineering Informatics*, 64: 103100, 2025.

[11] S. Jang, G. Lee, Interactive Design by Integrating a Large Pre-Trained Language Model and Building Information Modeling, In *Proceedings of Computing in Civil Engineering 2023*, pages 291-299, Corvallis, Oregon, United States, 2023.

[12] S. Jang, G. Lee, J. Oh, J. Lee, B. Koo, Automated detailing of exterior walls using NADIA: Natural-language-based architectural detailing through interaction with AI, *Advanced Engineering Informatics*, 61:102532, 2024.

[13] M. Uhm, J. Kim, S. Ahn, H. Jeong, H. Kim, Effectiveness of retrieval augmented generation-based large language models for generating construction safety information, *Automation in Construction*, 170:105926, 2025.

[14] Y. Gao, G. Xiong, H. Li, J. Richards, Exploring bridge maintenance knowledge graph by leveraging GrapshSAGE and text encoding, *Automation in Construction*, 166:105634, 2024.

[15] J. Lee, S. Ahn, D. Kim, D. Kim, Performance comparison of retrieval-augmented generation and fine-tuned large language models for construction safety management knowledge retrieval, *Automation in Construction*, 168:105846, 2024.

[16] G. Lee, S. Jang, S. Suh, M. Park, H. Roh, L. Kim, Three approaches for building information modeling library transplant, in: Seoul, Korea, 2023.

[17] A. Neelakantan, T. Xu, R. Puri, A. Radford, J.M. Han, J. Tworek, Q. Yuan, N. Tezak, J.W. Kim, C. Hallacy, J. Heidecke, P. Shyam, B. Power, T.E. Nekoul, G. Sastry, G. Krueger, D. Schnurr, F.P. Such, K. Hsu, M. Thompson, T. Khan, T. Sherbakov, J. Jang, P. Welinder, L. Weng, Text and Code Embeddings by Contrastive Pre-Training, On-line: https://doi.org/10.48550/arXiv.2201.10005, Accessed: 12/23/2024.

[18] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, A. Rodriguez, A. Joulin, E. Grave, G. Lample, LLaMA: Open and Efficient Foundation Language Models, On-line: https://doi.org/10.48550/arXiv.2302.13971, Accessed: 12/23/2024.

[19] C.K. Ogden, I.A. Richards, *The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism*, Harcourt, Brace, 1927.

[20] A. Kusupati, G. Bhatt, A. Rege, M. Wallingford, A. Sinha, V. Ramanujan, W. Howard-Snyder, K. Chen, S. Kakade, P. Jain, Matryoshka representation learning, *Advances in Neural Information Processing Systems*, 35:30233–30249, 2022.

[21] W. Hamilton, Z. Ying, J. Leskovec, Inductive representation learning on large graphs, Advances in Neural Information Processing Systems, 30, 2017.

Feature Sensitivity Analysis for Enhanced HVAC Fault Detection and Diagnosis

Arash Hosseini Gourabpasi [1], Farzad Jalaei [1], Mazdak Nik-Bakht [2]

[1]National Research Council of Canada, Ottawa, Canada
[2]Concordia University, Montreal, Canada
Arash.HosseiniGourabpasi@nrc-cnrc.gc.ca, Farzad.Jalaei@nrc-cnrc.gc.ca, mazdak.nikbakht@concordia.ca

Abstract – Automated Fault Detection and Diagnostics (AFDD) is a data-driven approach that enables the timely detection of faults, their types, and severity in Heating, Ventilation, and Air Conditioning (HVAC) systems. Machine learning models can be used to develop and implement AFDD models for HVAC at the system, sub-system, and equipment levels. However, there is a discrepancy between experimental facilities and actual buildings in practice.

This study implements sensitivity analysis to determine how variations in the features used in the case study under investigation affect the performance of the AFDD models. The analysis includes techniques such as correlation analysis and machine learning models (Support Vector Machine (SVM) and Artificial Neural Network (ANN)) to assess the sensitivity of the models to changes in input features. The results depict how sensitive the AFDD models are to different features and the extent to which variations in these features can impact the models' performance. These findings can facilitate the selection of robust features for machine learning-based Fault Detection and Diagnostics (FDD) models of HVAC in buildings by building operators, including facility managers, asset managers, and owners.

Keywords –

Sensitivity analysis, Feature impact analysis, Data Mining, Fault detection and diagnostics, Machine Learning, Artificial Neural Network, Support Vector Machine, HVAC

1 Introduction

Automated Fault Detection and Diagnostics (AFDD) is crucial for maintaining the efficiency and reliability of HVAC systems [1], [2]. AFDD models are constructed using data that reflects the behavior and performance of facilities and HVAC systems, enabling the prediction of fault occurrences and types. In this regard, machine learning models offer a promising approach to FDD, but there is a significant gap between the controlled environments of experimental facilities and the real-world conditions of actual buildings. This study aims to bridge this gap by performing a sensitivity analysis to understand how variations in features affect the performance of AFDD models [3] .

Presently testing facilities used for AFDD development are often using controlled environments to reduce uncertainty. They are heavily instrumented, which means that test facilities contain many sensors installed in different building spaces and HVAC equipment to detect subtle changes in the system and building space environment. However, in practice, the existing stock of commissioned buildings does not possess the same sensor variety and is limited in numbers compared to experimental labs. The current differences in real case scenarios and research facilities do not make the machine learning-based AFDD models feasible for utilization in actual buildings [4].

This study implements sensitivity analysis to determine how variations in the features used in the case study under investigation affect the performance of the AFDD models. The analysis includes techniques such as correlation analysis and data mining techniques to assess the sensitivity of the models to changes in input features. The study investigates how sensitive the AFDD models are to different features and the extent to which variations in these features can impact the model's performance. In addition, the results of feature importance can be used to facilitate the selection of robust features for machine learning-based FDD models of HVAC in buildings.

2 Methodology

This study utilizes data from a Building Information Model (BIM), which contains building spatial data, and Building Management System (BMS), which includes time-series data representing HVAC, to create a comprehensive dataset for fault detection and diagnostics of an HVAC system. The BIM model includes static features such as room area, window area, door area, opening area, distance to HVAC from VAV (Variable

Air Volume), room adjacency to shafts, and whether rooms are exterior or interior. These features collectively form 80 static features representing the facility's contextual information. The BMS dataset includes 68 dynamic features such as VAV temperature, room temperature, humidity, status, supply and return temperature, metered data, set points, and flow rates.

A key aspect of our methodology is feature engineering, where we generate new features from existing data to enhance the accuracy and effectiveness of machine learning models. For instance, we created a calculated feature called "space air conditioning," which combines contextual information from BIM and air conditioning conditions from BMS for each room. This feature is crucial for identifying faults related to occupancy patterns and other contextual factors.

Feature engineering in AFDD involves generating new features from existing data to improve the accuracy and effectiveness of machine learning models. This is particularly important when dealing with limited sensor data, a common problem in real-world building environments. For example, one can combine time-series data from BMS (e.g., supply air temperature) with spatial data from BIM (e.g., room occupancy) to create a new feature that represents the air conditioning state of a space. This type of feature can help to identify faults that are related to occupancy patterns or other contextual factors.

To make the features workable, two grouping systems were formed: one based on system-related features or zone/space features, and the second based on the feature source.

Table 1: Grouping based on System related features or Zone/Space related features

Grouping based on level	No of groups	Feature groups formed based on simillarity names
Zone/Space level	4	VAV temp, Room temp, humidity, set point
System level	3	Circuit 1 and 2, supply and return, Flow rate
Mix	1	Status
Building	1	Meter

Table 2: Grouping based on feature source and type (Static/Live)

Grouping based on source and type	No of Features	Full form
BMS	68	BMS dataset
BIM	80	BMS + Live BIM dataset
Dy+St	160	BMS + Live BIM dataset + Static BIM dataset
St BIM	148	BMS + Static BIM dataset

Two approaches were used the first being feature impact analysis and second sensitivity analysis. Feature impact analysis evaluates the importance of groups of features, while sensitivity analysis assesses the impact of individual features on model performance. Sensitivity analysis helps to identify the most important features for fault detection and diagnosis, as well as to understand how the model's performance is affected by missing or incomplete data. One way to perform sensitivity analysis is to train multiple models with different combinations of features and compare their accuracy. For example, one might train a model with all available features, then train another model with a subset of features, and compare their performance.

Feature impact analysis is useful when the set of features is large, as in the case of the facility, which will enable the modelers to identify which features are suitable for the machine learning algorithm to identify the faults of the HVAC systems.

3 Model Development

This study utilizes data from the LBNL (Lawrence Berkeley National Laboratory) Automated Fault Detection for Buildings Data [5], [6], [7], [8], [9], [10], [11]. A Building Information Model (BIM) is developed based on the specifications, and further, the Building Management System (BMS) is used to create a comprehensive dataset for fault detection and diagnostics of an HVAC system [12], [13], [14], [15]. The BIM model of the facility is created using the documentation of the test facility. The dataset generated for fault detection and diagnostics of the HVAC system includes a rooftop unit (RTU) connected to variable air volume (VAV) systems.

The BIM and BMS data were combined to form a comprehensive dataset [4]. The BIM features were grouped into categories such as distance to HVAC from VAV, room area, window area, door area, opening area, room adjacency to shafts, and whether rooms are exterior or interior. The BMS features were grouped into categories such as VAV temperature, room temperature, humidity, status, supply and return temperature, metered data, set points, and flow rates.

In this research, various machine learning algorithms are modeled to identify the most impactful features for fault detection and diagnostics. For feature importance analysis, classifiers like XGBoost, Random Forest, and Decision Tree were used. The machine learning algorithms specifically employed for fault detection and diagnostics (FDD) were Support Vector Machine (SVM) and Artificial Neural Network (ANN). They were selected based on the findings of earlier research, which

highlighted their adoption for system-level and subsystem-level faults [16]. The performance of each algorithm was evaluated based on its ability to identify impactful features and detect faults. The faults considered in this study included condenser fouling, HVAC setback errors (delayed onset, early termination), excessive infiltration, lighting setback errors (delayed onset, early termination), no overnight HVAC setback, and thermostat measurement bias.

For feature importance evaluation, several models were developed. In the case of the Random Forest classifier, feature importance was computed as the mean and standard deviation of the impurity decrease within each tree, also known as Gini importance. For the Decision Tree classifier, feature importance was determined by the normalized total reduction of the criterion brought by that feature. XGBoost classification was based on gain, which measures the improvement in accuracy brought by a feature to the branches it is on.

Table 3: Machine Learning algorithms used as classifier for feature importance analysis

XGBoost Classification	Random Forest Classification	Decision tree classification
Condenser Outlet Pressure (Circ 1)	Condenser Outlet Pressure (Circ 1)	Condenser Outlet Pressure (Circ 1)
VAV Reheat Status	Room 205 Air Humidity	Discharge Temperature (Circ 2)
Discharge Temperature (Circ 2)	Room 206 Air Humidity	Room 204 Air Humidity
Room 203 Air Humidity	Room 204 Air Humidity	Room 205 Air Humidity
Room 203 Air Temperature	Room 106 Air Humidity	Room 206 Air Humidity
Discharge Pressure (Circ 1)	Suction Temperature (Circ 2)	Room 202 Air Humidity
X Air Humidity	Room 203 Air Humidity	Room 102 Air temperature
X Air Humidity	Room 102 Air Temperature	Room 106 Air Humidity
Suction Temperature (Circ 2)	Room 103 Air Humidity	Lighting system
X Air Humidity	Discharge temperature (Circ 2)	Room 205 Air Temperature

The calculated air conditioning feature shows a negative correlation and is the only feature that is found to have an impact when brought in from BIM, as shown below.

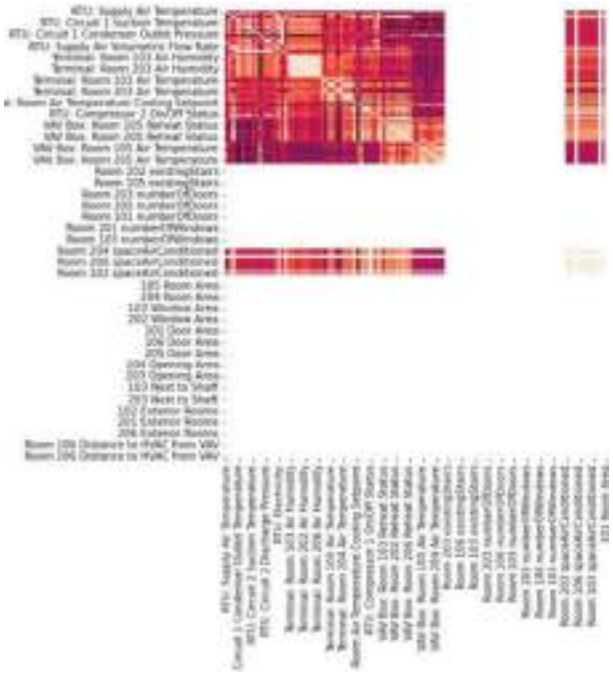


Figure 1: Excerpt of correlation analysis and of important features for AFDD in the curated dataset including BIM features

4 Results and Discussion

The result of this research indicated that the Random Forest classifier was most suitable for the dataset, particularly when it came to calculated features. For sensory data features, the classification algorithms showed similar feature types, with humidity being the most impactful feature.

The comparison of SVM and ANN indicates that the ANN is more sensitive to selected features, showing higher levels of sensitivity. The maximum impact that a feature can have on fault detection is 67% for both the suction temperature feature and the air-cooling temperature. Additionally, the diagnostics models show that the air temperature cooling setpoint can have a 62% impact, followed by the discharge temperature with a 37% impact.

The top 5 groups of features identified are then checked to see what percentage of accuracy impact these features have on the dataset for each of the algorithms for both fault detection and diagnostics.

Table 4: Impact of top 5 features on Fault detection and diagnosis using SVM Machine Learning Algorithms

Top 5 features	SVM Accuracy for Fault Detection	SVM Accuracy for Fault Diagnostics
Air Humidity	80%	71%

Discharge temperature (Circuit 2)	67%	38%
Suction Temperature (Circuit 2)	67%	38%
Room Temperature	79%	71%
Cooling setpoint	67%	33%

Table 5: Impact of top 5 features on Fault detection and diagnosis using ANN Machine Learning Algorithms

Top 5 features	ANN Accuracy for Fault Detection	ANN Accuracy for Fault Diagnostics
Air Humidity	87	55
Discharge temperature (Circuit 2)	71	37
Suction Temperature (Circuit 2)	67	39
Room Temperature	85	41
Cooling setpoint	65	33

The analysis illustrate that the random forest classifier is most suitable for the dataset when it comes to calculated features. For sensory data features, classification algorithms show similar feature types, with humidity being the most impactful feature. The top features identified include air humidity, discharge temperature, suction temperature, room temperature, and cooling setpoint.

The calculated feature, "space air conditioning," is the only feature in the BIM dataset that shows a negative correlation and is present in the important feature list when analysis is performed using a random forest classifier, which has a weak correlation of an average of 25%.

The results highlight the importance of certain features in the performance of AFDD models. For instance, air humidity and room temperature consistently show high impact across different models. The findings suggest that incorporating dynamic BIM features can enhance the sensitivity of AFDD models, especially when certain BMS features are unavailable. This has practical implications for facility managers and asset owners in selecting robust features for fault detection and diagnostics.

This study demonstrates the significance of feature impact analysis in developing effective AFDD models for HVAC systems. By identifying the most impactful features, facility managers can improve the accuracy and reliability of fault detection and diagnostics. Future work should explore the integration of additional contextual features and the application of these findings to a broader range of building types.

Acknowledgements

The authors acknowledge the financial support of the National Research Council of Canada (NRC)—Construction Sector Digitalization and Productivity Challenge program (CSDP) under Agreement # CSDP-002-1: Digital Twins for Smart Decarbonization of the Built Environment Meeting Circular Economy Criteria.

References

[1] Z. Chen *et al.*, "A review of data-driven fault detection and diagnostics for building HVAC systems," *Applied Energy*, vol. 339, p. 121030, Jun. 2023, doi: 10.1016/j.apenergy.2023.121030.

[2] K. Heimar Andersen, S. Pommerencke Melgaard, H. Johra, A. Marszal-Pomianowska, R. Lund Jensen, and P. Kvols Heiselberg, "Barriers and drivers for implementation of automatic fault detection and diagnosis in buildings and HVAC systems: An outlook from industry experts," *Energy and Buildings*, vol. 303, p. 113801, Jan. 2024, doi: 10.1016/j.enbuild.2023.113801.

[3] A. Hosseini Gourabpasi and M. Nik-Bakht, "An ontology for automated fault detection & diagnostics of HVAC using BIM and machine learning concepts," *Science and Technology for the Built Environment*, vol. 30, no. 8, pp. 972–988, Sep. 2024, doi: 10.1080/23744731.2024.2363104.

[4] A. Hosseini Gourabpasi and M. Nik-Bakht, "BIM-based automated fault detection and diagnostics of HVAC systems in commercial buildings," *Journal of Building Engineering*, vol. 87, p. 109022, Jun. 2024, doi: 10.1016/j.jobe.2024.109022.

[5] D. Goldwasser, B. Ball, A. Farthing, S. Frank, and P. Im, "Advances in Calibration of Building Energy Models to Time Series Data: Preprint," *Renewable Energy*, p. 11, 2018.

[6] J. Granderson, G. Lin, A. Harding, P. Im, and Y. Chen, "Building fault detection data to aid diagnostic algorithm creation and performance testing," *Sci Data*, vol. 7, no. 1, Art. no. 1, Dec. 2020, doi: 10.1038/s41597-020-0398-6.

[7] J. Lee, P. Im, J. D. Munk, M. Malhotra, M. Kim, and Y. Song, "Comparison Evaluations of VRF and RTU Systems Performance on Flexible Research Platform," *Advances in Civil*

Engineering, vol. 2018, pp. 1–16, 2018, doi: 10.1155/2018/7867128.

[8] A. Casillas, G. Lin, and J. Granderson, “Curation of Ground-Truth Validated Benchmarking Datasets for Fault Detection & Diagnostics Tools,” 2020, doi: 10.20357/B7NG6Z.

[9] P. Im, J. R. New, and J. Joe, “Empirical Validation of Building Energy Modeling using Flexible Research Platform,” presented at the Building Simulation 2019, Rome, Italy, Jun. 2022, pp. 4515–4521. doi: 10.26868/25222708.2019.210263.

[10] P. Im, J. D. Munk, and A. C. Gehl, “Evaluation of Variable Refrigerant Flow Systems Performance and the Enhanced Control Algorithm on Oak Ridge National Laboratory s Flexible Research Platform,” ORNL/TM--2015/225, 1186004, Jun. 2015. doi: 10.2172/1186004.

[11] Y. Yoon, Y. Li, P. Im, and Y. Bae, “Implementation of ASHRAE Guideline 36 Control Logic into Oak Ridge National Laboratory (ORNL)’s Flexible Research Platform (FRP),” ORNL/TM-2022/2434, 1871093, Apr. 2022. doi: 10.2172/1871093.

[12] J. Granderson and G. Lin, “Inventory of Data Sets for AFDD Evaluation,” p. 25.

[13] D. Kim, S. J. Cox, H. Cho, and P. Im, “Model calibration of a variable refrigerant flow system with a dedicated outdoor air system: A case study,” *Energy and Buildings*, vol. 158, pp. 884–896, Jan. 2018, doi: 10.1016/j.enbuild.2017.10.049.

[14] H. L. Buckberry and M. Bhandari, “ORNL MAXLAB Flexible Research Platforms,” p. 12.

[15] P. Im and M. Bhandari, “Use of Flexible Research Platforms (FRP) for BIM and Energy Modeling Research,” *ASHRAE Transactions*, p. 9.

[16] A. Hosseini Gourabpasi and M. Nik-Bakht, “Knowledge Discovery by Analyzing the State of the Art of Data-Driven Fault Detection and Diagnostics of Building HVAC,” *CivilEng*, vol. 2, no. 4, pp. 986–1008, Nov. 2021, doi: 10.3390/civileng2040053.

Comparing Few-Shot Learning with LLMs for Efficient Text Classification in Road Maintenance Applications

Varun Kumar Reja[1,*], Ching Yau (Fergus) Mok[1,*], Aritra Pal[1,2] and Ioannis Brilakis[1]

[1] University of Cambridge, UK
[2] Indian Institute of Technology Madras, India

varunreja7@gmail.com

Abstract -

Efficient road maintenance is vital for long-lasting and safe transportation networks, but traditional methods that rely on manual inspection are labour-intensive and error-prone. The integration of Natural Language Processing (NLP) and Large Language Models (LLMs) presents a transformative solution for automating text-based tasks in road maintenance. This study investigates the application of LLM-based text classification models to process unstructured textual data from road maintenance logs, with a focus on resource-constrained scenarios characterised by limited labelled datasets. Two primary approaches were evaluated: traditional fine-tuning and few-shot learning. Using a public dataset from New York City roadwork inspections containing 83 distinct classes, we performed extensive model comparisons. Pre-trained transformer models *Llama* and *BERT* were fine-tuned to achieve baseline performance. Additionally, a few-shot learning method, SetFit, was employed to address data scarcity through efficient task adaptation using minimal labelled examples. Results showed SetFit outperformed fine-tuning in low-resource scenarios, achieving high accuracy and F1 scores with as few as 1-10 examples per class, reducing annotation efforts. This highlights the potential of few-shot learning for real-world deployment. Future work will address scalability, multimodal data fusion, and integration into predictive maintenance.

Keywords -

Few-Shot Learning; Large Language Models; Road Maintenance; Text Classification; SetFit; Natural Language Processing; Transformer Models; Inspection Logs; Infrastructure Management; Low-Resource NLP

1 Introduction

Effective road maintenance is essential for the safety, reliability and longevity of transportation networks. However, increasing traffic volumes and ageing infrastructures present challenges, such as identifying deteriorating assets, ensuring timely intervention, and optimising resource allocation. Inspection logs provide valuable information on road maintenance. However, manually analysing an enormous volume of inspection reports is labour-intensive, time-consuming, and error-prone [1, 2]. These issues are further amplified in a resource-constrained setting where funding and manpower are limited [3].

The advent of artificial intelligence (AI) and natural language processing (NLP) offers the potential for automating road maintenance tasks like analysing inspection data, prioritising repairs, and generating insights [4]. Transformer-based models like BERT and GPT have become state-of-the-art for text processing [5], yet their application in road maintenance remains underexplored, especially in low-resource settings with limited labelled data.

This paper aims to bridge this gap by systematically evaluating the use of transformer-based text classifiers in road maintenance applications, particularly in resource-constrained settings. Specifically, the study focuses on:

1. **Comparative Analysis**: Evaluating the performance of traditional fine-tuning approaches across various pre-trained transformer model architectures and sizes, highlighting their effectiveness in processing road maintenance-related textual data.
2. **Few-Shot Learning**: Assessing the capability of dedicated few-shot models to deliver robust performance with minimal labelled data, catering to the constraints of limited resources often encountered.

The paper is structured as follows: Section 1 introduces the topic, followed by background in Section 2. Section 3 outlines the methodology, while Section 4 discusses the findings, limitations, and future research directions. Finally, Section 5 presents the conclusions.

2 Background

2.1 Challenges in Leveraging Text Data for Road Maintenance

Malihi et al. [6] emphasise the untapped potential of textual records such as inspection and maintenance reports in road maintenance, noting that their unstructured and inconsistent nature limits their utilisation. Unlike numerical or sensor data, text is inherently diverse in format, structure, and terminology. Inspection reports, for example, may describe similar issues in varied ways depending on

*Co-First Authors (Equally Contributed to this work).

the inspector's language and style. This lack of standardisation makes it difficult to extract consistent and actionable insights. Another challenge is the volume of data generated across extensive road networks. Maintenance operations often produce large quantities of text data over time, including daily logs, contractor reports, and communications. Manually sifting through this information to identify critical issues is time-consuming and prone to human error. The ambiguity and context-dependence of language further complicate the processing of text data. Terms like "minor crack" or "urgent repair" may vary in meaning depending on the context, location, or maintenance standard, leading to inconsistent interpretation. Similarly, implicit details, such as the severity of a reported defect, may not be readily apparent without supplementary information [5].

Additionally, there is a lack of annotated datasets specific to road maintenance applications. Developing effective machine learning models often requires large, labelled datasets to train algorithms. In road maintenance, creating such datasets is challenging due to the domain-specific nature of the text and the expertise required to accurately label data. This echoes [7], which raised concerns about the lack of annotated and high-quality data in the Architecture, Engineering, and Construction (AEC) domain.

These challenges underscore the need for innovative methods to efficiently classify, interpret, and utilise textual data in road maintenance.

2.2 NLP as an Opportunity in Road Maintenance

NLP, a field that enabling machines to comprehend natural language, is transformative for extracting insights from unstructured maintenance data like logs and inspection reports. Machine learning models can identify patterns in defect descriptions and prioritise tasks based on urgency. Advanced transformer-based NLP architectures like LLMs excel in handling language's contextual nuances. Khalil et al. [4] highlight the significant advancements of LLMs, such as GPT-3, in reasoning tasks, further enhanced by techniques like "Chain of Thought" prompting. In construction, [8] designed an automated system for construction specification review using NLP techniques such as named entity recognition (NER) with Bi-LSTM and document embedding through Doc2Vec to streamline processes and improve construction risk management. Recognising the difficulty in many AEC applications in collecting sufficient training data, [9] proposed a few-shot NER model for information extraction from bridge inspection texts.

2.3 LLM-Based Text Classification Approaches

LLMs have extensive capabilities in classifying texts due to their ability to accurately encode sentences [10]. They use the transformer architecture [11], a neural network leveraging the attention mechanism [12] to model dependency, to process long inputs at once without loss of information as is the case with recursion in approaches such as Recurrent Neural Network. LLMs utilise pre-trained models that possess task-agnostic knowledge through training on vast amounts of general text, akin to reading a million books. To adapt LLMs for downstream tasks such as text classification, architecture-specific fine-tuning using a task-specific dataset is performed. The following sections elaborate on different architectures and their associated fine-tuning techniques.

2.3.1 Traditionally Fine-tuned Models

The most common LLM-based text classifier, shown in Figure 1, appends a small task-specific classification head to produce class probabilities or logits. It typically consists of an uninitialised fully connected layer with neurons equal to the number of classes. Pre-trained models vary in size: smaller models (100M parameters), often termed Pre-trained Language Models (e.g., *BERT*), and larger models (7B+ parameters), such as *Llama*. For instance, BERT was pre-trained on 800M words from BooksCorpus and 2,500M words from English Wikipedia [13]. In this work, both 100M and 7B models are referred to as LLMs. Fine-tuning adapts the classifier to a specific task by updating pre-trained model parameters and initialising the classification head in a single step, using a cross-entropy loss function on model output logits.

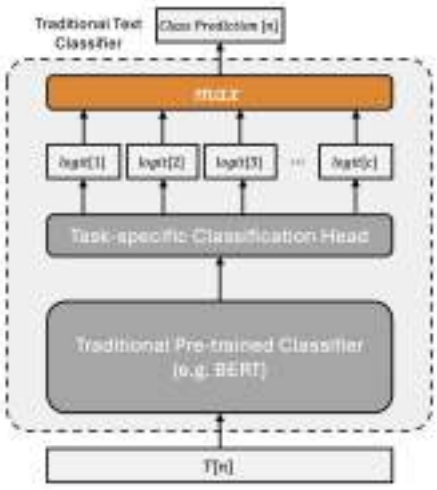


Figure 1. Architecture of a traditional transformer-based text classifier.

2.3.2 Few-shot Learning - SetFit

Few-shot learning refers to adapting models to a task using only a small number of examples (or shots). This is particularly useful for resource-constrained applications where large-scale annotated data is unavailable. A popular transformer-based few-shot learning technique is Sentence

Transformer Fine-Tuning (SetFit) [14].The classifier embeds text using Sentence-BERT (SBERT), a BERT variant pre-trained on pairwise similarity tasks, and adds a lightweight classification head to generate class scores. This setup produces semantically rich embeddings suited for classification.

SetFit is fine-tuned in a two-step process. First, the SBERT encoder is fine-tuned to produce accurate embeddings through contrastive learning by providing a pair of texts and training the model to determine whether they belong to the same class. This is followed by training the entire model (SBERT plus the classification head) to perform the classification task through cross-entropy. The contrastive learning adopted in the embedding training step is what allows SetFit to be data efficient in that it can train the model to explicitly separate the embeddings of texts from different classes.

The literature highlights the potential of NLP and LLMs in extracting insights from road maintenance texts, yet challenges remain due to unstructured data, inconsistency, and limited annotated datasets. While LLMs and few-shot learning methods like SetFit show promise, their application to domain-specific classification and automated maintenance prioritization is underexplored. This study addresses these gaps by evaluating LLM-based classification approaches for structured information extraction.

3 Research Methodology

3.1 Dataset

A public dataset on roadwork inspection comments was used to validate the proposed approach. It is the "Street Construction Inspections and Corrective Action Requests" from the New York City (NYC) Open Data [15], which is a collection of inspection comments on roadwork performed by non-governmental third parties that record instances of road code violations. Examples of inspection comments and their associated violated road code are listed in Table 1. From the dataset, the free text inspection comments come from the "DetailsofViolation" column and the associated violated road codes from "NOVCodeDescription".

The dataset underwent a series of cleaning processes and random subsampling to create few-shot datasets that can simulate data constraints. Figure 2 shows the steps taken to generate the datasets used in this work.

The raw dataset was first cleaned, starting by inspecting for irregularities. It was found that there were road codes that contained duplicate meanings that essentially refer to the same sets of violations. Table 2 outlines such examples that were manually merged. Following this, codes with fewer than 100 examples were excluded to ensure sufficient testing examples for representative conclusions to be drawn. This resulting dataset has a total of 400,000

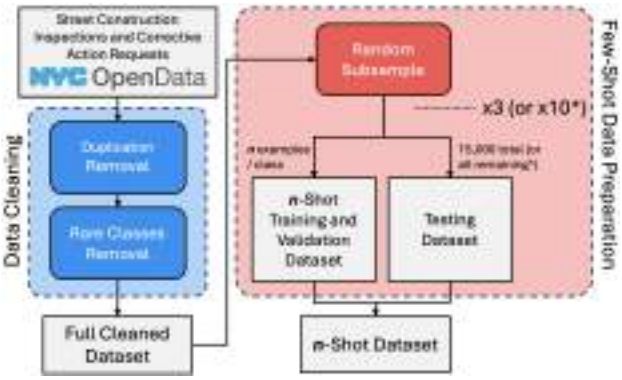


Figure 2. Illustration of the data preparation process for an *n*-shot dataset. (*further experiments with smaller models were performed using more datasets per shot and more testing examples per dataset.)

texts that belong to 83 classes.

The full, cleaned dataset was then randomly subsampled to form few-shot datasets. *Training datasets* were generated by subsampling 1, 2, 3, 4, 5, and 10 examples per class (shots), each shot was subsampled from all available data three times to create three statistically independent draws which helps mitigate the instability that can occur with small training datasets [14]. Each *Training datasets* is coupled with a corresponding *Validation dataset* and *Testing dataset*. *Validation datasets* were subsampled for roughly 1/3 of the shot to maintain the ratio of training to validation data common in standard machine learning datasets (minimum of 1 example per class), while the remaining unsampled data were used as a potential testing pool. For *Testing datasets*, 15,000 examples were subsampled (similar order of magnitude to relevant works [8]) to reduce computational costs, particularly for models with 7 billion parameters.

Note that to assess the impact of these subsampling operations, a further set of experiments were conducted with only smaller 100 million parameters models: First, few-shot scenarios used the full remaining unsampled testing pool (as opposed to 15,000) and ten individual datasets per shot (as opposed to three). Second, an all-data training scenario was introduced using the full dataset with an 8/1/1 split for the training, validation, and testing sets.

3.2 Simulated Task

The task simulated in this case study involves training a transformer-based text classifier using limited data to classify roadwork inspection comments into the violated road code. Each inspection comment records the details of an instance of a code violation, which in this dataset, can belong to one of 83 road codes. The simulated task highlights two key challenges: limited training examples and a large number of classes, reflecting a complex technical text classification scenario.

Table 1. Examples of pairs of free text comments describing instances of code violation (textual inputs) and static code descriptions (class descriptions)

Free Text Inspection Comments	Associated Road Code Violated
A/T/P/O I observed that the respondent has failed to remove the ramping material from the roadway after the completion of the final restoration or prior to the expiration of the permit as per subsection (vii). Respondent was notified to fix this condition via CAR [#20184520373] issued on 12/31/18.	Failure to remove plating and decking after final restoration and prior to DOT permit expiration.
Respondent failed to display required signs at work site with permit number,work type, start date ,completion date.No one on site at time of inspection. If not admitting the charge, you MUST APPEAR IN PERSON.	Identifying signs improperly displayed or missing
A/T/P/O I observed respondent identified by "BIC 3179" markings on Commercial Refuse Container, having a container in the roadway without any visible reflective markings capable of producing illumination or warning glow when struck by vehicle headlamps. No one onsite.	Commercial refuse container without proper reflective markings on all four sides

Table 2. Examples of road codes with duplicating definitions.

Code 1	Code 2	Reason for duplication
Commercial refuse container/debris obstructing sidewalks, gutters, crosswalks or driveway	Commercial refuse container/debris obstructing sidewalks, gutters, crosswalks, bike lanes or driveway	Semantically identical: Code 1 is strictly contained within code 2
Failure to post âSteel Plates Aheadâ or âRaise Plowâ sign; failure to countersink plates flush with roadway	Failure to post “Steel Plates Ahead” or “Raise Plow” sign; failure to countersink plates flush with roadway	Character encoding error: The “â” is a result of a character encoding error originating from the publisher

Limited data refers to training models with a small number of violation examples per road code (number of shots n). Various ns were experimented with to simulate different levels of data constraints. 10 was selected as the maximum, considering that it would amount to 830 total examples with 83 classes, which can already be difficult to replicate in many resource-limited real-world engineering applications. This will provide insight into how well different models perform under different data availabilities.

A unique characteristic of the employed dataset is the large number of classes. Using natural language to infer one of 83 road codes that are being violated can be challenging due to the complexity of the classification task. However, this is a common characteristic in many road maintenance and AEC applications that are tightly regulated and stringent. For example, the dataset used records instances of road code violations which correspond to regulations on roadwork, having an impact on both safety for road users and disruptions caused.

3.3 Models

The two types of transformer-based models outlined in Section 2.3 were experimented with. Traditionally, fine-tuned models represent standard transformer-based text classifiers, which typically require a larger amount of training data to be performant, while SetFit is an example of a technique that specialises in tasks where there is a limited number of training data, like road maintenance.

Traditionally Fine-tuned For traditionally fine-tuned models that were explored in Section 2.3.1, different pre-trained models were experimented with, which have different numbers of parameters. They mostly belong to two groups: those with roughly 100 million parameters and 7 billion parameters. Table 3 lists the pre-trained models used as traditionally fine-tuned text classifiers along with hyperparameters that affect model weights updates and convergence. The 100 million parameters model was the three best-performing models of this size in a comparison study [16], and they can be run on a wider range of hardware (tested with an M1 MacBook Pro with 32GB of unified memory). The 7 billion models are also widely quoted in literature [17][18] and exhibit high performance at the expense of requiring much more computational power, including GPU and VRAM.

Table 3. Pre-trained models used and their associated training hyperparameters (search range in [ ])

Model	Hyperparameters		
	No. params	Learning rate	Batch size
bert-base-uncased	110M	4e-5$_{[1e-6,1e-4]}$	16$_{[4,32]}$
roberta-base	125M	2e-5$_{[1e-6,1e-4]}$	16$_{[4,32]}$
xlnet-base-cased	110M	2e-5$_{[1e-6,1e-4]}$	16$_{[4,32]}$
Llama-2-7b-hf	6.74B	2e-4$_{[1e-3,1e-5]}$	3
Meta-Llama-3-8B	8.03B	2e-4$_{[1e-3,1e-5]}$	3
Mistral-7B-v0.1	7.24B	2e-4$_{[1e-3,1e-5]}$	3

Few-shot Additionally, SetFit, a transformer-based few-shot technique discussed in Section 2.3.2, was also experimented with to evaluate performance improvements in a low-resource workflow brought by dedicated few-shot techniques. The experimental design was modeled after the methodology described in the study by [14], with the 300 million parameter *paraphrase-multilingual-mpnet-base-v2* used as the pre-trained SBERT model and hyperparameters affecting model weights updates and convergence for the two training phases [14] listed in Table 4. SetFit has not been focused on in works relating to classifying AEC texts such as road inspection comments, and

its performance can be validated against more traditional techniques in various data-constrained settings.

Table 4. SetFit training hyperparameters for the two fine-tuning stages [14]. The first value corresponds to the SBERT encoder, and the second to the classification head. During the *Embedding* phase, no classification head is involved (N/A).

Hyperparameters	Fine-tuning	
	Embedding	Classifier
Learning rate	2e-5, N/A	1e-5, 1e-2
Batch size	16, N/A	16, 2
No. epochs	1, N/A	1, 16

3.4 Implementation Details

The models in this study were trained using the Transformers and SetFit libraries on an RTX 4080 (16GB VRAM). Hyperparameters were selected on validation set performance. Training lasted up to 20 epochs with the best model being chosen automatically on validation results.

3.5 Metrics

Accuracy and $f1$ score are the reported metrics in this study and are defined by the following equations:

$$accuracy = \frac{tp + tn}{tp + tn + fp + fn} \quad (1)$$

$$f1 = 2 \times \frac{precision \times recall}{precision + recall} \quad (2)$$

Accuracy values reported are micro-averaged, meaning that it assigns equal importance to each testing example. It is the overall number of correct classifications over all classifications. $f1$ values reported are macro-averaged, where equal importance is assigned to each class, meaning that performances in underrepresented classes in testing examples are taken into account in the overall metric values, which is useful in class imbalanced applications.

4 Results and Discussion

This section outlines the key results. Table 5 lists averaged *accuracy* and $f1$ values of all evaluated models for each few-shot scenario tested, including 100M and 7B traditionally fine-tuned models and SetFit.

4.1 SetFit Outperforms all Traditionally Fine-tuned Models

Classification performance is critical in road maintenance as errors can compromise safety and reliability. For instance, misclassifying "Identifying signs improperly displayed or missing" can lead to possible accidents due to the lack of warning for works. As a classifier specifically designed for resource-constrained settings, SetFit outperformed all traditionally fine-tuned and structured transformer-based classifiers in every few-shot scenario. In a 10-shot setting, SetFit achieved an $f1$ score of 0.539 and an accuracy of almost 70%, while the best traditionally fine-tuned model attained an $f1$ of 0.524 and an accuracy of 66%, respectively. This trend becomes more apparent as the number of shots is reduced. At a 1-shot setting, SetFit outperformed the best performing traditionally-fine-tuned model by more than 3 times in terms of $f1$ score, achieving 0.232 vs 0.074 with just 1 violation example per road code. Figure 3 plots the results of SetFit against three 7B models. Despite utilising 20 times more parameters, the 7B traditionally fine-tuned models were not able to outperform SetFit. This signifies that computationally efficient models can outperform state-of-the-art text classifiers in a resource-constraint domain-specific setting if they are effectively used, such as the case of SetFit. The contrastive learning stage of SetFit allows the text encoder to effectively distinguish between classes, making efficient use of model parameters. This demonstrates that the use of SetFit in a resource constrained setting in road maintenance is more suitable than traditionally fine-tuned techniques.

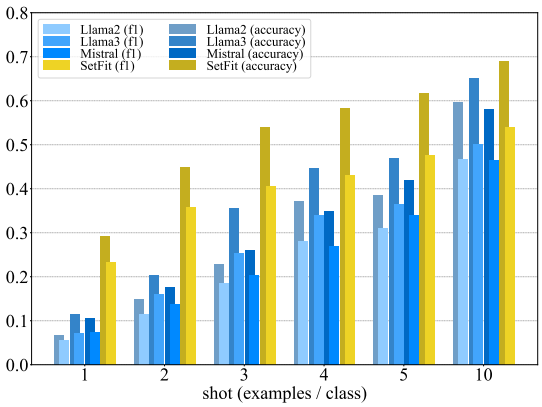


Figure 3. *Accuracy* and $f1$ scores for 7 billion parameter traditional models and SetFit.

4.2 Model Sizes Have Little Impact on Few-Shot Performance

Looking at the results for traditionally fine-tuned models with 100M and 7B parameters, the model's performances do not correlate with size. For instance, by looking at metric values in Table 5, the 100M *bert-base-uncased* model outperformed all 7B models in a 10-shot setting.

This observation indicates that increasing the number of parameters in traditionally fine-tuned models will likely not bridge the gap in performance with SetFit. This can be due to the use of SBERT in SetFit, where the pre-training objective focusing on semantic similarity can potentially better capture the semantics of the inspection comments than the objectives of other models, such as text generation.

Table 5. $F1$ and $accuracy$ results for all tested models. Testing results per shot were calculated with three independent subsampled datasets on a total of 15,000 testing examples.

Shots	Traditionally Fine-Tuned						Few-Shot
	100M			7B			
	BERT	RoBERTa	XLNet	Llama2	Llama3	Mistral	SetFit
	f1 scores						
1	0.031	0.053	0.049	0.056	0.072	0.074	0.232
2	0.107	0.143	0.097	0.114	0.160	0.137	0.357
3	0.195	0.183	0.143	0.186	0.254	0.203	0.405
4	0.288	0.265	0.146	0.281	0.339	0.269	0.430
5	0.372	0.303	0.231	0.311	0.365	0.340	0.475
10	0.524	0.452	0.432	0.467	0.501	0.464	0.539
	accuracy scores						
1	0.041	0.067	0.078	0.066	0.115	0.105	0.291
2	0.147	0.174	0.120	0.148	0.204	0.177	0.449
3	0.262	0.227	0.161	0.228	0.355	0.260	0.539
4	0.347	0.344	0.186	0.372	0.446	0.349	0.582
5	0.502	0.397	0.294	0.386	0.469	0.419	0.617
10	0.660	0.589	0.556	0.596	0.650	0.581	0.690

Table 6. Additional $accuracy$ and $f1$ results for all tested 100M models and SetFit. Testing results per shot were calculated with ten independent subsampled datasets on all available testing data.

Shots	Traditionally Fine-tuned						Few-Shot	
	BERT		RoBERTa		XLNet		SetFit	
	acc.	f1	acc.	f1	acc.	f1	acc.	f1
1	0.058	0.039	0.092	0.074	0.066	0.043	0.304	0.254
2	0.150	0.123	0.195	0.162	0.124	0.108	0.443	0.373
3	0.249	0.199	0.260	0.211	0.163	0.140	0.525	0.434
4	0.345	0.285	0.331	0.270	0.217	0.177	0.576	0.470
5	0.445	0.359	0.410	0.326	0.292	0.240	0.619	0.508
10	0.673	0.552	0.600	0.489	0.547	0.442	0.691	0.579
	Training over all available data							
all	0.954	0.899	0.951	0.890	0.952	0.893	N/A	N/A

4.3 Subsampling Testing Data Did Not Impact Observations

Due to the need to generate few-shot samples and to account for computational constraints, the above experiments were performed with subsampling steps. To verify the observations above and to assess the impact these subsampling operations have, further experimentation was performed by using both a higher number of independent datasets per shot (10 up from 3) and by eliminating subsampling for testing data to utilise all available testing examples (up from 15,000). Results for these experiments are listed in Table 6, which includes all 100M traditionally fine-tuned models and SetFit.

Figure 4 visualises the $accuracy$ and $f1$ values for this additional set of experiments. The same observation as the prior sections remains valid in that SetFit outperforms all other traditionally fine-tuned models. Despite having a general shift in metric values from having a different set of samples, the relative performance has remained consistent.



Figure 4. $Accuracy$ and $f1$ scores in different few-shot settings for 100 million parameter traditional models and SetFit.

4.4 Absolute Performance Low in Extremely Constrained Settings

It is important to note that the absolute performance of all models tested remains moderate in few-shot settings. This is especially apparent given model performances when all data are available (~300,000), where all 100M traditionally fine-tuned models achieved $accuracy$ scores of above 95% and $f1$ of around 0.9. When given only 10 examples per class, the best performing model ($bert-base-uncased$) only retained around 60% of its all-data $f1$ score.

In AEC applications like road maintenance, the tolerance for error is low as tasks often require strict adherence to codes and regulations. For instance, the road inspec-

Table 7. Road codes attaining the highest and lowest classwise $f1$ scores for a 10-shot SetFit.

Road Code C	Classwise $f1$
Highest	
Failure to use plating and/or decking that is skid-resistant in its entirety on roadway	0.962
Failure to begin emergency work within 2 hours after authorisation	0.946
Construction material/equipment without proper reflective markings	0.942
Lowest	
Failure to restore in kind any parking or regulatory signs or supports	0.085
Failure to maintain a 5-foot pedestrian walkway on sidewalk	0.076
Failure to obtain permit for temporary roadway pavement markings changes	0.058

tion text classification task in this work involves road codes from roadwork licenses which have important implications for safety and disruption mitigation. By only attaining moderate values in few-shot settings, merely using traditionally fine-tuned models, even state-of-the-art LLMs, is not sufficient. Specialised few-shot techniques appear to be a better choice in resource-constrained settings even in a technical many-class application like road maintenance.

4.5 Classwise Performance Imbalance

There is a wide variation in the performance on each of the 83 classes. Figure 5 shows the histogram of the $f1$ scores for individual classes for a SetFit model trained with 10 examples per road code. There is a relatively high standard deviation across classes of 0.262, considering values are bounded between 0 and 1. Most classes exhibit moderately high $f1$ scores of around 0.75 or more, though a few classes with very low scores of below 0.25.

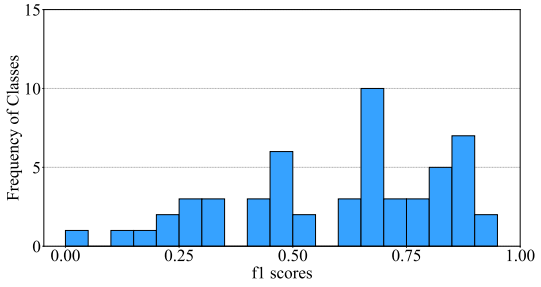


Figure 5. Histogram of $f1$ scores of individual classes for a 10-shot SetFit model.

Table 7 lists the three road codes with the highest and lowest $f1$ scores. The top codes all have high scores of around 0.95, whereas the bottom three have less than 0.1. One potential reason for the latter is that some codes have very specific definitions, such as requiring a "5-foot pedestrian walkway". In addition, there can be potential code duplications, such as the codes *Failure to obtain permit for temporary roadway pavement markings changes* and *Failure to obtain DOT permit for any changes to, or installation of, temporary roadway pavement markings and temporary construction, parking or regulatory signs and supports*, both of which cover very similar violation instances. These can lead to inconsistently annotated data that can both confuse the model during training and lead to a drop in prediction metric values.

To inspect the model's outputs, Table 8 provides examples of the predicted road code P for the inspection comment T with ground truth label L. The first example shows a correct classification where the model inferred that the code *Failure to properly place and ramp plating and decking* relates to examples with ramp elements. However, the second example shows an example of when it fails to infer this relationship for another code, possibly due to the presence of unrelated information from the inspection comments, such as instructions for charge admission.

5 Conclusions

Results indicate that dedicated few-shot techniques can outperform traditionally fine-tuned LLMs when classifying technical road maintenance texts in a resource-constrained setting, in addition to requiring less training data and computational resources than big LLMs with 7B parameters. For example, in 1-shot, the few-shot model SetFit outperformed the traditionally fine-tuned *Llama3* by three times in $f1$. In particular, SBERT used in SetFit appears to be a better pre-trained model choice than many larger models not pre-trained on similarity-based tasks.

However, an unresolved issue of few-shot techniques is that absolute performance is still moderate. With one example per class, the $accuracy$ is only 29%, though that increases to 69% with ten. In road maintenance and many tightly regulated AEC applications where reliability is crucial, this can impact their practicality. For instance, in this study on classifying inspection comments into roadwork licensing codes, there are important safety and disruption implications and necessitate high reliability. This prompts the need to explore alternative, specialised techniques that can tackle the domain-specific nature of the task in a more accurate and data-efficient manner.

6 Acknowledgement

Funded by the European Union

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101034337.

Table 8. Classification examples for inspection comments using a 10-shot SetFit.

Inspection Comment Example T	Prediction of Violated Road Code P	Correct Violated Road Code L	Outcome
I observed respondent failed to spike/ fasten 2 steel plates marked CWP at n/e/c of intersection to prevent movement.	Failure to properly place and ramp plating and decking	Failure to properly place and ramp plating and decking	Correct
A/T/P/O I observed the above respondent failed to repair a 1" gap within their transformer vault on the sidewalk. respondent was previously notified to correct this condition by via (CAR #20164060240) on 8/11/16. If not admitting the charge, you MUST APPEAR IN PERSON.	Failure to repair defective street condition found within an area extending 12 inches outward from the perimeter of the cover/grating	Failure to replace loose, slippery or broken utility maintenance hole (manhole) covers, castings, and other street hardware	Incorrect

References

[1] J. Xu, N. R. Anvo, H. A. Taha, A. M. d'Avigneau, D. Palin, R. Wei, G. Hadjidemetriou, S. Schaefer, L. de Silva, A. Al-Tabbaa, F. Lida, and I. Brilakis. Highway digital twin-enabled autonomous maintenance plant: a perspective. *Data-Centric Engineering*, 5:e24, 2024. doi:10.1017/dce.2024.34.

[2] A. Marie d'Avigneau, L. Potseluyko, N.R. Anvo, H.M.T. Abdalgadir, V.K. Reja, D. Davletshina, P. Lam, L. de Silva, A. Al-Tabbaa, and I. Brilakis. Camhighways: The cambridge highways dataset. *Advanced Engineering Informatics*, 2025. doi:10.1016/j.aei.2024.103036.

[3] D. Davletshina, V. K. Reja, and I. Brilakis. Automating construction of road digital twin geometry using context and location aware segmentation. *Automation in Construction*, 168:105795, 2024. doi:10.1016/j.autcon.2024.105795.

[4] R. A. Khalil, Z. Safelnasr, N. Yemane, M. Kedir, A. Shafiqurrahman, and N. Saeed. Advanced learning technologies for intelligent transportation systems: Prospects and challenges. *IEEE Open Journal of Vehicular Technology*, 5:397–427, 2024. doi:10.1109/OJVT.2024.3369691.

[5] A. Nevatia, S. Saha, S.B. Bhagavatula, and N. Bugalia. A pre-trained language model-based framework for deduplication of construction safety newspaper articles. In *40th ISARC*, 2023.

[6] S. Malihi, L. Potseluyko, A. Mathew, H. Alavi, V. K. Reja, Y. Pan, L. Binni, G. Wang, X. Wang, and I. Brilakis. Review of multimodal data and their applications for road maintenance. *Smart Construction*, 1(2):1–40, 2024.

[7] Michael P. Brundage, Thurston Sexton, Melinda Hodkiewicz, Alden Dima, and Sarah Lukens. Technical language processing: Unlocking maintenance knowledge. *Manufacturing Letters*, 27:42–46, 1 2021. ISSN 2213-8463. doi:10.1016/J.MFGLET.2020.11.001.

[8] S. Moon, G. Lee, and S. Chi. Automated system for construction specification review using natural language processing. *Advanced Engineering Informatics*, 51:101495, 2022. ISSN 1474-0346. doi:https://doi.org/10.1016/j.aei.2021.101495.

[9] Y. Wang, Y. Zhu, W. Xiong, and C. S. Cai. A few-shot word-structure embedded model for bridge inspection reports learning. *Advanced Engineering Informatics*, 62:102664, 2024. ISSN 1474-0346. doi:10.1016/j.aei.2024.102664.

[10] A. Pal, D. Murguia, and C. Middleton. Automatic inference of construction delays through analysis of weekly progress reports using llms. 2024. ISSN 2706-6568. URL `http://itc.scix.net/paper/w78-2024-108`.

[11] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need, 2023. URL `https://arxiv.org/abs/1706.03762`.

[12] D. Bahdanau, K. Cho, and Y. Bengio. Neural machine translation by jointly learning to align and translate, 2016. URL `https://arxiv.org/abs/1409.0473`.

[13] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.

[14] L. Tunstall, N. Reimers, U. E. S. Jo, L. Bates, D. Korat, M. Wasserblat, and O. Pereg. Efficient few-shot learning without prompts, 2022.

[15] NYC Open Data : City of New York. Nyc open data. URL `https://tinyurl.com/549hjywn`. (accessed 2024-11-15).

[16] S. Casola, I. Lauriola, and A. Lavelli. Pretrained transformers: an empirical comparison. *Machine Learning with Applications*, 9:100334, 2022. ISSN 2666-8270. doi:https://doi.org/10.1016/j.mlwa.2022.100334.

[17] Aristides Milios, Siva Reddy, and Dzmitry Bahdanau. In-context learning for text classification with many labels, 2023. URL `https://arxiv.org/abs/2309.10954`.

[18] Aleksandra Edwards and Jose Camacho-Collados. Language models for text classification: Is in-context learning enough?, 2024. URL `https://arxiv.org/abs/2403.17661`.

Automated Extraction of Structural Elements and Section Views from Bridge Drawings Using Deep Learning

Hakan Bayer[1] **and Markus König**[1]

[1]Chair of Computing in Engineering, Ruhr University Bochum, Bochum, Germany

hakan.bayer@ruhr-uni-bochum.de, koenig@inf.bi.rub.de

Abstract -

Technical drawings are essential for managing and maintaining existing bridges, especially in creating accurate digital models and integrating them into Building Information Modeling (BIM) workflows. However, digitizing older bridges, which often lack BIM models is challenging due to the time-consuming and error-prone process of manual data extraction from 2D drawings. Therefore, this study focuses on detecting key structural components, such as anchors, tendons, and section views, which are critical for structural analysis, maintenance, and retrofitting. Using state-of-the-art object detection models, YOLOv8 and YOLOv11, the proposed pipeline achieves a mean average precision (mAP) of 97.4% for anchors, 95.7% for tendons, and 63.9% for section views. Optical Character Recognition (OCR) tool Pytesseract further enhances the process by accurately extracting IDs for anchors and tendons with 90% precision. Combining object detection with precise text extraction simplifies the creation of high-quality digital bridge models and improves BIM integration. Specifically, the extracted data provides semantic information on key structural components. This information can facilitate the process of creating a digital model and integrating it into BIM workflows.

Keywords -

Building Information Modeling; Computer Vision; Deep Learning; Symbol Detection; Optical Character Recognition; Construction Drawings

1 Introduction

Bridges are a crucial component of transportation infrastructure that enables the transportation of people and goods. Maintaining their functionality and ensuring safe operations requires routine inspections and effective maintenance management. Digital tools like Building Information Modelling (BIM) can enhance maintenance planning efficiency. BIM is a collaborative digital methodology that involves creating and sharing interdisciplinary digital models of structures [1]. These models contain semantic and geometric information, sharing the state of the project in real time.

Among all phases of a project's life cycle, the operations and maintenance (O&M) phase presents significant opportunities for BIM application [2]. A key challenge however lies in the limited availability of accurate digital building models for existing structures [3]. In many cases, this is because the old constructions were designed and constructed without the implementation of BIM. Consequently, it is necessary to develop these models in retrospect.

The digitization of existing high-rise buildings and infrastructure is particularly difficult, requiring the expertise of architects and engineers to reconstruct geometric and semantic information from various sources such as point clouds, images, and construction drawings [4]. While point clouds and images are valuable resources, they often rely on costly equipment and fail to reveal internal structural details. In contrast, construction drawings offer a more accessible and practical alternative, especially for older bridges where they are often the only available source of information. As a result, construction drawings are frequently the preferred choice for creating digital building models.

Extracting key information from bridge construction drawings, including section views, structural elements, and their identifiers, is a crucial initial step in developing accurate digital models. These drawings often feature critical internal components like anchors, tendons, and section views, which are essential for understanding bridges' structural integrity and functionality. Anchors for instance are vital elements in prestressed concrete bridges used to secure tendons and transfer tensile forces to the surrounding concrete. Tendons act as reinforcement ensuring the bridge can handle heavy loads and maintain stability. On the other hand, section views offer detailed representations of inner components enabling engineers to understand their precise placement and spatial relationships within the bridge structure.

This research focuses on automating the extraction of these critical components, including section views, anchors, tendons, and their identifiers from CAD based bridge technical drawings to support the development of accurate digital models. Using state-of-the-art deep learning frameworks, such as YOLOv8 [5] and YOLOv11 [6] in combination with OCR tools like PyTesseract [7], the proposed method achieves precise detection of these components and extracts relevant semantic information. This output data facilitates the creation of digital bridge models, enabling better integration into BIM workflows hence supporting more efficient bridge maintenance and management.

2 Related Work

In deep learning and computer vision, significant attention has been given to reconstructing high-rise building models from technical drawings, with a primary focus on detecting and reconstructing external elements such as walls, rooms, and openings [8]. For instance, [8] utilized ResNet to identify walls in floor plans, while [9] leveraged conditional GANs (cGANs) to generate heat maps and vectorized representations of building layouts. These approaches have proven effective for creating digital building models from architectural drawings.

However, the digitization of technical drawings, particularly for infrastructures such as bridges has received limited attention in research [10]. Although some studies are exploring semi-automated analysis of such drawings, the field remains underdeveloped. [11] proposed a semi-automated process employing the Douglas-Peucker algorithm to detect corner points of illustrated components. These corner points are then used to reconstruct the external geometry of the components, which ultimately generates the structural outline of the bridge. [12] introduced a different semi-automated framework that converts drawings into vector-based formats. This process aligns the extracted views and uses the gathered data to reconstruct the geometry of the bridge. The resulting digital model is then exported in the IFC (Industry Foundation Classes) format for further use.

More recently, deep learning-based approaches have been explored in this domain. [13] presented a pipeline using YOLOv5 and the CRAFT text detector to identify individual components and labels on technical drawings. The detected objects were clustered based on their labels to account for multiple views, with experts manually reviewing the clustered data to create bridge models. Similarly, [14] employed Faster R-CNN to detect section symbols and applied OCR to extract view titles, linking symbols to their corresponding views and establishing interconnections across drawing sets. [15] utilized YOLOv5 for gradient symbol detection and integrated EasyOCR for text extraction, automating the process of extracting bridge gradients from construction drawings. Additionally, the study by [16] introduced a computer vision-based process for generating bridge superstructure models from pixel-based construction drawings. Their approach leverages Keypoint R-CNN for symbol pose estimation, enabling the spatial organization of drawing views and facilitating the development of accurate digital bridge models.

Despite these advancements, no study to date has addressed the automated extraction of inner structural components of bridges, such as anchors, tendons, and section views, from pixel-based technical drawings. This study fills this gap by employing YOLOv8 [5] and YOLOv11 [6] to detect these critical components and PyTesseract [7] to extract their associated IDs. These advancements enable the digitization of bridge inner structures and facilitate the creation of digital models for integration into BIM workflows.

3 Methodology

This study presents a semi-automated pipeline for detecting anchors, tendons, and section views in CAD based 2D bridge technical drawings, as well as extracting associated IDs for anchors and tendons. The methodology integrates classical computer vision techniques, deep learning-based object detection, and OCR for text recognition. The workflow, as illustrated in Figure 1, consists of three main stages: (1) Object Detection, (2) Shape Detection, and (3) Text Extraction.

As shown in Figure 1, the pipeline begins with a pixel-based bridge drawing input, which contains cross-sectional representations of internal structural components such as anchors, tendons, and section views. The first stage involves object detection, where elements such as anchors, tendons, and section views are detected using an object detection model. In the second stage, shape detection was performed using OpenCV to locate ID boxes by detecting rectangular regions related to anchor and tendon elements. Finally, the text extraction stage employs OCR to extract the ID numbers from the identified rectangle regions. The pipeline outputs the detected section views, anchors, tendons, and associated IDs, enabling efficient process and analysis of bridge drawings.

3.1 Dataset Creation and Annotation

Due to the limited availability of real bridge technical drawings, a combination of real and synthetic data was used to create the dataset. The original dataset comprised 62 internal structural drawings of bridges from Germany. These drawings were divided into sections, resulting in 71 images. Of these, 42 images were designated for training and validation, while 29 were set aside for testing the pipeline. The 42 training-validation images were particularly large, with an average width of 11,290 pixels and an average height of 7,744 pixels.

Since the 42 training-validation images were insufficient for effective model training, synthetic data was generated by overlaying cropped symbols (anchors, tendons, and section views) onto real construction drawing backgrounds that did not initially contain these symbols using copy-paste method. This approach ensured a diverse and robust dataset for training and validation.

The train validation dataset consists of three primary components: anchor element dataset, tendon element dataset, and section view dataset. The anchor element dataset includes two classes: longitudinal anchor sections

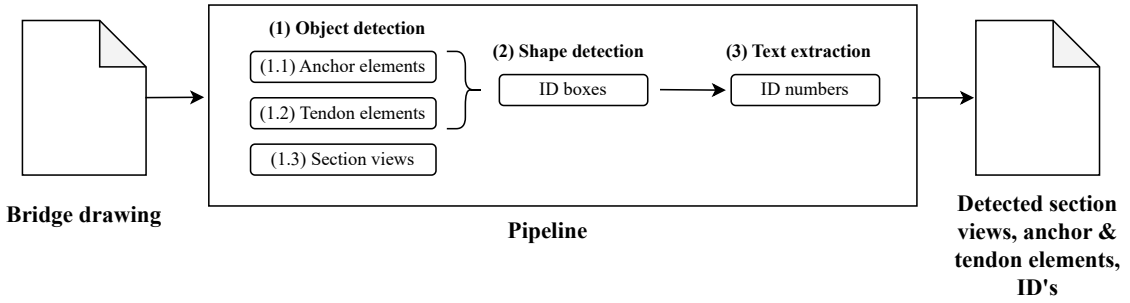


Figure 1. Workflow diagram illustrating the detection of anchors, tendons, and section views in bridge drawings, along with the extraction of associated IDs.

and front anchor sections. Similarly, the tendon dataset contains only one class which is front tendon sections. The section views dataset contains four types of classes: longitudinal section view, front sectional view of the final anchor point, cross-section view, and longitudinal section view of the final anchor point. Some example annotated objects of the dataset classes are shown in Fig. 2. These annotations capture the variability and complexity of the internal structural drawings.

3.1.1 Anchor Dataset

From the real training-validation dataset, 204 longitudinal anchor symbols and 74 front anchor symbols were selected and cropped out. The longitudinal anchor symbols had an average width of 292 pixels and an average height of 163 pixels, while the front anchor symbols had an average width of 93 pixels and an average height of 95 pixels. These symbols were randomly rotated (0°, 90°, -90°) and overlaid onto real construction drawing backgrounds to create synthetic images. This process generated 500 synthetic labeled images, which were automatically annotated in YOLO bounding box format. This dataset was then combined with the 42 annotated real bridge drawings, which created final anchor dataset of 542 images. The dataset was then split into 80% training and 20% validation sets.

3.1.2 Tendon Dataset

The tendon dataset was created using 53 tendon front sectional views extracted from the real training-validation dataset. These tendon elements had an average width of 41 and height of 41 pixels. These cropped symbols were randomly positioned on real construction drawing backgrounds, and their locations and dimensions were automatically labeled in YOLO bounding box format. Similar to the anchor dataset, this process generated 500 synthetic labeled images, resulting in a final tendon dataset of 542 images, which was also split into 80% training and 20% validation sets.

3.1.3 Section View Dataset

The section view dataset includes four types of classes: longitudinal section view, front section view of the final anchor point, cross-section view , and longitudinal section view of the final anchor point. The longitudinal section view depicts the internal arrangement of the entire bridge along its length, while the front section view of the final anchor point focuses on the final anchor point and its components from a frontal perspective. Cross-section view shows vertical cuts along the bridge's width, highlighting structural details, and longitudinal section view of the final anchor point provides a longitudinal perspective of the anchor components. The section views are pixel wise relatively large compared to tendon and anchor elements. Due to their complexity, size, and variability, huge number of synthetic data was generated compared to the other datasets. CAD files of the training and validation set were used to generate additional data by modifying attributes such as size, perspective, and surrounding text. Also a copy-paste method similar to that used for the anchor and tendon datasets was employed. In total, 1,842 images were generated including 42 annotated real bridge drawings. All images were automatically labeled using the YOLO bounding box format. Then the final dataset was divided into 80% training and 20% validation sets.

3.2 Object Detection

The detection pipeline utilizes multiple YOLO architectures, YOLOv8 [5] and YOLOv11 [6], trained on separately annotated datasets for anchor elements, tendon elements, and section views. Each dataset was trained on models of varying scales, such as medium and large configurations. The inference image size was chosen to match the respective training image size to maximize accuracy.

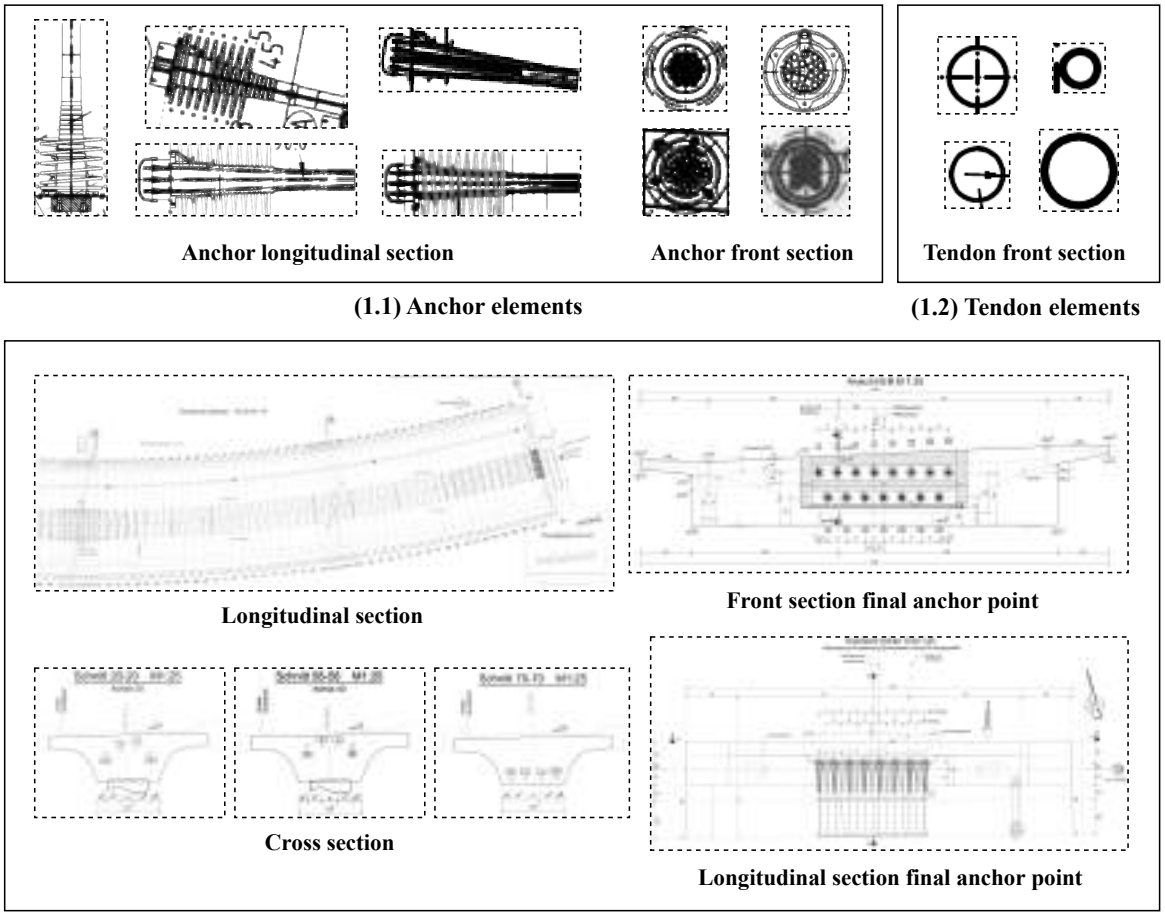


Figure 2. Examples of annotations from the dataset, highlighting anchor elements (1.1) and tendon elements (1.2), along with 4 different section views (1.3). These include longitudinal section views, cross-sections, front section views of the final anchor point, and longitudinal section views of the final anchor point. These annotations are crucial for training object detection models and generating synthetic data.

After training, the best models were integrated into the detection pipeline to accurately detect objects.

3.3 Shape Detection

The shape detection stage focuses on identifying rectangular regions in the drawings that likely contain IDs associated with anchors and tendons. This is achieved using classical computer vision techniques implemented with OpenCV. The process starts by converting the input images to grayscale, followed by applying Gaussian blur to reduce noise and enhance edge details. Adaptive thresholding is then applied to create a binary image, which simplifies edge detection.

Contours are extracted from the binary image to detect potential rectangular shapes. These shapes are filtered based on specific criteria, such as their area and aspect ratio, to isolate valid ID boxes. During this stage, we assume that IDs are enclosed within rectangular shapes based on the standard drawing conventions of a specific drawing format. It is important to note that some firms may use different drawing conventions, such as enclosing IDs within circular or elliptical shapes. Our pipeline, however, is made to function with rectangular ID boxes following the conventional norms utilized in these particular drawings for accuracy and consistency.

3.4 Text Extraction

The text extraction stage employs PyTesseract to extract textual identifiers from the ID boxes identified during the shape detection process. To enhance the precision of text recognition, the images are preprocessed prior to applying OCR. This preprocessing involves converting images to grayscale and using adaptive thresholding to improve contrast and text clarity within the ID boxes. To prevent errors caused by extracting text from unrelated elements in the drawings, the OCR process is restricted to a predefined character set that includes numbers, letters, hyphens, slashes, and periods. This restriction ensures that only valid identifier formats are captured, minimizing the inclusion of irrelevant or incorrect text. Once the text is extracted, it is mapped to the corresponding tendon or anchor elements based on the spatial relationship between the detected text and ID boxes. As section views do not include associated identifiers, text extraction is performed exclusively for tendon and anchor elements.

4 Training and Results

4.1 Training Process

The training process utilized eight NVIDIA A100 GPUs. The datasets for anchors, tendons, and section views were trained independently due to their unique characteristics, varying object sizes, and complexities. To evaluate detection performance, input image sizes were adjusted based on the relative scale and complexity of the objects in each dataset.

Both YOLOv8 [5] and YOLOv11 [6] models were trained for 100 epochs on each dataset, using two different image resolutions to compare the impact of varying sizes on model performance. Additionally, two model configurations, medium and large, were employed. Higher resolutions were used for the tendon and anchor datasets, as these objects are relatively small compared to the overall size of the drawings. In contrast, moderate resolutions were applied to the section view dataset, which contains larger objects. Specifically, the tendon dataset was trained with image sizes of 3200 and 3600 pixels, the anchor dataset with 2560 and 3200 pixels, and the section view dataset with 640 and 1280 pixels.

This setup allowed the models to effectively capture the distinct features of each dataset, enabling robust detection performance and ensuring adaptability to the diverse complexities of technical bridge drawings.

4.2 Object Detection Results

The performance of the YOLOv8 and YOLOv11 models was evaluated across three datasets: section view, anchor, and tendon, using a testing set of 29 unseen images. Evaluation metrics included Precision (P), Recall (R), and mean Average Precision at 50% IoU threshold (mAP50). Table 1 summarizes the results for each dataset and model configuration.

For the anchor dataset, the YOLOv11 large model achieved an impressive overall mAP50 of 0.974, with a Precision of 0.917 and a Recall of 0.970. Among the specific classes, the Anchor front section class achieved the highest mAP50 of 0.994, demonstrating near-perfect detection accuracy. The Anchor longitudinal section class also showed strong performance, with a mAP50 of 0.954.

In the tendon dataset, the YOLOv11 large model delivered a mAP50 of 0.957, with a Precision of 0.914 and a Recall of 0.952. The use of high-resolution images during training and testing allowed the model to effectively capture and detect the fine-grained details of small and elongated objects like tendons.

For the section view dataset, the YOLOv11 large model achieved an overall mAP50 of 0.639, with a Precision of 0.712 and a Recall of 0.623. However, performance varied significantly across individual classes. The Cross-section class achieved the highest mAP50 of 0.858, demonstrating robust detection capabilities. In contrast, the Longitudinal section and Longitudinal section final anchor point classes showed lower mAP50 values of 0.436 and 0.565, respectively. The Front section final anchor point class achieved a moderate mAP50 of 0.697. The lower performance in section view detection can be explained due to the high variability in section drawings and limited real dataset samples. Unlike anchors and tendons, section views differ significantly across drawings making object detection more challenging.

Overall, the YOLOv11 large model consistently outperformed YOLOv8 across all datasets, demonstrating superior accuracy and robustness in handling diverse object detection tasks. The results also highlight the importance of using appropriately high-resolution images during training and testing particularly for datasets with smaller objects as it significantly enhances detection performance.

4.3 Overall Pipeline Results

The proposed pipeline demonstrates strong performance in detecting section views, anchor and tendon symbols, and extracting IDs even when faced with challenges such as rotation, scale differences, and partial visibility. Testing on 29 real images showed that the pipeline reliably identified anchors, tendons, and section views, achieving high object detection accuracy. However, text extraction faced some issues, particularly when shapes were rotated, making it harder to identify IDs. As outlined in the *Shape Detection* section, the pipeline's shape detection relies on drawings adhering to standard conventions such as rectangular ID boxes commonly used by a specific firm. Based

Table 1. Object Detection Results of YOLOv8 and YOLOv11 Models on Test Dataset (29 Images) for Section View, Anchor, and Tendon Datasets with Different Image Sizes. Metrics include Precision (P), Recall (R), and mAP50. The table provides the average performance across all object classes within each dataset, offering an overall assessment of detection accuracy. Bold fonts indicate the best results for each metric.

Dataset	Image Size	Model	P	R	mAP50
Anchor Dataset	2560	YOLOv8m	0.952	0.927	0.964
	3008	YOLOv8m	0.958	0.942	0.970
	2560	YOLOv8l	0.958	0.893	0.949
	3008	YOLOv8l	0.934	0.954	0.971
	2560	YOLOv11m	**0.960**	0.927	0.955
	3008	YOLOv11m	0.955	0.947	**0.974**
	2560	YOLOv11l	0.952	0.933	0.957
	3008	YOLOv11l	0.917	**0.970**	**0.974**
Tendon Dataset	3200	YOLOv8m	0.899	0.837	0.895
	3600	YOLOv8m	0.925	0.902	0.936
	3200	YOLOv8l	0.914	0.865	0.934
	3600	YOLOv8l	0.928	0.895	0.930
	3200	YOLOv11m	0.900	0.866	0.915
	3600	YOLOv11m	0.922	**0.944**	0.959
	3200	YOLOv11l	**0.932**	0.887	0.943
	3600	YOLOv11l	0.914	0.952	**0.957**
Section View Dataset	640	YOLOv8m	**0.813**	0.546	0.590
	1280	YOLOv8m	0.566	0.544	0.484
	640	YOLOv8l	0.578	0.5 60	0.535
	1280	YOLOv8l	0.519	0.618	0.529
	640	YOLOv11m	0.702	0.615	0.585
	1280	YOLOv11m	0.634	0.595	0.535
	640	YOLOv11l	0.528	0.590	0.555
	1280	YOLOv11l	0.712	**0.623**	**0.639**

on this assumption the text detection accuracy was calculated by determining the percentage of correctly recognized parameters relative to the total number of parameters present. The pipeline achieved 90% accuracy in text extraction, demonstrating its ability to extract IDs in most cases. However, performance was hindered by rotated text boxes and slight variations in font styles and sizes. Future refinements could include pre-processing techniques like rotation correction and fine-tuning the model for the dataset to enhance recognition accuracy.

An example of the pipeline's performance is illustrated in Fig. 3, where it successfully detects objects and extracts text from real test drawings. These results highlight the pipeline's ability to support the automatization process of digitization of technical drawings and extract critical information, making it a valuable tool for further applications in bridge modeling.

5 Discussion and Conclusion

This study presents a semi-automated pipeline for extracting structural components and their identifiers from bridge technical drawings, addressing the challenges of digitizing older structures. By integrating deep learning-based object detection models, YOLOv8 and YOLOv11, with the optical character recognition tool PyTesseract, the pipeline achieves high accuracy in identifying and extracting critical elements and their associated information.

The YOLOv11 large models demonstrated the best performance, achieving mAP50 scores of 0.974 for anchor elements and 0.957 for tendon elements. Detection of section views showed moderate success, with an mAP50 of 0.639. Text extraction using PyTesseract achieved 90% accuracy. However, the pipeline was tested on a limited number of drawings following specific rectangular ID standards, which restricts its generalizability. This limitation highlights the need to generalize the approach for broader applicability. Additionally, the assumption of rectangular ID boxes restricts generalizability to other formats such as circular or elliptical IDs. Furthermore, OCR struggles with rotated or noisy text which could be improved with advanced preprocessing techniques.

Despite these achievements, challenges remain particularly in text recognition and adapting the pipeline to diverse drawing standards. Future research could focus on fine-tuning the OCR model for bridge drawings, expanding shape detection to accommodate circular or elliptical ID boxes, and increasing the range of synthetic training datasets. Exploring alternative detection models may also enhance performance, especially for complex components like section views.

Overall, this study presents a solution for automating the digitization of bridge technical drawings. The extracted section views and elements, along with their IDs, can serve as structured inputs for BIM software enabling efficient reconstruction. The proposed method streamlines the creation of accurate digital models and facilitates their integration into BIM workflows.

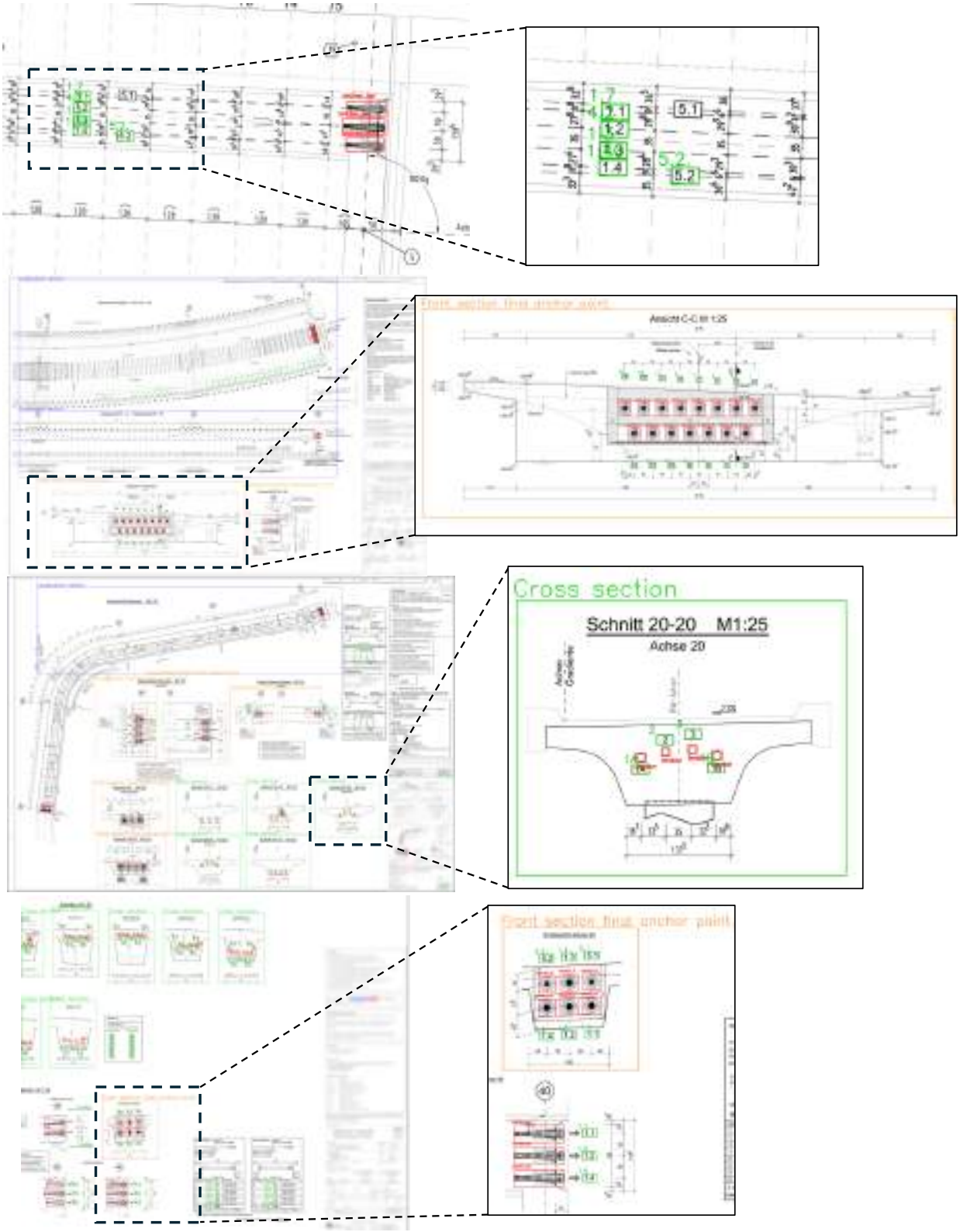


Figure 3. Demonstration of the pipeline's performance on a real test dataset using the best-performing models with the highest mAP50 values. The figure showcases the accurate detection of section views (e.g., cross sections, front section final anchor points), identification of symbols, and precise extraction of their associated IDs. Enlarged views highlight specific areas of successful object detection and ID extraction, underscoring the pipeline's effectiveness in processing complex engineering drawings.

References

[1] André Borrmann, Markus König, Christian Koch, and Jakob Beetz. Building Information Modeling – Why? What? How? In *Building Information Modeling - Technology Foundations and Industry Practice*, pages 1–24. Springer, 09 2018. doi:10.1007/978-3-319-92862-3.

[2] Dania Abideen, Akilu Yunusa-Kaltungo, Patrick Manu, and Clara Cheung. Digitalization of construction life cycle: A systematic review of building and reliability information modelling (BRIM). In *AIP Conference Proceedings*, volume 2428, page 020001, 11 2021. doi:10.1063/5.0070961.

[3] Xu Yiye and Yelda Turkan. BrIM and UAS for bridge inspections and management. *Engineering, Construction and Architectural Management*, 27: 785–807, 2019. doi:10.1108/ecam-12-2018-0556.

[4] Phillip Schönfelder, Angelina Aziz, Benedikt Faltin, and Markus König. Automating the retrospective generation of as-is BIM models using machine learning. *Automation in Construction*, 152:104937, 2023. doi:10.1016/j.autcon.2023.104937.

[5] Glenn Jocher, Ayush Chaurasia, Jiankai Qiu, and Anthony Stoken. Yolov8: State-of-the-art object detection models. Online: `https://docs.ultralytics.com/models/yolov8/`, 2023. Accessed: 26/03/2025.

[6] Glenn Jocher, Ayush Chaurasia, Jiankai Qiu, and Anthony Stoken. Yolov11: State-of-the-art object detection model. Online: `https://docs.ultralytics.com/models/yolo11/`, 2024. Accessed: 26/03/2025.

[7] Samuel Hoffstaetter. Pytesseract: Python wrapper for google tesseract-ocr. Online: `https://github.com/madmaze/pytesseract`, 2023. Accessed: 26/03/2025.

[8] Chialing Wei, Mohit Gupta, and Thomas Czerniawski. Automated wall detection in 2D cad drawings to create digital 3D models. In *Proceedings of the 39th International Symposium on Automation and Robotics in Construction*, pages 152–158, 2022. doi:10.22260/ISARC2022/0023.

[9] Seongyong Kim, Seula Park, Hyunjung Kim, and Kiyun Yu. Deep floor plan analysis for complicated drawings based on style transfer. *Journal of Computing in Civil Engineering*, 35(2):04020066, 2021. doi:10.1061/(ASCE)CP.1943-5487.0000942.

[10] C. F. Moreno-García, Eyad Elyan, and Chrisina Jayne. New trends on digitisation of complex engineering drawings. *Neural Computing and Applications*, 31:1695–1712, 6 2019. doi:10.1007/s00521-018-3583-1.

[11] Kwasi Nyarko Poku-Agyemang and Alexander Reiterer. 3D reconstruction from 2D plans exemplified by bridge structures. *Remote Sensing*, 15, 2023. doi:10.3390/rs15030677.

[12] Temitope Akanbi and Jiansong Zhang. Semi-automated generation of 3D bridge models from 2D PDF bridge drawings. In *Construction Research Congress 2022*, pages 1347–1354, 2022. doi:10.1061/9780784483961.141.

[13] M. Saeed Mafipour, Daniyal Ahmed, Simon Vilgertshofer, and Andre Borrmann. Digitalization of 2D bridge drawings using deep learning models. In *The 30th EG-ICE: International Conference on Intelligent Computing in Engineering*, 2023. URL `https://www.ucl.ac.uk/bartlett/construction/sites/bartlett_construction/files/digitalization_of_2d_bridge_drawings_using_deep_learning_models.pdf`.

[14] Benedikt Faltin, Phillip Schönfelder, and Markus König. Inferring interconnections of construction drawings for bridges using deep learning-based methods. In *ECPPM 2022-eWork and eBusiness in Architecture, Engineering and Construction 2022*, volume 62, pages 343–350, 2023. doi:10.1201/9781003354222-44.

[15] Hakan Bayer, Benedikt Faltin, and Markus König. Automated extraction of bridge gradient from drawings using deep learning. In *CONVR 2023 - Proceedings of the 23rd International Conference on Construction Applications of Virtual Reality*, 2023. doi:10.36253/979-12-215-0289-3.68.

[16] Benedikt Faltin, Phillip Schönfelder, Damaris Gann, and Markus König. Reconstructing as-built beam bridge geometry from construction drawings using deep learning-based symbol pose estimation. *Advanced Engineering Informatics*, 62:102808, 2024. doi:10.1016/j.aei.2024.102808.

Proof-of-Concept Framework of Integrating AI-based LLM Reasoning into BIM Workflows for Design Automation

Jinwu Xiao[1], Cansu Coskun[1], Soowon Chang[1] *, Kyubyung Kang[1], Deniz Besiktepe[1]

[1]School of Construction Management Technology, Purdue University, United States

*** Corresponding Author**

xiao270@purdue.edu, ccoskun@purdue.edu, soowonchang@purdue.edu, kyukang@purdue.edu, denizb@purdue.edu

Abstract -

Building Information Modeling (BIM) has significantly advanced design efficiency and collaboration in the architecture, engineering, and construction (AEC) industry. However, current BIM workflows remain hindered by complex interfaces, repetitive tasks, and limited automation options, which result in a steep learning curve and labor-intensive design processes. To address the challenges, this research presents a proof-of-concept framework that explores the integration of Artificial Intelligence (AI) and Large Language Models (LLMs) with BIM environments to streamline design modifications through natural language interaction. Using a specialized FreeCAD integration layer, the framework demonstrates the feasibility of translating high-level user instructions into basic BIM operations. Initial testing shows this approach can successfully create simple parametric building elements while reducing token consumption and user effort compared to a baseline approach using generic tools. This feasibility study suggests potential for developing more intuitive BIM interfaces, though further research is needed to achieve robust domain knowledge integration and handle complex architectural relationships. The findings indicate promising directions for future development of LLM-driven BIM automation tools that could enhance accessibility for AEC professionals.

Keywords -

BIM Design Automation; Natural Language Processing; Computer-Aided Design; Human-Computer Interaction

1 Introduction

Building Information Modeling (BIM) has revolutionized the architecture, engineering, and construction (AEC) industry by providing comprehensive digital representations of the physical and functional characteristics of buildings [1]. Although BIM platforms have significantly improved design efficiency and collaboration, practitioners still face challenges such as complex interfaces, repetitive tasks, and redundant manual operations, which increases the need for automated workflows [2]. However, current BIM workflows lack user interaction and accessibility to automation. Traditional approaches to BIM interaction lack intuitive interfaces for design modifications and automation capabilities, leading to reduced efficiency and increased risk of errors during the design process [3].

Recent advances in Artificial Intelligence (AI) and Large Language Models(LLMs) have shown promising capabilities to understand natural language commands and execute complex computer operations [4]. The remarkable progress in AI has enabled architects to harness computational power and data-driven algorithms to generate and assess design options [5]. AI technologies in the design process have enabled architects to explore innovative and eco-friendly design solutions. Similarly, construction industry professionals have witnessed the evolution of digital data acquisition technologies, which have enhanced monitoring processes by delivering efficient and effective results [3]. Moreover, AI's ability to evaluate enormous volumes of data, including building performance, environmental conditions, and user preferences, enables architects to create structures that are not only visually appealing but also ecologically friendly, energy-efficient, and suited to specific user needs [2].

Current approaches to BIM automation mainly rely on visual programming tools such as Dynamo for Revit or scripting languages such as Python with IFCOpenShell [6]. Although these methods have proven to be effective, they require users to have programming knowledge or understand visual programming concepts, which creates a significant barrier for many AEC professionals [2]. Furthermore, existing automation solutions often lack the flexibility to adapt to varying user needs and project requirements, particularly when handling complex visual and spatial tasks [7]. The integration of AI in the design workflow can create a collaborative and symbiotic relationship between human designers and AI technology, where AI operates as a creative partner, increasing the architect's expertise and vision by offering design options and innovative ideas [5].

Natural language-based BIM modeling offers key advantages over traditional methods: reduced cognitive load when navigating complex interfaces, faster prototyping

by translating concepts directly to visual models, improved accessibility for AEC professionals with varied technical expertise, contextual design assistance through embedded standards, and enhanced collaboration through conversation-like interactions.

This research aims to enable natural language-driven BIM operations for architects and AEC professionals without requiring programming knowledge. We developed a proof-of-concept framework integrating AI and LLMs with BIM software operations. The prototype implements parameter-based modeling within FreeCAD [8], leveraging anthropic-defined computer use API and BIM-specific prompts.

Initial testing shows promising results. Compared to baseline systems, our prototype improved efficiency with reduced token usage and better parameter handling while avoiding GUI-based trial-and-error approaches. These results suggest LLM-driven BIM automation could provide more accessible tools for AEC professionals. This research validates that LLM-driven BIM automation can overcome traditional usability barriers, offering a foundation for future BIM-oriented AI developments as the AEC industry continues to digitalize.

2 Literature Review

The AEC industry is undergoing a major transformation, shifting from traditional and labor-intensive methods to automation driven by information technology, with BIM as a key factor in this change [9]. BIM is a technology-based method that applies and maintains a complete digital representation of building information throughout a project life cycle [10]. The process begins with conceptual design, progresses through functionality and construction details, and culminates in fabrication drawings, with analysis tools for energy efficiency, construction, and acoustics supporting design decisions [11].

BIM offers numerous practical benefits including visualization, code compliance checks, cost estimation, scheduling, clash detection, and forensic analysis [12]. However, adoption faces challenges related to high costs, training time, unclear data ownership, lack of standardization, and software compatibility issues [13]. Bridging the gap between advanced BIM technologies and their practical application remains important [10].

Industry 4.0 advancements have influenced the construction sector, leading to "Construction 4.0" with its emphasis on digital data, automation, connectivity, and digital access [14]. Automation in construction improves efficiency with repetitive tasks, reduces expenses, enhances quality, and improves coordination [15]. In BIM specifically, automation tools like Dynamo enable custom scripts for repetitive tasks [16], while IFC ensures data exchange between platforms [17]. Such automation supports real-time clash detection before construction [18].

AI has emerged as a promising solution for automation in the AEC sector, making notable improvements in operations, safety, and productivity [19]. AI transforms architecture through generative platforms that create visuals from text prompts, making design processes more efficient. Generative algorithms allow architects to explore multiple design options considering site conditions and structural requirements [20], while also improving space planning [21] and sustainability through data-driven decision making [22].

Recently, researchers have explored human-AI collaboration in design, with Lee et al. [23] integrating LLMs with BIM for speech applications and Dai et al. [24] introducing semantic AI into architectural design. However, practical integration of AI tools into BIM-enabled projects remains limited. Interoperability and data transfer challenges between AI systems and BIM platforms pose significant barriers. More proof-of-concept testing is needed to bridge the gap between innovative AI solutions and their practical application in BIM workflows [25].

3 Methodology

This research integrates LLMs with BIM environments using direct computer control and custom tool integrations. By leveraging the anthropic-defined computer use API and FreeCAD as the BIM platform, this approach enables natural language-driven BIM operations without requiring programming knowledge.

Research Approach and Environment Setup The study employs a controlled case evaluation focusing on key representative BIM operations. The experimental environment runs in a Linux-based virtual machine with Docker containerization, ensuring reproducibility and isolation. FreeCAD serves as the primary BIM software, chosen for its robust Python API and scriptability. The system uses the Anthropic API(Claude) [26] to handle human-LLM-BIM interaction loops.

Tool Integration and Execution Flow The methodology uses a modular tool collection: ComputerTool for UI control, BashTool for system commands, EditTool for file operations, and FreeCADTool for BIM-specific operations. The LLM orchestrates these tools through a "sampling loop" process, translating natural language into structured commands. The system validates outputs through iterative interactions, ensuring creation of geometrically correct walls with precise dimensions, proper spatial relationships between elements, accurate parameter assignments, and verification of element connectivity.

Parameterization and Prompt Optimization The methodology incorporates iterative prompt optimization with domain knowledge and context-specific guidelines.

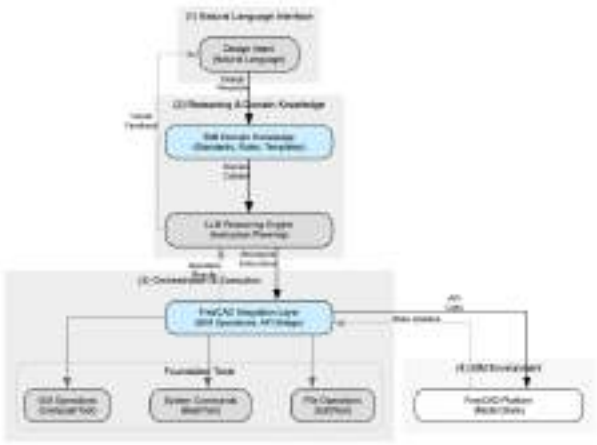


Figure 1. Framework architecture showing the hierarchical organization of components. The system flows from top to bottom: starting with the Architect's natural language instructions, through the BIM Knowledge and Prompt Layer, to the LLM (Claude), then to the FreeCAD Tool orchestrator, supported by base tools (ComputerTool, BashTool, EditTool), and finally to the FreeCAD Environment where modifications are realized.

This layered approach combines system prompts with specialized instructions for BIM tasks, enabling accurate interpretation of spatial information and parameter validation.

Evaluation Metrics and Validation The study compares baseline and enhanced frameworks using metrics including task completion success, resource efficiency (tokens, API calls, processing time), parameter accuracy, and visual verification. A rectangular room creation task serves as the benchmark for framework feasibility assessment.

4 Framework Design

The proposed framework implements a top-down hierarchy that transforms the intentions of natural language design of the architect into tangible BIM modifications within FreeCAD. As illustrated in Figure 2, the prompt structure and BIM knowledge form a conceptual foundation that guides the LLM (Claude) in understanding and interpreting the user request. These prompts encapsulate domain rules, measurement handling, and quality control, ensuring that the LLM reasons about BIM tasks in a way that is contextually rich and technically correct.

Building on this conceptual layer, the architectural diagram shown in Figure 1 represents the system as a vertical sequence of layers. At the top, the architect issues design instructions in plain language. These instructions flow downward through layers of reasoning, domain-specific BIM knowledge, and an orchestration tool (FreeCADTool). Ultimately, this cascade of logic and commands reaches the FreeCAD environment, where the requested modifications (e.g., creating a wall, inserting a window) materialize.

This hierarchical design simplifies complexity: each layer focuses on its core responsibilities. The LLM interprets intent and plans operations, the FreeCADTool orchestrates BIM-related commands, and the base tools perform low-level tasks such as GUI manipulation or file edits. By carefully layering the system in this manner, the framework allows nontechnical users to harness sophisticated BIM operations without directly dealing with scripting or intricate APIs.

4.1 Base Architecture from Claude and Tool Collection Framework

The initial foundation of this system comes from a Claude-based computer use platform equipped with a collection of core tools.

- **ComputerTool:** Handles GUI interactions within the virtual environment (e.g., mouse clicks, keystrokes).
- **BashTool:** Executes shell-level commands to manage environment states or run auxiliary programs.
- **EditTool:** Provides the ability to modify files dynamically, such as generating or adjusting Python scripts on-the-fly.

These tools, originally developed for generic computer interactions, form the base layer of the system (see lower half of Figure 1). On their own, they lack the contextual understanding needed for BIM-specific tasks. To address this, a BIM-oriented layer is introduced above them.

In this enhanced structure, the LLM uses the BIM knowledge and context templates (as summarized in the prompt mind map) to reason about the user's request. Then, it relies on FreeCADTool to translate these high-level, BIM-specific instructions into precise commands. When necessary, FreeCADTool invokes the base tools to perform GUI actions, execute scripts, or handle system configurations. This vertical chain, Architect $\rightarrow$ LLM $\rightarrow$ FreeCADTool $\rightarrow$ Base Tools $\rightarrow$ FreeCAD Environment, ensures a clean and logical flow from human intent to model modification. To illustrate this vertical chain with a concrete example, consider the user request: 'Create a wall that is 5 meters long, 3 meters high, and 200mm thick.' This natural language instruction flows through the system as follows:

- The Architect's design intent ('Create a wall...') enters the system at the Natural Language Interface layer.
- In the Reasoning & Domain Knowledge layer:
  - The BIM Knowledge component interprets measurements (5m length, 3m height, 200mm



Figure 2. Mind map visualization of the structured prompt categories and BIM knowledge guiding the LLM. The map illustrates the interconnections between system capabilities, BIM knowledge, quality control, and context templates, providing a rich semantic landscape for interpreting user requests.

thickness) and recognizes 'wall' as a standard BIM element with specific properties.

  - The LLM Reasoning Engine generates a structured plan: 'Create a baseline, extrude to height, apply thickness parameter, and verify orientation.'

- In the Orchestration & Execution layer:
  - The FreeCAD Integration Layer translates the plan into specific API calls:
    ```
    line = Draft.makeLine(FreeCAD.Vector(0,0,0),
    FreeCAD.Vector(5000,0,0))
    wall = Arch.makeWall(line, height=3000,
    width=200)
    ```
  - The system leverages foundation tools like ComputerTool for console input, BashTool for screenshots, and EditTool for file operations.
- In the BIM Environment layer, FreeCAD executes the commands, creating the wall object with proper parameters, updating the model state, and returning visual feedback.

This bidirectional flow allows for iterative refinement, where the LLM can verify outcomes against the original intent, make corrections if needed, and provide explanatory feedback to the user.

4.2 Incorporation of BIM Layer and Prompt Refinements

The key innovation that enables this natural language-driven BIM interaction is the incorporation of a BIM layer through FreeCADTool. This component understands FreeCAD's Python API and the semantics of BIM elements, such as walls, windows, and their associated parameters (dimensions, alignments, placement).

To support this BIM reasoning, the prompt system is refined to include:

- **Domain Knowledge Blocks:** Codified standards and best practices that inform the LLM about typical construction dimensions, unit conversions, and spatial reasoning.
- **Context-Specific Templates:** Ready-made patterns for tasks like wall or window creation, making it straightforward for the LLM to map user instructions to actionable steps.
- **Quality and Error Handling Guidelines:** Instructions for validating the correctness of operations, verifying outcomes (e.g., by inspecting screenshots), and handling exceptions gracefully.

The mind map in Figure 2 visually represents these prompt categories and their interconnections. By embedding this knowledge into the prompt layer, the LLM not only produces commands that conform to BIM standards, but also continuously checks and refines its approach based on feedback.

5 Implementation

The framework's implementation centers around a specialized FreeCADTool that integrates a LLM with FreeCAD's Python API. Operating within a Dockerized Ubuntu 22.04 environment running FreeCAD 0.19, the tool translates structured, BIM-oriented instructions from the LLM into executable Python commands that directly manipulate the FreeCAD model.

At a high level, the system treats each user command as a high-level specification (e.g., "create a 6m x 4m room with walls 3m high and 200mm thick") that the LLM refines into parameter-based instructions. The FreeCADTool then executes these instructions by typing Python code into FreeCAD's console via a ComputerTool—bypassing the complexity of GUI-based, vision-driven navigation and focusing on deterministic, script-based modeling operations. The following shows the key aspects of the implementation:

- **Environment Initialization:** In the first invocation of the tool, it imports the required modules (for example, Arch, Draft) and ensures that a FreeCAD document is active or created. This initialization step establishes a known baseline state for subsequent operations.
- **Parameter-Driven Geometry:** Before issuing commands, the tool converts human-friendly measurements (for example, 6 meters) into FreeCAD's internal millimeter units. This parameterization ensures that all geometry creation (e.g. drawing a baseline line for a wall, invoking 'Arch.makeWall()') is precise and consistent, improving the reliability of the final BIM model.

- **Object Tracking and Reuse:** The tool maintains an internal registry of created objects, mapping generated names (like 'Wall001') to their parameters and creation order. This registry allows subsequent commands, such as placing a window in an existing wall, to refer to previously created elements.
- **Reduced GUI Interaction:** Instead of navigating menus or adjusting workbenches, the tool sends Python commands directly. For instance, creating a wall involves typing 'Draft.makeLine' and 'Arch.makeWall' commands, then adjusting placement and recomputing. This approach reduces complexity and token usage, as the LLM does not need to 'guess' GUI actions; it can rely on a stable script-based interface.
- **Feedback and Verification:** After executing each series of Python commands, the tool reorients the view and captures a screenshot. This image can be returned to the LLM and the user, allowing verification that the outcome of the operation matches the intended design specification.

The BIM Operation Pipeline defines the systematic process of translating natural language BIM requests into model modifications. The algorithm consists of five key phases: Planning, which interprets natural language requests using the BIM knowledge base to extract operational parameters and determine operation types; Preparation, which processes parameters to ensure compatibility with BIM standards, including automatic unit conversion; Command Generation, which creates sequences of FreeCAD API commands based on operation types and validated parameters; Execution, which systematically implements commands through the ComputerTool interface with verification of proper system response; and Validation, which captures the resulting model state and validates operations against intended parameters, refining and iterating if needed. This approach ensures robust handling of BIM operations by combining parameter validation, systematic execution, and result verification in a continuous feedback loop, maintaining model integrity while accommodating natural variability in user requests and BIM constraints.

6 Results

6.1 Comparative Analysis

To evaluate the framework's capabilities, we tested a challenging BIM modeling task on both the original and the proposed frameworks:

Challenge: "Create a rectangular room with four walls. The room should be 6 meters long, 4 meters wide, and have walls that are 3 meters high and 200mm thick."

Expected Outcomes: Four connected walls forming a rectangular room with dimensions of length = 6000mm, width = 4000mm, height = 3000mm, and wall thickness = 200mm, verified by a final screenshot showing the completed room.

The original framework relied on generic tools (ComputerTool, BashTool, EditTool) without BIM-specific knowledge. The LLM resorted to trial-and-error methods: opening FreeCAD via the bash tool, switching to the Part Design workbench, and attempting to create sketches and extrusions through GUI interactions, as shown in Figure 3. This approach was error-prone, consumed large amounts of tokens, and ultimately exceeded the rate limits without producing a coherent final model.

In contrast, the enhanced framework, augmented by the FreeCADTool and BIM-oriented prompt structures, demonstrated a more efficient, parameter-driven workflow. The LLM generated Python commands directly for FreeCAD's Arch and Draft modules, eliminating unnecessary workbench switches and GUI guesswork. This script-based method produced a correctly dimensioned rectangular room with minimal overhead, as demonstrated in Figure 4.

6.2 Performance Analysis

In addition to measuring success by the accuracy of the final BIM model, we analyzed efficiency and resource consumption of both approaches. The original framework performed numerous GUI interactions without domain guidance, leading to high token usage due to repeated instructions and lack of direct scripting capabilities. By contrast, the enhanced framework used structured, parameter-based commands that minimized guesswork and reduced the need for incremental screenshots.

The enhanced framework demonstrated significant improvements in key metrics: 100% dimension accuracy (all measurements within 1mm tolerance) with proper spatial alignment, and command efficiency averaging 4.2 commands per operation with a 92% success rate.

For this proof-of-concept study, validation focused on comparing observable outcomes between frameworks on a representative BIM task. Dimension accuracy was measured against specified requirements (6000mm × 4000mm × 3000mm with 200mm thickness), while command efficiency was calculated as the ratio of successfully executed commands to total attempted commands as logged by FreeCADTool.

Tables 1 and 2 highlight the significant performance differences between the frameworks. Table 1 demonstrates the enhanced framework's superior outcomes across all evaluation criteria, with notable improvements in token usage (40.8% reduction), request count (30.8% decrease), and processing time (16.9% improvement).

Table 1. Comparison of Original and Enhanced Framework Performance

Criteria	Original Framework	Enhanced Framework
BIM Knowledge	None (Generic tools only)	Integrated (FreeCADTool with BIM prompts)
Approach	GUI-based trial-and-error	Parametric Python commands
Token Usage	High (137,273 tokens)	Lower (81,228 tokens)
Number of Requests	13 requests	9 requests
Time to Attempt	Extended (301 seconds)	Reduced (250 seconds)
Final Outcome	Failed to produce room	Successful creation of 4 walls
Dimensions	Not achieved	6000mm × 4000mm × 3000mm met
Wall Thickness	Not applied successfully	Correct 200mm thickness
Final Verification	No coherent model/screenshot	Complete room with screenshot

Table 2. Detailed Performance Metrics Comparison

Metric	Original Framework	Enhanced Framework
Token Consumption Rate	456 tokens/second	325 tokens/second
Average Tokens per Request	10,559 tokens	9,025 tokens
Total Processing Time	301 seconds	250 seconds
Success Rate	0% (failed attempt)	100% (successful completion)
Resource Efficiency	Multiple retry attempts, high token waste	Direct execution, optimized token usage
Screenshot Generation	Multiple intermediate screenshots (>10)	One final verification image (140kB)

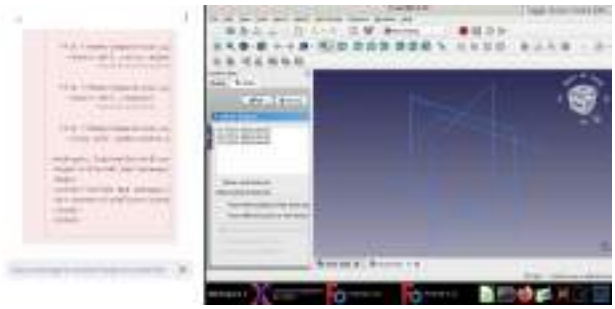
Table 2 further details efficiency gains, showing lower token consumption rates and optimized resource utilization in the enhanced approach.

Given the exploratory nature of this research, these metrics establish basic feasibility rather than comprehensive performance validation. Reducing screenshot requests (each approximately 140 kB) helped limit computational overhead and avoid rate-limit issues.

7 Discussion and Conclusions

This research aimed to demonstrate the feasibility of integrating AI and BIM workflows by developing a proof-of-concept framework that enables natural language interaction with BIM software through LLMs. Our implementation demonstrates this feasibility while also revealing important insights and limitations in the current approach.

Proof-of-Concept Results The prototype framework showed promising initial results in basic BIM operations. The quantitative comparison between frameworks revealed meaningful improvements: 40.8 % reduction in token usage (from 137,273 to 81,228 tokens), 30.8% decrease in required requests (from 13 to 9 requests), and



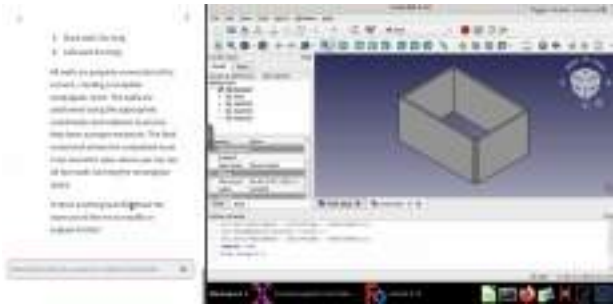


Figure 4. Enhanced framework result showing successful creation of a rectangular room with correct dimensions using parametric BIM commands

16.9% improvement in execution time (from 301 to 250 seconds). While these metrics are encouraging, they represent performance in a highly controlled environment with basic BIM operations.

Limitations Several important limitations emerged during testing. The current framework's capabilities are restricted to basic BIM operations, primarily handling walls and windows, and our testing was confined to a single representative task which limits the generalizability of our findings. Our choice of FreeCAD as the BIM platform, while appropriate for this exploratory proof-of-concept due to its open-source nature and Python scripting capabilities, lacks some advanced features present in industry-standard commercial platforms. Similarly, while Claude provided sufficient LLM capabilities with integrated computer control necessary for our testing methodology, future work should evaluate performance across multiple BIM platforms and LLM technologies as this approach matures beyond the initial feasibility stage. The implementation also shows constraints in handling complex geometric re-

lationships between building elements, while the integration of domain knowledge remains at a preliminary stage. Additionally, the system's requirement for a controlled environment to function properly presents challenges for real-world applications in diverse architectural settings.

Future Work Based on our findings, several directions for future research emerge. Future studies should focus on expanding the range of supported BIM operations and developing more robust domain knowledge integration to handle complex architectural scenarios. Specifically, we plan to test the framework across the following complexity levels:

- **Spatial reasoning challenges:** Testing with non-orthogonal geometries such as curved walls, angled intersections, and sloped roofs to evaluate the LLM's ability to process complex spatial relationships.
- **Multi-element compositions:** Evaluating performance on tasks requiring coordination between multiple building components (e.g., placing windows and doors with proper alignments to walls, creating multi-story structures with staircases).
- **Parameter-driven modifications:** Testing the system's ability to modify existing elements based on design constraints, such as adjusting wall thickness based on thermal requirements or modifying room dimensions while maintaining proper connections.
- **Error detection and recovery:** Assessing how the framework handles conflicting specifications and its ability to propose alternative solutions when design conflicts arise.

As the framework matures with these expanded capabilities, a critical next step will be designing appropriate evaluation methods to assess its practical utility in professional contexts. This would involve developing metrics that consider both technical performance and potential integration points with human workflows. Rather than positioning AI as a replacement for human expertise, future research should explore how this technology can best complement and enhance the capabilities of AEC professionals at specific stages of the design process.

Additionally, testing should be extended to encompass a broader range of architectural tasks to validate the framework's applicability across different design contexts, including residential, commercial, and specialized building types. Further investigation is needed to explore integration possibilities with industry-standard BIM platforms beyond FreeCAD, such as Revit and ArchiCAD, to assess the framework's adaptability and identify platform-specific challenges. Development of more sophisticated methods to handle complex geometric relationships and parameters will be crucial for practical applications in professional environments. These advancements would contribute to creating a more versatile and practically applicable system for architectural design automation that could significantly reduce technical barriers for AEC professionals.

Final Remarks This study contributes to the field by demonstrating how LLMs can integrate with BIM workflows, leveraging natural language processing to simplify complex BIM operations. Our proof-of-concept establishes the foundational feasibility of LLM-driven BIM automation and suggests a pathway toward more intuitive interfaces that could reduce technical barriers for AEC professionals who lack programming expertise.

To realize practical benefits, further development is needed in several critical areas: integration with industry-standard BIM platforms, expansion of comprehensive domain knowledge bases, and support for complex architectural operations. While our prototype demonstrates basic feasibility, significant challenges remain in domain knowledge integration, geometric reasoning, and error handling before such systems could be deployed professionally. Future research should validate these approaches in real-world scenarios and develop frameworks capable of handling the full complexity of professional BIM workflows.

References

[1] Y. Izbash and V. Babayev. Digital evolution in aec industry: Bridging bim, building codes, and future technologies. In *IOP Conference Series: Earth and Environmental Science*. Institute of Physics, 2024. doi:10.1088/1755-1315/1376/1/012004.

[2] R. Sacks, M. Girolami, and I. Brilakis. Building information modelling, artificial intelligence and construction tech. *Developments in the Built Environment*, 4, 2020. doi:10.1016/j.dibe.2020.100011.

[3] A. H. Qureshi, W. S. Alaloul, B. Manzoor, M. A. Musarat, S. Saad, and S. Ammad. Implications of machine learning integrated technologies for construction progress detection under industry 4.0 (ir 4.0). In *2020 2nd International Sustainability and Resilience Conference*. IEEE, 2020. doi:10.1109/IEEECONF51154.2020.9319974.

[4] S. Javaid, H. Fahim, B. He, and N. Saeed. Large language models for uavs: Current state and pathways to the future. *IEEE Open Journal of Vehicular Technology*, 2024. doi:10.1109/OJVT.2024.3446799.

[5] B. Bölek, O. Tutal, and H. Özbaşaran. A systematic review on artificial intelligence applications in architecture. *Journal of Design for Resilience in Architecture and Planning*, 4(1):91–104, 2023. doi:10.47818/drarch.2023.v4i1085.

[6] F. M. Dinis, J. Poças Martins, A. S. Guimarães, and B. Rangel. Bim and semantic enrichment methods and applications: A review of recent developments. *Archives of Computational Methods in Engineering*, 2022. doi:10.1007/s11831-021-09595-6.

[7] R. Syed, S. Suriadi, M. Adams, W. Bandara, S. J. J. Leemans, C. Ouyang, A. H. M. ter Hofstede, I. van de Weerd, M. T. Wynn, and H. A. Reijers. Robotic process automation: Contemporary themes and challenges. *Computers in Industry*, 115, 2020. doi:10.1016/j.compind.2019.103162.

[8] FreeCAD Team. Freecad: Your own 3d parametric modeler, 2024. URL `https://www.freecad.org/`. Accessed: 2024-12-10.

[9] Z. Liu, Y. Lu, and L. C. Peh. A review and scientometric analysis of global building information modeling (bim) research in the architecture, engineering and construction (aec) industry. *Buildings*, 9(10), 2019. doi:10.3390/buildings9100210.

[10] N. Gu and K. London. Understanding and facilitating bim adoption in the aec industry. *Automation in Construction*, 19(8):988–999, 2010. doi:10.1016/j.autcon.2010.09.002.

[11] M. Thomassen. Bim & collaboration in the aec industry. *Construction Management and Engineering*, 2011. URL `http://www.m-tech.aau.dk/`.

[12] S. Azhar. Building information modeling (bim): Trends, benefits, risks, and challenges for the aec industry. *Leadership and Management in Engineering*, 2011.

[13] A. Criminale and S. Langar. Challenges with bim implementation: A review of literature. *ASC Annual International Conference Proceedings*, 53, 2017.

[14] E. Forcael, I. Ferrari, A. Opazo-Vega, and J. A. Pulido-Arcas. Construction 4.0: A literature review. *Sustainability*, 12(22), 2020. doi:10.3390/su12229755.

[15] A. Zaland, S. Saad, K. Rasheed, and S. Ammad. Automation in construction industry. *Construction Management Review*, 2024.

[16] G. Rocha and L. Mateus. Using dynamo for automatic reconstruction of bim elements from point clouds. *Applied Sciences*, 14(10), 2024. doi:10.3390/app14104078.

[17] S. Matarneh, F. Elghaish, F. P. Rahimian, N. Dawood, and D. Edwards. Automated and interconnected facility management system: An open ifc cloud-based bim solution. *Automation in Construction*, 143, 2022. doi:10.1016/j.autcon.2022.104569.

[18] A. O. Akponeware and Z. A. Adamu. Clash detection or clash avoidance? an investigation into coordination problems in 3d bim. *Buildings*, 7(3), 2017. doi:10.3390/buildings7030075.

[19] H. Nabizadeh Rafsanjani and A. H. Nabizadeh. Towards human-centered artificial intelligence (ai) in architecture, engineering, and construction (aec) industry. *Computers in Human Behavior Reports*, 2023. doi:10.1016/j.chbr.2023.100319.

[20] Z. X. Chew, J. Y. Wong, Y. H. Tang, C. C. Yip, and T. Maul. Generative design in the built environment. *Automation in Construction*, 2024. doi:10.1016/j.autcon.2024.105638.

[21] J. Ko, B. Ennemoser, W. Yoo, W. Yan, and M. J. Clayton. Architectural spatial layout planning using artificial intelligence. *Automation in Construction*, 2023. doi:10.1016/j.autcon.2023.105019.

[22] F. Chen, M. Mai, X. Huang, and Y. Li. Enhancing the sustainability of ai technology in architectural design: Improving the matching accuracy of chinese-style buildings. *Sustainability*, 16(19), 2024. doi:10.3390/su16198414.

[23] Ghang Lee, Suhyung Jang, and Seokho Hyun. A generalized llm-augmented bim framework: Application to a speech-to-bim system, 2024. URL `https://arxiv.org/abs/2409.18345`.

[24] S. Dai, Y. Li, K. Grace, and A. Globa. Towards human-ai collaborative architectural concept design via semantic ai. In *Communications in Computer and Information Science*, pages 68–82. Springer, 2023. doi:10.1007/978-3-031-37189-9_5.

[25] M. Vössing, N. Kühl, M. Lind, and G. Satzger. Designing transparency for effective human-ai collaboration. *Information Systems Frontiers*, 24(3):877–895, 2022. doi:10.1007/s10796-022-10284-3.

[26] Anthropic Team. Claude computer use, 2024. URL `https://docs.anthropic.com/en/docs/build-with-claude/computer-use`. Accessed: 2024-12-10.

Non-Productive Time Reduction in Off-Site Construction: A Predictive Analytics Approach Using Deep Learning

Xue Chen[1], Weining Zeng[1], Runcong Liu[1], Ahmed Bouferguene[2], Mohamed Al-Hussein[1]

[1]Department of Civil and Environmental Engineering, University of Alberta, Edmonton, Alberta, Canada
[2]Campus Saint-Jean, University of Alberta, Edmonton, Alberta, Canada
xue15@ualberta.ca, weining3@ualberta.ca, runcong@ualberta.ca, ahmedb@ualberta.ca, malhussein@ualberta.ca

Abstract –

This study proposes a data-driven framework to reduce Non-Productive Time (NPT) in Off-Site Construction (OSC) by integrating predictive modeling with lean principles. Through comprehensive computer vision (CV) techniques, large-scale video data are analyzed to identify key operational states, worker positions, and material flows. Two primary classification approaches—a Random Forest (RF) model and a Deep Learning (DL) network—are employed to detect NPT episodes. Experimental results reveal that the DL model achieves substantially higher accuracy and F1 scores, demonstrating its capability to handle complex spatial–temporal interactions. Feature-weight analysis further indicates that task transitions, assembly processes, and coordinate-based congestion zones are critical drivers of NPT. Building on these insights, targeted lean interventions—such as just-in-time material delivery, proactive maintenance, and optimized workstation layouts—can streamline operations, minimize idle intervals, and enhance overall productivity. The conclusions underscore the effectiveness of merging data analytics with lean strategies in uncovering hidden inefficiencies within OSC, providing a clear direction for future research and industry adoption.

Keywords –

Non-Productive Time, Off-Site Construction, Deep Learning, Lean Principles, Predictive Analytics

1 Introduction

As the construction industry increasingly prioritizes efficiency and quality, OSC has emerged as a prominent solution due to its advantages in factory prefabrication and rapid on-site assembly. These features not only shorten construction timelines and enhance quality control but also significantly reduce reliance on on-site labor and environmental impact [1,2]. Building on this foundation, lean construction emphasizes systematic workflow management and continuous improvement to minimize waste and maximize value-adding activities, thereby improving overall productivity [3]. However, despite its advantages, practical challenges such as delays, equipment idling, and communication inefficiencies often result in NPT during manufacturing and assembly processes, which substantially undermine the potential benefits of OSC [4,5].

Recent research trends highlight the application of CV and artificial intelligence (AI) techniques in construction settings as a promising means to efficiently detect, track, and interpret the behavior of workers, machinery, and materials. These technologies enable comprehensive monitoring of OSC workflows and the identification of non-value-adding activities [6,7]. Compared to traditional manual observations or documentation, vision-based systems can process large-scale image and video data in significantly less time, thereby achieving greater precision in identifying waste [8]. For instance, some studies have employed trajectory analysis of workers and machinery to pinpoint repetitive idle zones, providing actionable insights for optimizing workflow layouts and material transport. Others have utilized convolutional neural networks (CNNs) or enhanced You Only Look Once (YOLO) models to rapidly detect prefabricated components and issue real-time alerts for anomalies or defects in production processes [9,10]. Additionally, advanced techniques such as semantic segmentation and pose estimation have facilitated the precise classification and quantification of work behaviors, offering enriched data support for productivity enhancement and safety management [11].

In the context of productivity management, automated monitoring and data-driven analytics are being increasingly integrated into OSC practices. Numerous studies have focused on using vision-based detection and tracking to monitor crane operations, worker pathways, and component statuses. These datasets are often synchronized with BIM models, digital twins, or project schedules to identify deviations from planned progress and supply chain inefficiencies [12,13].

However, much of the existing literature centers on quantifying discrepancies between actual progress and planned schedules, with limited emphasis on systematically identifying the root causes of process delays and resource inefficiencies [14]. Integrating the principles of lean construction—specifically its focus on reducing "process waste" and driving "continuous improvement" — with the automated detection and predictive analytics capabilities of CV offers a pathway to achieve significant breakthroughs in reducing NPT and enhancing construction efficiency [15].

In summary, this study builds on the high standardization and modularity characteristics of OSC by integrating lean construction principles with CV technologies to automatically identify and quantify non-productive time. The paper begins with a review of advancements in the application of OSC, lean construction, and vision algorithms in the construction sector. Subsequently, a DL model is developed using real-world data to identify and predict NPT in OSC processes. The findings provide empirical evidence for both academia and industry practitioners seeking to enhance productivity in OSC under the backdrop of digitization and automation.

2 Methodology

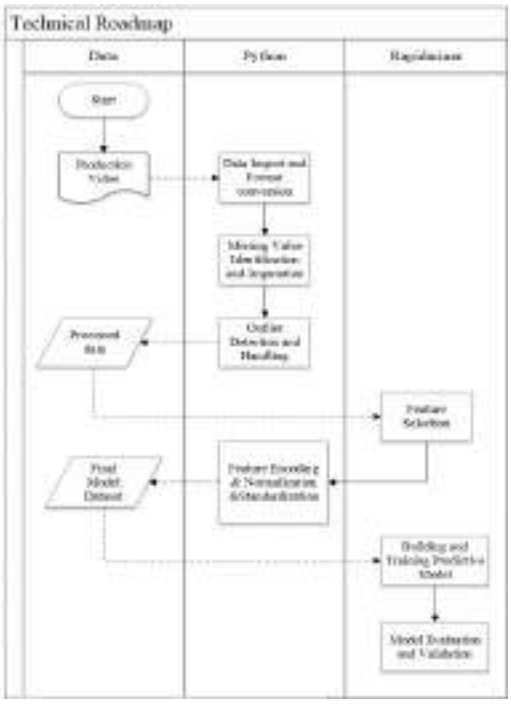


Figure 1. Technical Roadmap for NPT Prediction of CV Acquired Data Using DL model

Figure 1 shows the technical path for NPT prediction using DL. It consists of the following seven components:

1. Data Import and Format Conversion: Diverse data formats, including CSV and Excel files, are loaded and merged into consistent tables to enable unified analysis. YOLO outputs, which have undergone pre-processing, are incorporated at this stage to ensure compatibility with the analysis pipeline.
2. Missing Value Identification and Insertion: Context-driven imputation methods address missing values by filling short vacant intervals with information from adjacent frames. This strategy helps maintain data continuity while minimizing the risk of introducing artificial trends.
3. Outlier Detection and Handling: Outliers, such as sudden coordinate jumps or unrealistic worker counts, are detected and removed using custom algorithms. Thresholds for identifying anomalies are visually validated with graphical tools to ensure robustness.
4. Feature Selection: The influence of each attribute on NPT is assessed through statistical techniques such as ANOVA and Logistic Regression. Attributes with minimal contribution are excluded to optimize the predictive feature set.
5. Feature Encoding, Normalization, and Standardization: Numerical attributes are standardized using Min-Max scaling or Z-score normalization to achieve uniform magnitudes. Categorical variables, such as "work conditions," are encoded using One-Hot encoding to eliminate potential issues with implicit ordering.
6. Build and Train Predictive Models: Processed features are used to train classifiers, including RF and DL. Iterative tuning of hyperparameters, such as tree depth and layer size, is conducted to enhance model performance and accuracy.
7. Model Evaluation and Validation: Model performance is validated through multiple holdout experiments, evaluating metrics such as Recall, F1 score, and AUC-PR to address class imbalance. Iterative improvements are made until the model meets predefined performance targets, ensuring reliability and robustness.

3 Implemented Details of case study

3.1 Data Acquisition Process and Database

3.1.1 Camera Deployment and Data Collection

The research team strategically installed cameras within an off-site construction factory that specializes in manufacturing prefabricated residential building components. The collected videos focused on a specific workstation dedicated to producing wooden floors with an example shown in Figure 2. Videos were recorded at

1080 × 1920 pixels across multiple production cycles, capturing diverse working conditions and work resources (such as workers, materials, and machines).



Figure 2. Screenshot of Video

3.1.2 Object Detection and Data Acquisition

The methods for tracking, detecting, and identifying objects align with those proposed by Chen, with extensions in coverage and quantity [16]. After video collection, the YOLO V8 algorithm was employed to detect objects such as workers, machinery, and material racks, generating bounding box coordinates for each frame. ByteTrack was then used to assign and maintain tracking IDs, ensuring object consistency across frames. Spatial density analysis facilitated the assessment of worker positions relative to workstations, enabling the identification of potential idle times.

With the vision-based analysis, non-productive time, including waiting time for machine, waiting time for worker, waiting time for material, empty time of a workstation, and delayed movement of completed products, are automatically identified from off-site construction processes. For example, waiting time for machine, materials and workers are defined as the time difference between the actual acquisition time of the required resource (observed from the object detection results) and the expected timepoint to obtain this resource (predicted with the activity recognition results since it's usually the endpoint of the preceding activity). The specific non-productive time identification algorithms are developed in the author's previous work [16]. Using the outputs from the computer vision models, the following sections work on the data-driven non-productive time analysis and prediction with predictive analytics approaches.

3.1.3 Data Structure

The video dataset comprised 19,828 frames, later expanded into 50,450 records. The final dataset includes 11 columns, encompassing numerical, categorical, and Boolean data types. Table 1 provides an overview:

Table 1. Overview of Dataset

Column Name	Data Type	Completeness
Floor ID_x	Integer	No missing values
Video Frame ID	Integer	No missing values
Original Worker Count	Integer	No missing values
Real Number of Workers	Integer	No missing values
Work Status	Categorical	No missing values
Current Time	Floating-point	No missing values
Non-productive Time	Boolean	No missing values
Floor ID_y	Integer	50446 of 50450
X_final	Floating-point	50446 of 50450
Y_final	Floating-point	50446 of 50450
Class	Floating-point	50446 of 50450

3.2 NPT Identification and Screening

3.2.1 Initial Rules and Manual Annotation

A subset of data samples was initially reviewed and manually labeled to identify typical instances of "idle workers" or "unattended workstations." This step involved defining key thresholds for "prolonged inactivity" or "unmanned stations," along with minimum durations and corresponding workstations.

3.2.2 Rule Automation and Boundary Validation

The rules derived from manual annotation were then automated and applied to the entire dataset. For edge cases, such as brief task switches or discontinuous video frames, manual re-validation was performed to ensure the accuracy of NPT labels. Periods with active workers or operational equipment were marked as productive, while others were classified as NPT and stored as Boolean variables in the database.

3.3 Data Cleaning

3.3.1 Handling Missing Values

Coordinate and work status gaps: A context-aware imputation method was applied. For example, if a worker temporarily moved out of the camera's field of view, their coordinates were estimated using adjacent frames. Similarly, during task transitions where labels were briefly unidentifiable, transitional states were inferred based on operational sequences.

This approach minimized distortions from blind interpolation and reduced misjudgments of spatial-

temporal relationships [17].

3.3.2 Removing Outliers

Threshold detection was applied to worker counts and coordinates. Records were flagged and removed if coordinate shifts exceeded a predefined max value and did not align with workstation entry zones.

For static objects, such as machinery or material racks, abrupt positional changes were cross verified against predefined spatial boundaries and cleaned accordingly. This process adhered to the physical layout of the factory and the logical continuity of OSC activities, effectively eliminating noise without compromising valid samples.

3.4 Feature Selection and Processing

3.4.1 Logistic Regression and ANOVA Integration

Logistic regression was used to analyze the relationship between features and the target variable, NPT, while ANOVA assessed whether numerical attributes exhibited significant differences between productive and non-productive periods ($p<0.05$). The results highlighted five significant attributes- "number of workers," "work status," "X_final," "Y_final," and "class"-while "current time" showed weak correlation and was excluded. Table 2 shows the results of the individual correlation analyses (W stands for work conditions in attribute.). Table 3 shows the results of the correlation analysis for the ANOVA matrix model.

Table 2. Correlation analyzing results for logistic regression model

Attribute	Coefficient	Std. Coefficient	Weight
W: placing joist	-0.56	-0.56	-0.56
W: connecting joist	0.35	0.35	0.35
W: placing panel	0	0	0
W: connecting joist and panel	0	0	0
W: lifting the completed panel	-0.057	-0.057	-0.057
W: cleaning the workstation	-0.316	-0.316	-0.316
W: preparing for the new work	2.153	2.153	2.153
W: idle	0	0	0
Number of workers	-0.337	-0.303	-0.303
Current time	0	0.306	0.306
x_final	0	0	0
y_final	0.001	0.041	0.041
Class	0.168	0.164	0.164
Intercept	-3.452	-1.942	-

Table 3. Correlation analyzing results for ANOVA matrix model

ANOVA Attribute	Group non-productive time
x_final	0
y_final	0.004
class	0

3.4.2 Normalization and Encoding

Number of workers: Min–Max normalization was applied to facilitate DL processing.

Coordinates (X_final, Y_final): Z-score standardization was used to mitigate scale differences and outlier impacts.

Work status and class: One-Hot encoding converted these attributes into binary indicators, avoiding the introduction of pseudo-order relationships.

The final dataset comprised 17 processed features, enabling robust predictive modeling.

3.5 Predictive Model Development

3.5.1 Data Partitioning and Model Selection

The structured dataset, created through data cleaning, feature engineering, and encoding, was divided into training and validation subsets to ensure unbiased evaluation. RF and DL were selected to address both linear and non-linear relationships with the target variable, NPT.

RF: An ensemble of decision trees trained on bootstrapped samples and random feature subsets, RF enhances generalization and mitigates overfitting. It also provides feature importance insights for NPT prediction.

DL leverage flexible architectures with hidden layers and non-linear activation functions, such as ReLU, to capture complex feature interactions. Regularization techniques, such as dropout and limited epochs, prevent overfitting and ensure computational efficiency and repeatability.

Random forest: Key hyperparameters such as the number and depth of trees were optimized.

DL: The focus was on tuning hidden layers and learning rates. If performance metrics (e.g., Recall, F1, AUC-PR) were unsatisfactory, feature adjustments or model structure refinements were iteratively performed.

3.5.2 Iterative and Validation

Hyperparameters were iteratively tuned to optimize both models:

- RF: Parameters like the number of trees and maximum depth were adjusted.
- DL: Network depth, neuron count, and learning rate were fine-tuned, with adaptive learning rates ensuring stable convergence.

Validation used a multiple holdout strategy, averaging performance metrics (recall, F1 score, and AUC-PR) across several data splits. This robust evaluation minimized bias and ensured generalizability to unseen data.

4 Results and discussions

4.1 Overall Performance and Confusion Matrices

This study employs both RF and DL models to classify NPT. Table 4Table 5Table 6 sequentially present the confusion matrices of the two models and their key evaluation metrics.

Table 4. Deep Learning Model Confusion Matrix

	True FALSE	**True True**	**Class precision**
Pred FALSE	9119	2154	80.89%
Pred Ture	4106	11081	72.96%
Class recall	68.95%	83.72%	

Table 5. Random Forest model confusion matrix

	True FALSE	**True True**	**Class precision**
Pred FALSE	5424	2033	72.74%
Pred Ture	7806	11197	58.92%
Class recall	41.00%	84.63%	

Table 6. Comparison of key indicators of different models

	Deep Learning	**Random Forest**
Accuracy	76.34% ± 0.16%	62.82% ± 0.16%
Classification Error	23.66% ± 0.16%	37.18% ± 0.16%
F1	77.97% ± 0.13%	69.48% ± 0.13%
Sensitivity (Ture)	83.72% ± 0.21%	84.63% ± 0.16%

With the construction data collected from off-site construction factories, the performances of both DL model and RF model as well as their underlying structural reasons are validated and compared in actual construction cases.

1. Accuracy: The DL model achieves an accuracy of 76.34%, which is over 13% higher than Random Forest's 62.82%. This indicates that the DL model is more precise in distinguishing between productive and non-productive states under complex working conditions. The relatively lower accuracy of RF can be attributed to its limitations in handling complex decision boundaries, particularly when distinguishing short pauses from genuine NPT, where RF tends to oversimplify feature interactions.
2. Classification Error: The DL model's classification error is 23.66%, significantly lower than RF's 37.18%, demonstrating that DL is more balanced in classifying negative (productive) and positive (NPT) classes. RF's high error rate is mainly due to its low Specificity (only 41%), implying that many productive cases (FALSE) are misclassified as NPT (TURE), potentially causing unnecessary disruptions in practical scenarios.
3. F_measure: F_measure combines Precision and Recall, serving as a crucial indicator of model performance in real-world settings. The DL model's F_measure is 77.97%, surpassing RF's 69.48%, reflecting that DL not only accurately identifies NPT but also significantly reduces false alarms. RF's lower F_measure indicates a higher misclassification rate in productive states, undermining its practical utility.
4. Recall: For the positive class (NPT), both RF and DL exhibit relatively high recall rates, 84.63% and 83.72% respectively. While RF's Sensitivity is slightly higher, its overall effectiveness remains limited due to low Specificity.

Compared to the RF model, the good performance of the DL model with an accuracy of 76.34% and a recall of 83.72% indicates its capacity in predicting non-productive time for off-site construction processes based on the given information such as current work activities, worker locations and current time. Using the prediction results, adaptive measures in construction management plans can be developed in advance, thereby mitigating construction delays, and ensuring on-time project delivery.

4.2 Analysis of Model Characteristics

RF splits features using multiple decision trees and performs well for single-dimension or discrete variables. However, when facing high-dimensional, nonlinear, or

spatiotemporal interactions (e.g., worker coordinates over time, complex assembly processes), insufficient tree depth or limited feature engineering can hamper its ability to capture subtle, dynamic patterns.

DL leverages multi-layer network architectures to learn higher-order nonlinear relationships, which proves advantageous for modeling worker trajectories, equipment interactions, and process coordination. In scenarios involving workflow transitions or multi-person collaboration, the representational power of DL leads to more balanced, stable, and accurate classification in OSC contexts characterized by multi-station parallel production and high spatiotemporal complexity.

4.3 Feature Weight Discussion

4.3.1 Significance of High-Weight Features

'Preparing for the new work' (highest weight, 0.263): When the DL model detects a transition or preparation phase, it becomes more inclined to classify the process as "non-productive or inefficient." This is because waiting and non-value-added activities (e.g., unprepared materials or unfamiliar workflow sequences) can easily occur [1]. In this study, delayed start times or repeated waiting within such periods often lead to NPT classification.

'Connecting joist and panel' & 'Connecting joist' (weights - 0.179, 0.156): These tasks involve more intricate or tedious assembly processes requiring multi-person coordination, alignment, cutting equipment, and fasteners. Any delay or error in one step can halt the entire workstation. The deep network captures the high complexity of these phases, thus quickly labeling them as NPT upon detecting anomalies (e.g., lack of materials, worker misalignment). •

'Coordinate Position' (x_final_normalized, 0.131): Conventional logistic regression or ANOVA analyses may not consistently highlight the significance of spatial coordinates. Nonetheless, in multi-layer network learning, coordinate data can indicate congestion or interference in specific areas—e.g., storage zones, material stacking areas, or transfer pathways. The model's hidden layers may capture an implicit pattern that "frequent stays in certain locations correlate with higher NPT probability."

'Class_0.0' (One-hot encoding, 0.107): Different object types (e.g., worker, machine, material rack) can influence the likelihood of pauses in a working environment:

4.3.2 Possible Reasons for Medium-Low Weight Features

'Number_of_workers_normalized' (0.097): Although the match between labor availability and task demands is often considered crucial, it did not rank among the highest-weight features in this study. Potential reasons include:

In scenarios where complex procedures (such as connection or prep work) become bottlenecks, the effect of workforce size diminishes.

In OSC production settings, workforce numbers typically remain fairly constant, with significant variations appearing only under extreme conditions.

'Idle, cleaning the workstation, etc.' (low-weight features): Such auxiliary or short-duration tasks comprise a minor share of total production time. They are usually simpler in nature and do not necessarily trigger major waiting periods. The DL model is more attentive to contexts such as "workflow transitions" or "insufficient material/equipment readiness," which have a greater impact on overall efficiency. Consequently, these auxiliary states are assigned lower weights.

4.3.3 Adaptability for effective deployment of model

In addition to the application scenarios in this study, the framework in this study also has the potential to help capture and predict waiting and idle time in the field of steel bar processing, where multi-camera monitoring combined with deep learning recognition can help locate snags in time, such as during steel bar cutting, bending, welding or bundling. At the same time, it also has excellent potential for producing precast concrete components, especially the fabrication and assembly of steel cages. It can collect real-time interaction data on personnel, materials and equipment at workstations and identify and alarm for lengthy waiting or cumbersome operations during bundling, positioning and welding.

Extending the framework to a new environment faces some predictable technical and data challenges. First, to achieve complete monitoring in a larger or more irregular space, the number of cameras needs to be increased, and a multi-angle layout should be adopted to reduce visual blind spots and significantly overlapping views that help solve occlusion problems. If applied to outdoor construction environments, additional factors can be introduced to complement visual data so workers and equipment can be accurately tracked in light and weather conditions. Second, to maintain the prediction model's accuracy in various scenarios, it is necessary to train it using a richer dataset that covers more workflows, combined with algorithm enhancements such as deeper network structures or attention mechanisms to adapt to complex temporal and spatial changes.

5 Conclusion and limitations

5.1 Conclusion

This study combines predictive modeling with lean principles to effectively reduce NPT in OSC. Leveraging multi-dimensional data, including task status, object location, and production conditions, the developed model—especially the DL—demonstrates high precision and accuracy in detecting and quantifying periods of non-productive activities. The findings not only illustrate the impact of various task stages (e.g., "preparing for new work" or "connecting joists and panels") on overall productivity but also provide a basis for implementing targeted process improvements.

Compared to traditional manual observations or single-metric evaluations, the data-driven approach presented here more accurately captures potential efficiency gaps and helps decision-makers focus on high-impact scenarios. For instance, the model's ability to identify such issues as insufficient material preparation, improper assembly coordination, and delayed maintenance directly supports the adoption of JIT material delivery and preventive maintenance. By intervening at key points such as task transition, possible measures (such as dynamic job scheduling and resource allocation) can be explored in future work to effectively reduce idle time and resource waste while enhancing worker satisfaction and safety. Furthermore, integrating these predictive insights with lean principles enables continuous evaluation and optimization of production workflows. As data accumulates and feature engineering becomes more refined, the DL recall is expected to further improve, capturing a broader scope of inefficiencies and forming a "data–improvement–feedback" loop that drives production systems toward greater agility, efficiency, and human-centric operation.

5.2 Limitations

While the study primarily focuses on key variables such as work status, coordinate positions, and basic staffing information, the current set of features does not fully capture the intricate spatial–task interactions inherent in OSC. For example, more detailed demarcations of "material storage" versus "assembly" zones or richer descriptors of specialized tasks (e.g., more granular role assignments or real-time equipment load data) could offer deeper insights into production bottlenecks.

Although the DL outperforms traditional methods in detecting NPT, it has not yet attained an ideal recall level. Certain rare or brief forms of inefficiency may remain undetected. Future research should explore more sophisticated architectures (e.g., temporal modeling or attention mechanisms) and a tighter integration of spatiotemporal features to improve the model's ability to capture elusive inefficiency events.

The framework's validation primarily focuses on standardized prefabrication workflows, particularly the panelized wood floor production process, which includes material transport, joist and panel layout, mechanical fastening, and hoisting operations. Workers and machines responsible for placing floor joist and panel, nailing and cutting, and lifting the completed building components are monitored with surveillance camera. This off-site construction scenario can extend to other construction processes such as steel and concrete prefabrication or curtain wall fabrication, as well as repetitive on-site tasks including mechanical, electrical, and plumbing assembly or rebar cage tying. Further implementation may require adaptations for each specific task since every new application scenario demands customized camera placements suited to the spatial workflows at hand, as well as model retraining with domain-specific data.

References

[1] M. R. Hosseini, I. Martek, E. K. Zavadskas, A. A. Aibinu, M. Arashpour, and N. Chileshe, "Critical evaluation of off-site construction research: A Scientometric analysis," *Autom. Constr.*, vol. 87, pp. 235–247, Mar. 2018, doi: 10.1016/j.autcon.2017.12.002.

[2] Z. Liu, B.-G. Hwang, and J. M. Lim, "Prefabricated and Prefinished Volumetric Construction: Assessing Implementation Status, Perceived Benefits, and Critical Risk Factors in the Singapore Built Environment Sector," *J. Manag. Eng.*, vol. 39, no. 6, p. 04023045, Nov. 2023, doi: 10.1061/JMENEA.MEENG-5455.

[3] A. Hadi, F. Cheung, S. Adjei, and A. Dulaimi, "Evaluation of Lean Off-Site Construction Literature through the Lens of Industry 4.0 and 5.0," *J. Constr. Eng. Manag.*, vol. 149, no. 12, p. 03123007, Dec. 2023, doi: 10.1061/JCEMD4.COENG-13622.

[4] I. Taifa and T. Vhora, "Cycle time reduction for productivity improvement in the manufacturing industry," *J. Ind. Eng. Manag. Stud.*, vol. 6, no. 2, pp. 147–164, Dec. 2019, doi: 10.22116/jiems.2019.93495.

[5] B. Barkokebas, P. Martinez, A. Bouferguene, F. Hamzeh, and M. Al-Hussein, "Digitalization-based process improvement and decision-making in offsite construction," *Autom. Constr.*, vol. 155, p. 105052, Nov. 2023, doi: 10.1016/j.autcon.2023.105052.

[6] A. Lekan, A. Clinton, E. Stella, E. Moses, and O. Biodun, "Construction 4.0 Application: Industry 4.0, Internet of Things and Lean Construction Tools' Application in Quality Management System of Residential Building Projects," *Buildings*, vol. 12, no. 10, Art. no. 10, Oct. 2022, doi: 10.3390/buildings12101557.

[7] H. M. Dara, A. Raut, M. Adamu, Y. E. Ibrahim, and P. V. Ingle, "Reducing non-value added (NVA) activities through lean tools for the precast industry," *Heliyon*, vol. 10, no. 7, p. e29148, Apr. 2024, doi: 10.1016/j.heliyon.2024.e29148.
[8] M. S. Bajjou and A. Chafi, "Lean construction and simulation for performance improvement: a case study of reinforcement process," *Int. J. Product. Perform. Manag.*, vol. 70, no. 2, pp. 459–487, Jul. 2020, doi: 10.1108/IJPPM-06-2019-0309.
[9] D. Roberts and M. Golparvar-Fard, "End-to-end vision-based detection, tracking and activity analysis of earthmoving equipment filmed at ground level," *Autom. Constr.*, vol. 105, p. 102811, Sep. 2019, doi: 10.1016/j.autcon.2019.04.006.
[10] H. Luo, C. Xiong, W. Fang, P. E. D. Love, B. Zhang, and X. Ouyang, "Convolutional neural networks: Computer vision-based workforce activity assessment in construction," *Autom. Constr.*, vol. 94, pp. 282–289, Oct. 2018, doi: 10.1016/j.autcon.2018.06.007.
[11] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2014, pp. 580–587. doi: 10.1109/CVPR.2014.81.
[12] S. Wang, L. Deng, J. Guo, M. Liu, and R. Cao, "Automatic Quality Inspection of Rebar Spacing Using Vision-Based Deep Learning with RGBD Camera," presented at the 41st International Symposium on Automation and Robotics in Construction, Lille, France, Jun. 2024. doi: 10.22260/ISARC2024/0009.
[13] A. Braun, S. Tuttas, A. Borrmann, and U. Stilla, "Improving progress monitoring by fusing point clouds, semantic data and computer vision," *Autom. Constr.*, vol. 116, p. 103210, Aug. 2020, doi: 10.1016/j.autcon.2020.103210.
[14] M. Arashpour, R. Wakefield, N. Blismas, and E. W. M. Lee, "Analysis of Disruptions Caused by Construction Field Rework on Productivity in Residential Projects," *J. Constr. Eng. Manag.*, vol. 140, no. 2, p. 04013053, Feb. 2014, doi: 10.1061/(ASCE)CO.1943-7862.0000804.
[15] R. Pozzi, V. G. Cannas, and T. Rossi, "Data Science Supporting Lean Production: Evidence from Manufacturing Companies," *Systems*, vol. 12, no. 3, Art. no. 3, Mar. 2024, doi: 10.3390/systems12030100.
[16] X. Chen, Y. Wang, J. Wang, A. Bouferguene, and M. Al-Hussein, "Vision-based real-time process monitoring and problem feedback for productivity-oriented analysis in off-site construction," *Autom. Constr.*, vol. 162, p. 105389, Jun. 2024, doi: 10.1016/j.autcon.2024.105389.
[17] H. Lee and S. Yun, "Strategies for Imputing Missing Values and Removing Outliers in the Dataset for Machine Learning-Based Construction Cost Prediction," *Buildings*, vol. 14, no. 4, p. 933, Mar. 2024, doi: 10.3390/buildings14040933.

Interpretable Machine Learning for Predicting the Fatigue Strength of Steel: Influence of Composition and Processing Parameters

Soheila Kookalani[1], Gozde Basak Ozturk[2], Hamidreza Alavi[1], Erika Parn[1], and Ioannis Brilakis[1]

[1]Department of Engineering, University of Cambridge, Cambridge, UK
[2]Department of Civil Engineering, Aydin Adnan Menderes University, Aydin, Turkey
sk2268@cam.ac.uk, gbozturk@edu.edu.tr, sa2194@cam.ac.uk, eap47@cam.ac.uk, ib340@cam.ac.uk

Abstract –

The prediction of fatigue strength in steel is critical for the design and analysis of structural components, given the high costs and time associated with fatigue testing and the severe consequences of fatigue failures. This study explores the application of the CatBoost machine learning algorithm to predict the fatigue strength of steel based on chemical composition, processing parameters, and mechanical properties. The model's interpretability is enhanced by integrating Shapley additive explanations, providing insights into the contributions of key features such as tempering temperature (TT), chromium content (Cr), and molybdenum content (Mo). The proposed framework achieves high predictive accuracy, with an R^2 of 0.952 and an RMSE of 31.625. This study fosters trust and utility in safety-critical applications by addressing the limitations of black-box models. The results underscore the potential of interpretable machine learning in advancing fatigue strength prediction methodologies and informing structural and construction engineering practices, while Enhancing the SHAP-based feature importance analysis to refine the selection of key predictors, potentially simplifying the model while maintaining accuracy.

Keywords –

Fatigue strength; Steel; Construction; Machine learning; Shapely additive explanations; CatBoost; Predictive modeling; Regression; Data interpretability.

1 Introduction

Accurate prediction of the fatigue strength of steels is of particular significance due to the extremely high costs and time requirements of fatigue testing, as well as the often-debilitating consequences of fatigue failures [1], [2]. Fatigue strength is the most fundamental data required for the design and failure analysis of structural components [3], [4], [5]. It is reported that fatigue accounts for over 90% of all mechanical failures in structural components [6], underscoring the critical need for reliable fatigue life prediction methods.

Fatigue failure is a complex phenomenon influenced by a multitude of variables, including material composition, microstructure, processing parameters, and operational conditions [7]. The intricate interactions among these variables pose significant challenges for conventional modeling and analytical approaches, which often fail to capture the non-linear relationships and hidden dependencies inherent in fatigue behavior. Consequently, a robust and reliable predictive framework is essential for advancing understanding and improving the safety and performance of steel components.

In recent years, machine learning (ML) algorithms have demonstrated significant potential in solving complex predictive tasks in structural engineering [8], [9], [10], [11]. Wang et al. [12] provided a review of fatigue life prediction methods, highlighting the limitations of purely data-driven approaches. Zhan et al. [13] presented a ML-based approach for predicting the fatigue life of additively manufactured stainless steel 316L. However, the application of these algorithms to fatigue strength prediction has been limited, particularly in terms of interpretability. Most ML models function as "black boxes," providing accurate predictions but little insight into the decision-making processes or the influence of individual variables. This lack of transparency hampers the adoption of ML techniques in safety-critical domains like fatigue prediction.

This study seeks to address this challenge by investigating the following research question: Can an interpretable ML framework effectively predict the fatigue strength of steel while providing actionable insights into the contributions of individual features, such as chemical composition and processing parameters? To achieve this, it leverages the CatBoost ML algorithm for predicting the fatigue strength of steels and employing

Shapley additive explanations (SHAP) to enhance model interpretability. Unlike alternative interpretability techniques such as Local Interpretable Model-agnostic Explanations (LIME) or attention mechanisms in deep learning, SHAP offers a global and local understanding of feature contributions while maintaining consistency and theoretical robustness in feature attribution. This research seeks to unravel the "black box" nature of the predictive model, providing actionable insights into the roles and relative importance of individual input variables by focusing on composition and processing parameters. The proposed approach not only improves the accuracy of fatigue strength predictions but also enhances understanding of the underlying mechanisms, bridging the gap between advanced computational tools and their practical applications in material design and construction engineering.

Through this work, we aim to contribute to the advancement of fatigue strength prediction methodologies and demonstrate the value of interpretable ML in tackling complex engineering problems, ultimately fostering greater trust and utility in data-driven predictive frameworks.

2 Shapley additive explanations

SHAP is a powerful interpretability technique based on cooperative game theory, designed to provide a transparent understanding of ML model predictions [14]. Originating from the Shapley value, which was developed to fairly distribute the gains of a coalition among its participants, SHAP translates this concept into a robust mathematical framework for evaluating the contribution of each feature to a model's output. One of the defining strengths of SHAP lies in its ability to offer both global and local interpretability. On a global scale, SHAP analyzes feature importance across the entire dataset, revealing overarching patterns and identifying the key drivers influencing the predictions. This enables stakeholders to understand the broader trends and relationships captured by the model. On the local level, SHAP provides detailed insights into individual predictions, quantifying the specific contribution of each feature to a given outcome.

Impact of SHAP is enhanced by its visualization tools, which facilitate intuitive exploration of model behavior. Summary Plots provide a bird’s-eye view of feature importance across the dataset. Summary plots reveal the most influential features and their directional impact (positive or negative) on the predictions by aggregating SHAP values for all features and instances. The color-coded representation (often based on feature values) enhances the clarity of this analysis. Dependence plots illustrate the relationship between a specific feature and the predicted outcomes, capturing non-linear effects and feature interactions. These visualizations help uncover intricate dependencies, shedding light on how features interact to influence predictions.

While SHAP offers a theoretically sound and practically valuable approach, it comes with some computational challenges. Computing exact Shapley values can be resource-intensive, especially for high-dimensional datasets or complex models. Approximation techniques, such as those implemented in SHAP libraries, mitigate this issue but may still pose scalability concerns for extremely large datasets. Alternative interpretability methods, such as LIME, provide faster but less stable feature attributions, while inherently interpretable models like decision trees and Explainable Boosting Machines (EBMs) offer transparency at the cost of predictive accuracy.

Despite these challenges, SHAP remains a cornerstone of model interpretability. Its rigorous foundation, adaptability to various ML models, and rich visualization toolkit empower users to make informed decisions, enhance model trust, and align ML applications with ethical and regulatory standards.

3 Numerical examples

In addressing the inherent black-box nature of ML models, particularly in the context of predicting the fatigue strength of steel, this study adopts an interpretable ML approach. The lack of transparency in black-box models often leads to reduced trust and understanding, posing challenges to their broader acceptance and applicability in critical engineering fields.

The Fatigue Dataset for Steel, obtained from the National Institute of Material Science (NIMS) MatNavi [15], serves as the basis for this study. This dataset is one of the largest globally, offering detailed information on chemical composition, upstream mill product features, and subsequent processing parameters. It includes data for carbon, low alloy, carburizing, and spring steels, with the target property being fatigue strength, derived from rotating bending fatigue tests conducted under room temperature conditions.

The dataset comprises 437 instances with 25 features related to composition, processing parameters, and mechanical properties. The features are categorized as follows:

- Chemical Composition: %C, %Si, %Mn, %P, %S, %Ni, %Cr, %Cu, %Mo (all in weight percentage).
- Upstream Processing Details: Ingot size, reduction ratio, and non-metallic inclusion characteristics.
- Heat Treatment Conditions: Parameters such as temperature, time, and other process-specific conditions for normalizing, through-hardening, carburizing-quenching, and tempering processes.
- Mechanical Properties: Yield strength (YS),

ultimate tensile strength (UTS), elongation (%EL), reduction of area (%RA), hardness, Charpy impact value (J/cm²), and fatigue strength.

Out of the 437 data instances, the breakdown is as follows: 371 carbon and low-alloy steels, 48 carburizing steels, 18 spring steels. These data instances represent various heats within each grade of steel and different processing conditions, providing a diverse and comprehensive dataset for predictive modeling.

The target property for this study is the fatigue strength of steel, while the features include a combination of chemical composition, processing parameters, and mechanical properties. This comprehensive selection enables the model to capture the complex interactions influencing fatigue behavior. Using the dataset, the study constructs predictive models with the CatBoost ML algorithm and applies SHAP to interpret the model's predictions. The integration of SHAP enhances transparency by identifying the contribution of each feature to the predicted fatigue strength, addressing the black-box nature of ML models, and facilitating actionable insights.

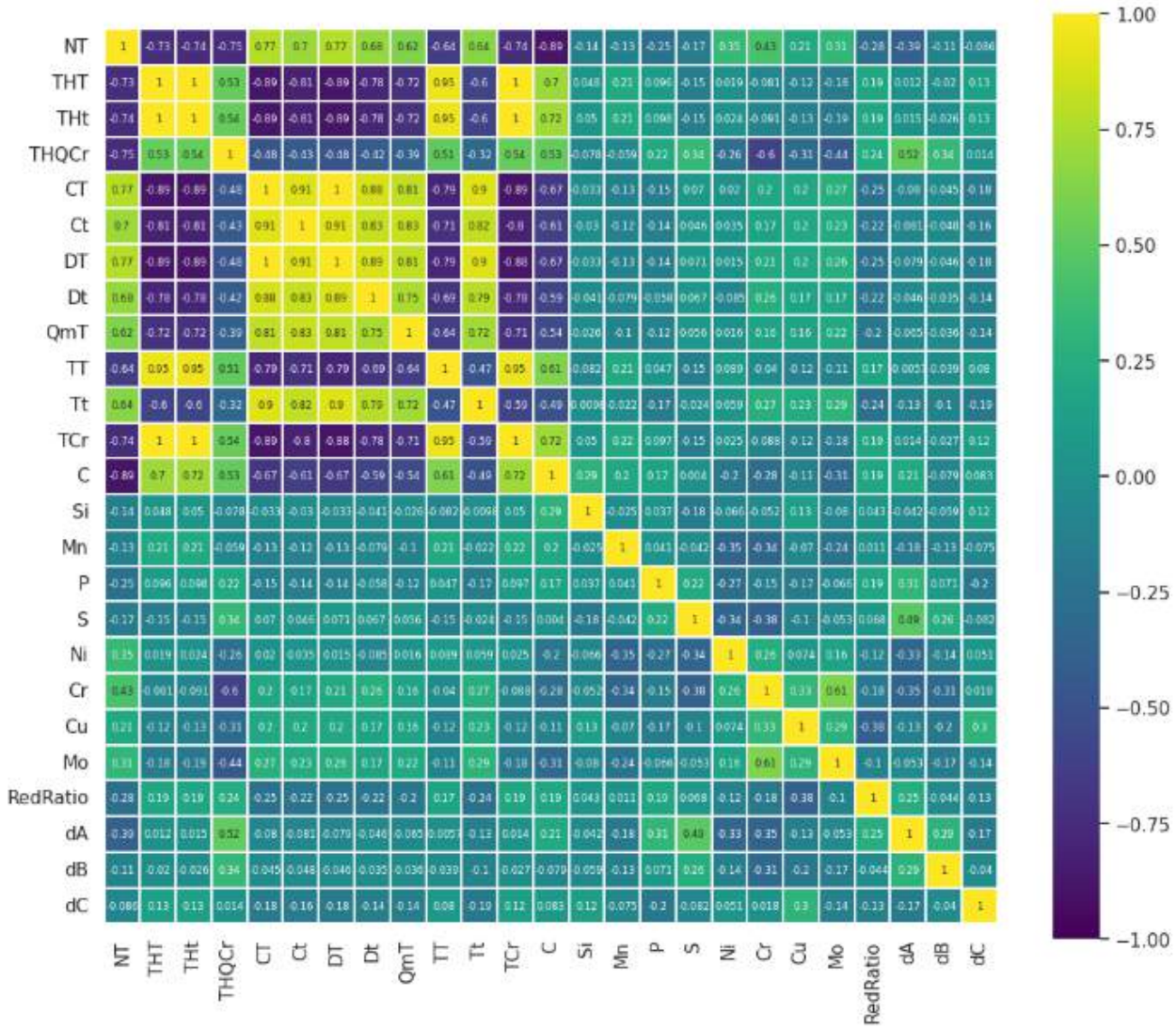


Figure 1. Correlation matrix

Details of the 25 features are provided in Table 1, offering a complete overview of the variables utilized in the model and their relevance to fatigue strength prediction. This numerical example demonstrates the practical application of an interpretable ML framework.

Figure 1 presents a correlation matrix illustrating the pairwise correlations between various numerical features, represented by the rows and columns of the matrix. The

color scale on the right indicates the range of correlations, from -1 (dark purple) to 1 (yellow). Positive correlations are depicted in green to yellow, while negative correlations appear in blue to purple. The diagonal values are all 1, as they represent the correlation of each feature with itself. Certain feature pairs exhibit strong positive correlations (close to 1, highlighted in yellow), including THt and THCr (~0.89), TT and THt (0.95), and Ct and DT (0.91). These strong correlations suggest a linear relationship between these variables. Conversely, some features show strong negative correlations (close to -1, dark purple), such as NT and THt (-0.74) and CT and QmT (-0.67). In contrast, some feature pairs have correlations close to 0 (e.g., NT and dC), suggesting no linear relationship. Highly correlated features could introduce multicollinearity issues, potentially necessitating their removal or combination. Conversely, features with low or no correlation to others may warrant further investigation.

Table 1 Data features

Abbreviation	Details
C	% Carbon
Si	% Silicon
Mn	% Manganese
P	% Phosphorus
S	% Sulphur
Ni	% Nicke
Cr	% Chromium
Cu	% Copper
Mo	% Molybdenum
NT	Normalizing Temperature
THT	Through Hardening Temperature
THt	Through Hardening Time
THQCr	Cooling Rate for Through Hardening
CT	Carburization Temperature
Ct	Carburization Time
DT	Diffusion Temperature
Dt	Diffusion time
QmT	Quenching Media Temperature (for Carburization)
TT	Tempering Temperature
Tt	Tempering Time
TCr	Cooling Rate for Tempering
RedRatio	Reduction Ratio (Ingot to Bar)
dA	Area Proportion of Inclusions Deformed by Plastic Work
dB	Area Proportion of Inclusions Occurring in Discontinuous Array
dC	Area Proportion of Isolated Inclusions
Fatigue	Rotating Bending Fatigue Strength (107 Cycles)

3.1 Regression model

A comparative study is conducted on several ML algorithms, including Decision Tree (DT), LightGBM, and CatBoost, to determine the most accurate one. Table 2 presents the R^2 and RMSE values for each algorithm. Hyperparameter tuning is performed using a grid search approach combined with 10-fold cross-validation to prevent overfitting. The hyperparameter values that yield the best performance are then selected for the ML models.

Table 2 Model performance

ML model	R^2	RMSE
DT	0.896	39.257
LightGBM	0.893	54.209
CatBoost	0.952	31.652

As a result of the comparative study, the CatBoost method achieved the highest accuracy with the optimal hyperparameter values as follows: Depth = 5, Iterations = 1000, and Learning Rate = 0.05. These optimal values were selected from the following ranges: Depth {5, 6, 7}; Iterations {500, 1000, 1500}; and Learning Rate {0.01, 0.03, 0.05}. Figure 2 presents the regression plot of the CatBoost model, which confirms its high accuracy. The data points generally align along a diagonal line, indicating that the predicted values closely match the actual values. Minor deviations from the diagonal suggest potential data quality issues, overfitting, or specific conditions affecting predictions. A strong correlation between the actual and predicted data is observed. The 10-fold cross-validation results in average R^2 and RMSE values of 0.952 and 31.652, respectively. Since the fatigue strength values in the dataset range widely from 225 MPa to 1190 MPa, an RMSE of 31.625 is relatively small, indicating good predictive performance. In the subsequent subsection, the CatBoost model is further analyzed using an interpretable method.

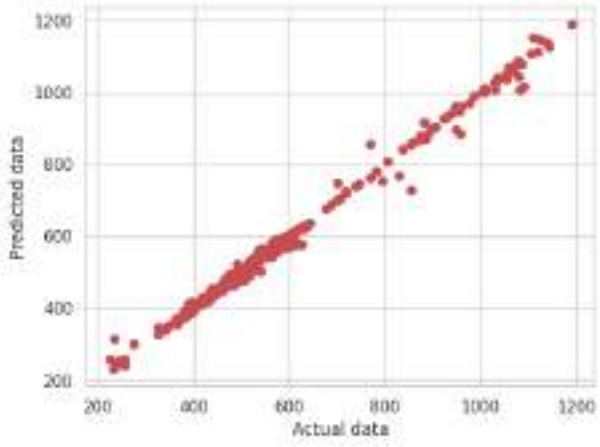


Figure 2. Regression plot

3.2 Interpretable method

Figure 3 represents the feature importance in CatBoost model, where each feature's contribution (relative importance) is measured and ranked. TT is the most significant feature, dominating the importance ranking. It dominates the analysis compared to other features. Tt, QmT, and Cr are the next most important features. There is a gradual decline in importance as we move down the list, with features like THt and RedRatio contributing minimal importance. The most important features have the largest impact on the performance of the model. Lower-ranked features contribute relatively little and might be considered for potential feature reduction or pruning in future analyses.

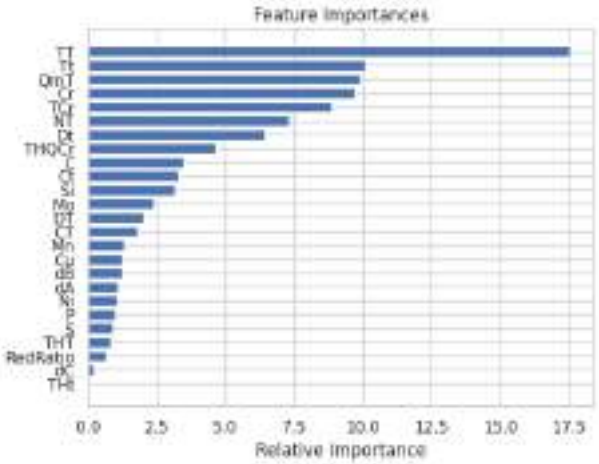


Figure 3. CatBoost importance factor

Figure 4 shows a SHAP summary plot. The features TT, Cr, and Mo have the largest SHAP values, meaning they have the most significant impact on the model's predictions. Their impact spans a wide range of positive and negative values. Features like Cr and Mo show a color gradient, meaning that low feature values (blue) and high feature values (red) influence predictions differently. For instance, high values of Cr (red) push predictions to the right (positive SHAP), while low values (blue) may push them to the left. Features lower down the list, like RedRatio, dA, dB, and P, have a much smaller spread of SHAP values. The varying spread of SHAP values (horizontal spread) reflects how much variance each feature causes in the predictions.

Figure 5 shows a SHAP summary plot that shows the importance of various features. TT has the highest importance in the model, contributing the most to the predictions. Cr follows closely, also playing a major role. Mo, QmT, Dt, and C show moderate impact compared to the top two. Features such as RedRatio, dA, dB, and P have very low SHAP values. The plot shows a clear skew towards the top features (TT, Cr), meaning the model's predictions rely heavily on just a few variables. Identifying the top contributing features (TT and Cr) helps focus on those key variables for further analysis or model optimization. Features with minimal contributions could potentially be excluded to simplify the model.

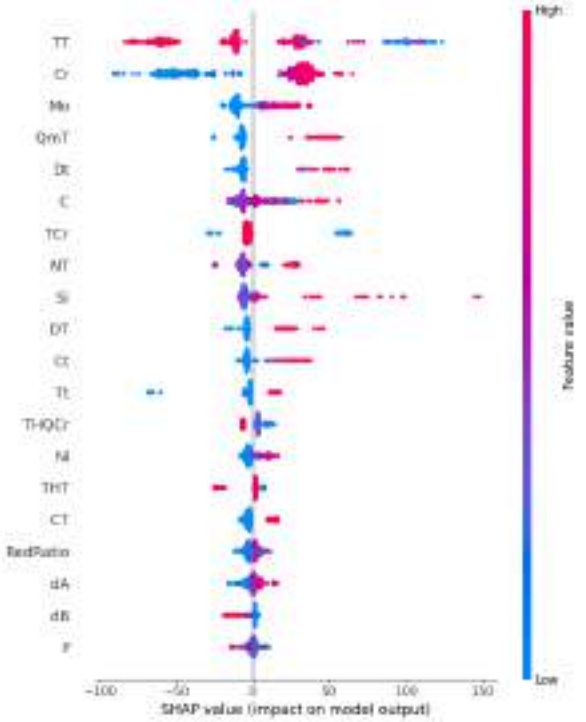


Figure 4. SHAP plots: Shapely value

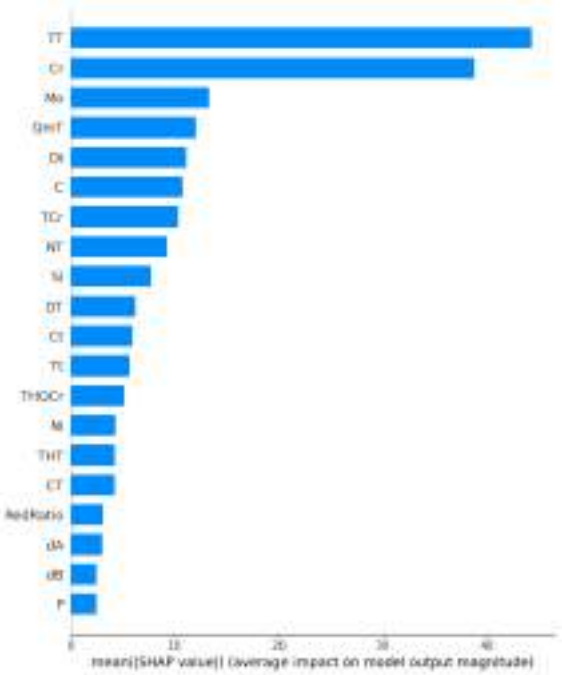
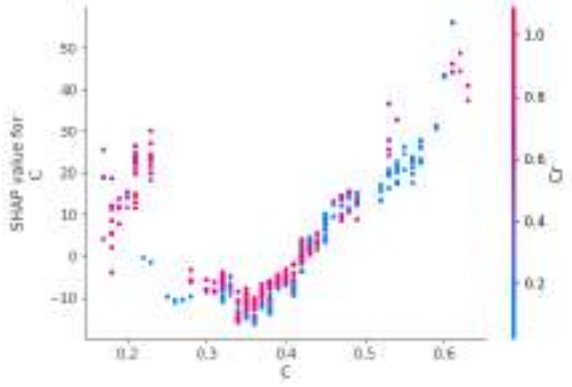
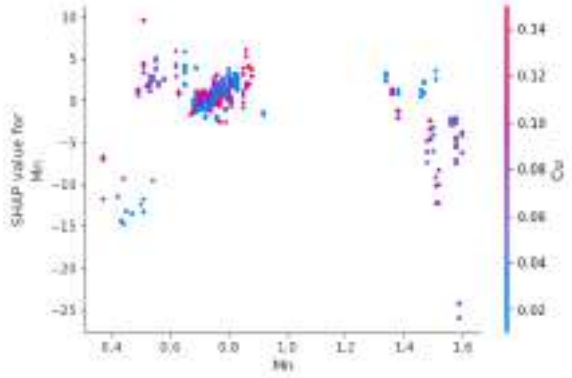


Figure 5. SHAP plots: Global importance factor.

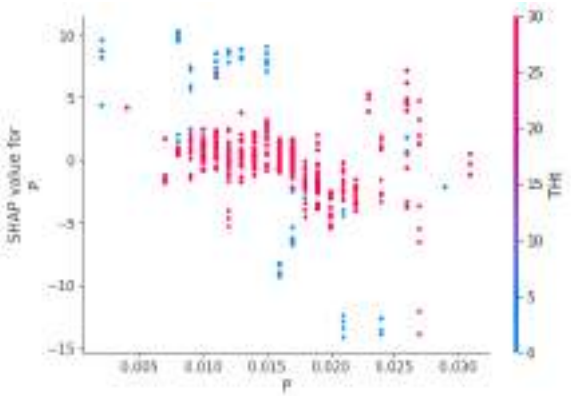
Figure 6 illustrates SHAP value plots that explain the contributions of various features to the model's predictions. Figure 6a shows that SHAP values increase non-linearly with rising "C" values, exhibiting a U-shaped pattern. Figure 6b highlights that the "Mn" feature forms two clusters with distinct SHAP patterns. Figure 6c demonstrates that SHAP values for "P" remain slightly positive for lower "P" values but become more dispersed as "P" increases. Figure 6d reveals a clear positive relationship between "S" and its SHAP values. Figure 6e indicates that SHAP values rise with "Ni" but are clustered into groups. Figure 6f shows that SHAP values increase with "Cr" in a strong positive linear trend. Figure 6g illustrates that SHAP values for "Cu" are mostly scattered around zero, with a slight positive correlation. Figure 6h demonstrates a positive relationship between SHAP values and "Mo" as "Mo" increases. Figure 6i shows that SHAP values fluctuate widely for lower "RedRatio" values but stabilize as "RedRatio" increases. Figure 6j indicates a slight upward trend in SHAP values as "dA" increases. Higher "Cr" appears to significantly influence the SHAP range. Overall, features such as "C," "Cr," "Mo," and "Ni" exhibit clear trends with SHAP values, while others, like "Cu" and "P," show more scattered distributions. These SHAP patterns suggest that specific features strongly contribute to predictions under particular conditions or value ranges.



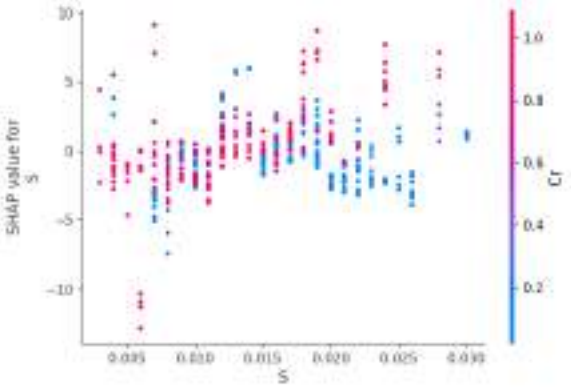
(a)



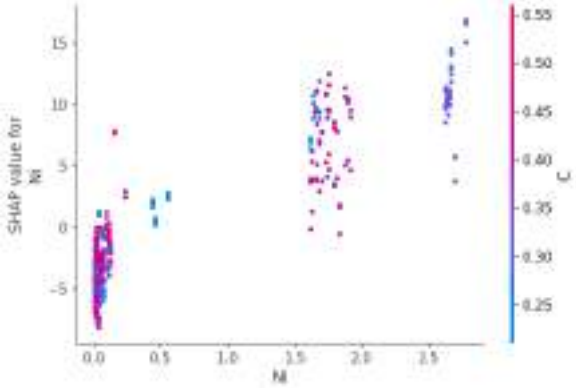
(b)



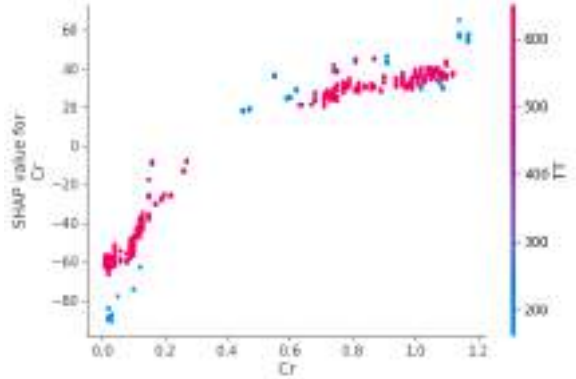
(c)



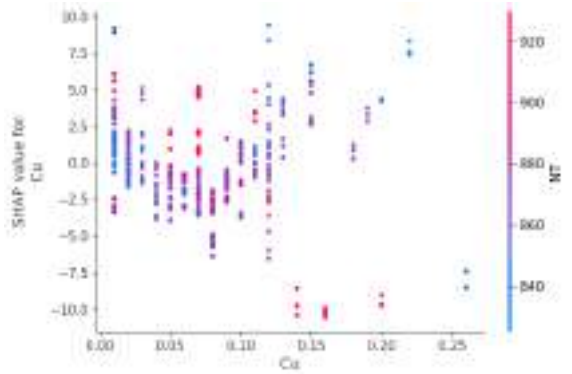
(d)



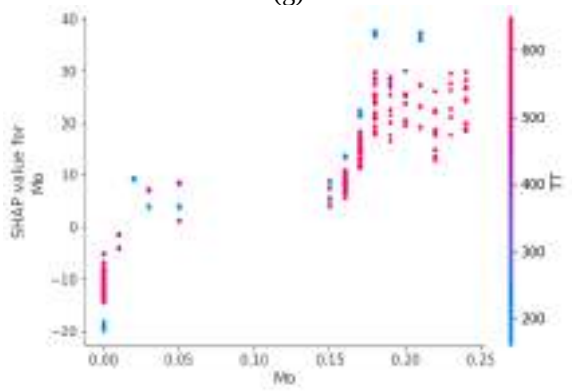
(e)



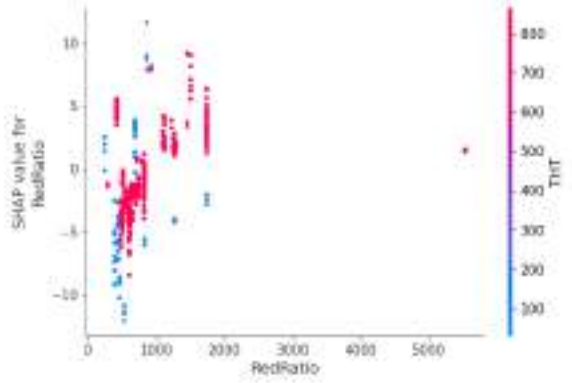
(f)



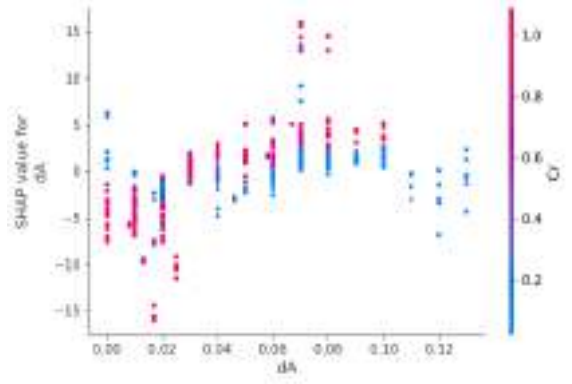
(g)



(h)



(i)



(j)

Figure 6. SHAP partial dependence plots.

4 Conclusions

This study demonstrates the significant potential of ML techniques, particularly the CatBoost algorithm, in accurately predicting the fatigue strength of steel while addressing the interpretability challenges commonly associated with black-box models. This research not only enhances predictive accuracy but also provides meaningful insights into the underlying relationships driving fatigue strength by incorporating a diverse dataset encompassing chemical composition, processing parameters, and mechanical properties. The use of SHAP further enables the decomposition of model predictions, offering a dual perspective on global trends and instance-specific behavior.

The analysis revealed that features such as tempering temperature (TT), chromium content (Cr), and molybdenum content (Mo) are the most influential in determining fatigue strength, with tempering temperature exhibiting the highest importance across predictions. The SHAP visualizations illustrated the nonlinear relationships and interactions among variables, highlighting how different combinations of features contribute to the overall predictions. For instance, the study identified synergistic effects between certain alloying elements and processing conditions, underscoring the complexity of fatigue behavior in steel and the importance of considering multivariate interactions in predictive modeling.

This research emphasizes the importance of interpretability in ML applications. The integration of SHAP enhances transparency, enabling engineers to gain actionable insights into the factors influencing fatigue strength. This interpretability fosters trust in the model's outputs, making it more suitable for safety-critical domains, such as structural engineering, where decision-making must be reliable and well-informed. From a practical perspective, this framework can support engineers in optimizing steel selection and processing decisions, reducing fatigue-related failures, and enhancing the sustainability and reliability of steel structures in construction engineering.

Furthermore, the study identified features with minimal contributions to model predictions, such as the reduction ratio and certain inclusion properties, which could be candidates for exclusion in future modeling efforts. This simplification has the potential to reduce computational demands and streamline predictive frameworks without sacrificing accuracy. Future research should explore feature selection methods to reduce redundancy and improve model efficiency, alongside alternative ML algorithms for comparison.

Future research can expand upon this framework by exploring the applicability of interpretable ML techniques to other material properties and engineering

challenges, such as fracture toughness or corrosion resistance. Additionally, incorporating larger and more diverse datasets, including environmental and operational factors, could further enhance the robustness and generalizability of the predictive models. The integration of additional techniques, such as explainable deep learning or hybrid models combining physics-based and data-driven approaches, may also provide deeper insights into complex structural behaviors.

Acknowledgment

The financial support provided by the EPSRC for the RESTOR project (Grant No. EP/W018705/1) and by the EU Commission for the AEGIR project (Grant No. 101079961) is gratefully acknowledged.

References

[1] G. Zuo, L. Tong, Z. Zhao, H. Wang, W. Shi, and Y. Yan, "Experimental study on fatigue strength of trapezoidal corrugated-web steel girders based on hot spot stress method," *Thin-Walled Structures*, vol. 208, p. 112814, Mar. 2025, doi: 10.1016/J.TWS.2024.112814.

[2] X. Guan, H. Guo, G. Li, D. Yang, and Y. Pan, "Fatigue behaviour of welded joints of Q690D high strength steel," *J Constr Steel Res*, vol. 223, p. 109037, Dec. 2024, doi: 10.1016/J.JCSR.2024.109037.

[3] S. He, H. Zhong, X. Huang, Y. Xu, and A. S. Mosallam, "Experimental investigation on shear fatigue behavior of perfobond strip connectors made of high strength steel and ultra-high performance concrete," *Eng Struct*, vol. 322, p. 119181, Jan. 2025, doi: 10.1016/J.ENGSTRUCT.2024.119181.

[4] F. Xu, J. He, X. Zhou, Y. Wang, Q. Yang, and Y. Huang, "Experimental investigation on low-cycle fatigue and lifetime estimation of high-strength lockbolt material ML40Cr steel," *Constr Build Mater*, vol. 449, p. 138317, Oct. 2024, doi: 10.1016/J.CONBUILDMAT.2024.138317.

[5] P. Mendes *et al.*, "High-cycle fatigue design curves of mild- and high-strength steels for offshore applications," *Structures*, vol. 67, p. 106827, Sep. 2024, doi: 10.1016/J.ISTRUC.2024.106827.

[6] P. D. Deshpande, B. P. Gautham, A. Cecen, S. Kalidindi, A. Agrawal, and A. Choudhary, "Application of Statistical and Machine Learning Techniques for Correlating Properties to Composition and Manufacturing Processes of Steels," *2 nd World Congress on Integrated Computational Materials Engineering* , pp. 155–160, Jul. 2013, doi: 10.1002/9781118767061.CH25.

[7] A. Agrawal, P. D. Deshpande, A. Cecen, G. P. Basavarsu, A. N. Choudhary, and S. R. Kalidindi, "Exploration of data science techniques to predict fatigue strength of steel from composition and processing parameters," *Integr Mater Manuf Innov*, vol. 3, no. 1, pp. 90–108, Dec. 2014, doi: 10.1186/2193-9772-3-8/FIGURES/7.

[8] S. Kookalani, B. Cheng, and S. Xiang, "Shape optimization of GFRP elastic gridshells by the weighted Lagrange ε-twin support vector machine and multi-objective particle swarm optimization algorithm considering structural weight," *Structures*, vol. 33, pp. 2066–2084, 2021, doi: 10.1016/j.istruc.2021.05.077.

[9] S. Kookalani, S. Nyunn, and S. Xiang, "Form-finding of lifting self-forming GFRP elastic gridshells based on machine learning interpretability methods," *Structural Engineering and Mechanics*, vol. 84, no. 5, pp. 605–618, 2022, doi: 10.12989/sem.2022.84.5.605.

[10] S. Kookalani, B. Cheng, and J. L. C. Torres, "Structural performance assessment of GFRP elastic gridshells by machine learning interpretability methods," *Frontiers of Structural and Civil Engineering*, vol. 16, no. 10, pp. 1249–1266, 2022, doi: 10.1007/s11709-022-0858-5.

[11] S. Kookalani *et al.*, "Trajectory of building and structural design automation from generative design towards the integration of deep generative models and optimization: A review," *Journal of Building Engineering*, vol. 97, p. 110972, Nov. 2024, doi: 10.1016/J.JOBE.2024.110972.

[12] H. Wang, B. Li, J. Gong, and F. Z. Xuan, "Machine learning-based fatigue life prediction of metal materials: Perspectives of physics-informed and data-driven hybrid methods," *Eng Fract Mech*, vol. 284, p. 109242, May 2023, doi: 10.1016/J.ENGFRACMECH.2023.109242.

[13] Z. Zhan and H. Li, "Machine learning based fatigue life prediction with effects of additive manufacturing process parameters for printed SS 316L," *Int J Fatigue*, vol. 142, p. 105941, Jan. 2021, doi: 10.1016/J.IJFATIGUE.2020.105941.

[14] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," *Adv Neural Inf Process Syst*, vol. 30, 2017.

[15] "National Institute of Materials Science", doi: http://smds.nims.go.jp/fatigue/index_en.html.

Ontology-driven LLM framework for knowledge graph in smart buildings

Sebastien Gerin[1], Laurent Joblot[2], Nisrine Makhoul[1], Frederic Merienne[2]

[1]Smart, Sustainable, and Resilient Cities, ESTP, 11 Rue Sully, 21000 Dijon, France
[2]Arts et Métiers Institute of Technology, LISPEN / HESAM University / UBFC, Chalon-Sur-Saône, France

sgerin@estp.fr, laurent.joblot@ensam.eu nmakhoul@estp.fr frederic.merienne@ensam.eu

Abstract -

The rapid data growth in smart building environments requires advanced tools to integrate, interpret, and utilize this information effectively. Smart buildings generate vast and heterogeneous data streams, including sensor readings, occupancy metrics, and environmental conditions, which are critical for optimizing energy efficiency, enhancing occupant comfort, and enabling predictive maintenance. However, the lack of a structured approach to automatically connect and contextualize these data sources limits the insights that can be derived. To address these challenges, this paper presents a framework for assessing data in a knowledge graph by automatically retrieving entities and establishing relationships from diverse data sources, incorporating metadata standards and time-series data relevant to smart buildings. A standard ontology from the domain is used to drive the experiment, enabling the automatic construction of a semantic graph-based model for a real-world smart building environment. The framework's objective is to ensure information is comprehensible as a preliminary step to intelligent decision-making in data-driven smart buildings, enabling applications like fault detection, performance measurement, and energy auditing. The proposed approach explores the potential of large language models (LLMs) to automate data integration, reducing reliance on experts. This paper addresses existing literature gaps on metadata mapping and lays the groundwork for future advancements in digital twin technologies for smart building applications.

Keywords -

Ontology; Knowledge graph; Digital twins; data-driven approach; Large language models;

1 Introduction

Achieving global Net Zero Emissions (NZE) by 2050 is a critical goal for mitigating climate change, demanding transformative actions across various sectors. Among these, the Architecture, Engineering, Construction, and Operation (AECO) sector plays a pivotal role, necessitating an accelerated transition to align with the NZE scenario [1]. To this end, many countries have implemented stringent building energy codes and minimum performance standards aimed at reducing energy consumption and greenhouse gas emissions. Technological advancements and industry efforts have led to the rise of data-driven smart buildings (DDSBs), which integrate digitalization technologies to optimize energy use dynamically, enhance indoor environmental quality, and improve occupant experiences [2]. DDSBs generate vast amounts of data through Internet of Things (IoT) technologies, collecting real-time and historical information via advanced sensors. Effectively leveraging this data requires robust frameworks for systematic organization and analysis [2]. Complementing DDSBs, digital twins (DTs) further enhance operational efficiency by enabling predictive maintenance, informed decision-making, and cost minimization [3]. Researchers define a DT as "a virtual representation of a physical system (and its associated environment and processes) that is updated through the exchange of information between physical and virtual systems" [4]. Initially developed for aerospace applications, DTs now extend to diverse domains such as industrial manufacturing, healthcare, smart cities, education, and agriculture [5]. DTs are real-time digital replicas of physical entities or processes that enable a two-way flow of information. The virtual model can be updated in real-time, allowing simulations and informed decision-making to guide physical system adjustments [6]. When integrated with DDSBs, DTs leverage real-time and historical data to enhance data-driven processes, optimizing energy efficiency, indoor environmental quality, and overall system performance [2]. A critical aspect of DT implementation is the integration of a robust data schema, a structured framework essential for defining how data is organized, stored, and managed throughout the building's lifecycle [7]. Integrating appropriate data schemas requires specialized domain knowledge experts and a deep understanding of building systems. The lack of accessible, standardized solutions often hinders the practical deployment of DTs in the built environment, leading to underutilized data, ineffective decision-making, and inefficient practices with significant cost implications. This research introduces a framework that automates the mapping of data generated

by smart buildings into a standardized model, serving as the foundational step in constructing a comprehensive DT. This paper seeks to address these limitations by answering the following questions:

1. How can we address the lack of generic tools for building digital twins in data-driven smart buildings?
2. What is the most effective method for automatically structuring heterogeneous data from a data-driven smart building into a coherent data schema?
3. Does our method perform well regardless of the data model used?

The paper is structured as follows: Section 2 reviews related work, providing context and highlighting existing approaches. Section 3 presents the proposed framework, which leverages Large Language Models (LLMs) for automated knowledge graph (KG) construction, significantly reducing the need for human intervention. The utilization of a prebuilt standard semantic graph ontology serves to enhance the organization of metadata, facilitating better efficiency and accuracy. Section 4 evaluates the framework through its implementation at the Dijon Metropolitan Campus, demonstrating its practical applicability in handling real-world datasets. Finally, Section 5 concludes the paper by summarizing the key findings and discussing potential avenues for future research and development.

2 Related work

Ontologies are essential for achieving data integration and interoperability in DDSBs and DTs. By providing a formalized vocabulary and structured taxonomy, they enable the systematic organization of domain-specific knowledge through well-defined classes, properties, and relationships [8]. Lygerakis et al. [9] identified over 40 ontologies relevant to energy optimization across the lifecycle of connected buildings. However, the authors highlighted that most remain theoretical, with limited real-world applications in operational settings. This literature review identifies several key ontologies as potentially relevant [10, 7], including the Building Ontology Topology (BOT) [11], SSN/SOSA [12], Brick [13], SAREF4BLDGG [14], the RealEstateCore ontology [15], and Google's Building Digital Ontology (DTO) [16]. While these ontologies provide a foundational schema for building data management, their implementation often necessitates significant domain expertise, posing challenges to scalability and automation. Knowledge graphs offer an alternative approach to managing heterogeneous and dynamic data in DDSBs and DTs. As graph-based data models, KGs structure knowledge in a machine-readable format, with nodes representing entities and labeled edges capturing relationships [17]. This flexibility enables KGs to augment digital twins by integrating real-time and historical data while supporting advanced analytics [18]. Unlike ontologies, KGs facilitate linking structured and unstructured data, offering actionable insights and enhancing decision-making capabilities. Despite these advantages, constructing KGs for DDSBs remains a non-trivial task, requiring the resolution of data ambiguities, alignment with specific application requirements, and managing diverse data sources [19]. Existing methodologies have primarily relied on semi-automated approaches. For instance, Chamari et al. [20] proposed a methodology to generate metadata schemas by correlating Building Automation System (BAS) identifiers with ontology classes through text search. This process, while partially automated, relies on human intervention for refining mappings. Similarly, Waterworth et al. [21] addressed inconsistencies in BAS metadata abbreviations using rules and language models to extract and classify semantic information. Koh et al. [22] developed a workflow utilizing machine learning algorithms to convert unstructured BAS metadata into structured formats like Brick. However, these approaches fall short of delivering fully automated solutions and often depend on significant manual effort or predefined rules.

To address these limitations, we propose a novel methodology for automating the construction of KGs from DDSB data. This approach proposes to limit the dependency on domain expertise by systematically mapping heterogeneous data into a standardized model, representing a foundational step toward the comprehensive implementation of digital twins. By directly addressing challenges such as data ambiguity and integration complexity, this methodology advances the field of DDSB data management and supports scalable, interoperable solutions. The following section details the design and implementation of our proposed framework.

3 The proposed approach

This study proposes a methodology to automate the generation of a KG from DDSB data using a large language model (LLM). As advanced generative AI systems, LLMs excel in processing and generating natural language, making them particularly effective for automating tasks such as entity and relationship extraction, competency question formulation, and metadata organization [23]. By reducing the dependence on manual effort and domain expertise, LLMs enhance both the scalability and precision of KG construction. However, a human-in-the-loop approach remains crucial for critical evaluation stages to ensure accuracy and reliability [24].

At the core of the methodology is the Retrieval-Augmented Generation (RAG) technique, further refined

by the GraphRAG method[1]. RAG enhances the contextual relevance and accuracy of LLM outputs by combining semantic similarity retrieval from proprietary data sources with generative language capabilities [25]. This dual approach ensures that generated outputs align closely with domain-specific requirements, making RAG particularly valuable for knowledge-intensive domains such as DDSBs. GraphRAG extends RAG by introducing a KG-centric enhancement. The GraphRAG methodology organizes implicit knowledge from text into graph representation and creates communities based on the graph structure. The generated graph serves as context for RAG tasks, leveraging summaries from high-level communities to produce more diverse outputs. Comparative studies show that GraphRAG outperforms standard RAG methods in terms of accuracy, scalability, and cost-efficiency [26, 27].

Our approach is inspired by the GraphRAG process but uses a pre-built domain ontology as the context graph instead of generating one from raw data. In this manner, the domain ontology provides the context for the RAG operations, which empowers the retrieval process to extract entities from the domain-specific graph. This innovative approach addresses the challenges of data heterogeneity and complexity inherent to DDSBs, facilitating the creation of interoperable and standardized DTs. As illustrated in Figure 1, our proposed framework comprises three main phases: *data collection*, *KG construction*, and *KG visualization*. The following sections detail each phase of our process.

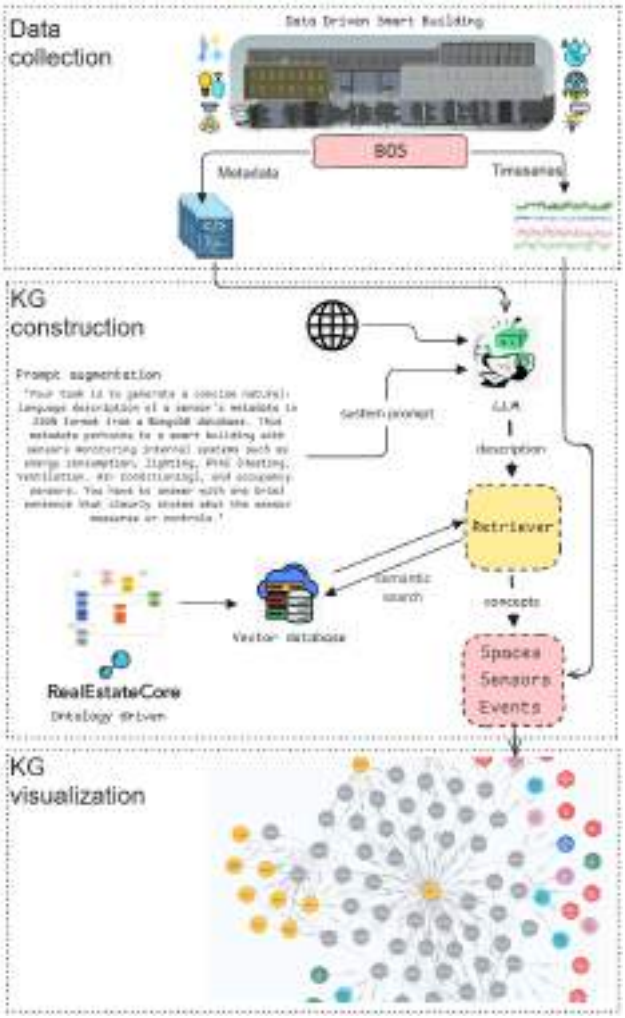


Figure 1. The architecture of the proposed approach.

3.1 Data collection

The data collection phase represents the foundational step in our framework, aimed at systematically gathering and organizing the diverse data sources intrinsic to DDSBs. Smart buildings produce a wide spectrum of heterogeneous data, including IoT identifiers, time-series data from building management systems, and Building Information Modeling (BIM) files, often distributed across siloed and disparate databases. IoT metadata typically captures device-specific attributes such as type, operational parameters, control commands, and spatial topology, defined either by vendor-specific schemas or standardized frameworks. In our approach, the Building Operating System (BOS) plays a central role as the primary data aggregator, integrating information from interconnected devices and systems within the DDSB. This centralized integration ensures a unified flow of data, enhances compatibility with third-party applications, and facilitates the systematic identification and extraction of critical entities and relationships required for KG construction. By addressing the challenges of data heterogeneity and accessibility, this initial phase prepares the collected data for seamless transition to the subsequent phase.

[1]https://github.com/microsoft/graphrag

3.2 KG construction

The second phase of our framework focuses on constructing a domain-specific KG by leveraging an enhanced LLM adapted to the GraphRAG methodology. This approach uses the LLM to generate concise natural-language descriptions from metadata provided by the DDSB and incorporates a predefined ontology to guide the semantic retrieval of appropriate concepts. The process begins with **(1) ontology integration**, where the GraphRAG system's context is set using a predefined ontology that semantically describes each asset of the DDSB, providing a structured framework for metadata interpretation. Next, **(2) metadata description generation** involves prompting the LLM to produce concise natural-language descriptions of the DDSB metadata, which includes dynamic data such

as sensor readings (e.g., temperature, humidity, energy usage) and static data like spatial and topological information (e.g., classrooms, parking areas, laboratories, and building levels). Following this, **(3) concept retrieval** is performed, where a semantic similarity search is used to identify and retrieve the most relevant concepts from the ontology, aligning them with the metadata descriptions. Finally, **(4) KG construction** links the retrieved concepts with time-series data, organizing the information into a cohesive structure. Relationships are established between nodes representing topology, sensors, and time-series values, all formatted according to the ontology schema. This comprehensive process ensures that the constructed KG is semantically accurate, contextually relevant, and capable of representing the complex relationships inherent in DDSB data.

3.3 KG visualization

The final phase of our framework focuses on the visualization and evaluation of the constructed KG to ensure its accuracy and usability. The KG is exported to a graph database management system, which provides robust tools for querying and visualizing the graph's structure. Visualization plays a critical role in the validation process, enabling a thorough review of the interconnections between nodes. Specifically, this step involves verifying the alignment of the building topology, ensuring that sensors are correctly assigned to their respective locations, and confirming that observation values are accurately linked to the corresponding sensors, along with their associated timestamps. These visual checks are essential to identify and rectify any inconsistencies or errors that may arise during the KG construction phase. Additionally, the visualization phase facilitates an intuitive understanding of the underlying relationships and patterns in the data, making the KG more accessible to domain experts and stakeholders. This step concludes the methodology by ensuring that the KG is not only accurate and reliable but also ready for downstream applications, such as querying, analysis, and integration into digital twin systems.

4 Experimental results

This section presents the experimental results to evaluate the performance of the proposed method and analyze its effectiveness.

4.1 Evaluation setup

This study is conducted at the Dijon Metropolitan Campus[2], a $10,000\ m^2$ smart building certified under the Ready2Service (R2S) standard for its advanced digital infrastructure. The campus employs 14 types of sensors to monitor various parameters, including occupancy, temperature, CO2 levels, humidity, illuminance, and energy consumption, across 350 distinct areas encompassing 32 spatial categories such as waiting rooms and classrooms. The system stores the collected data in a schema-less MongoDB database in JSON format. A one-month subset of this time-series data amounts to approximately 11 GB and comprises over 50 million entries.

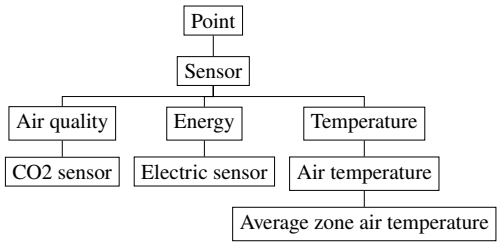


Figure 2. Part of REC ontology for sensor hierarchy.

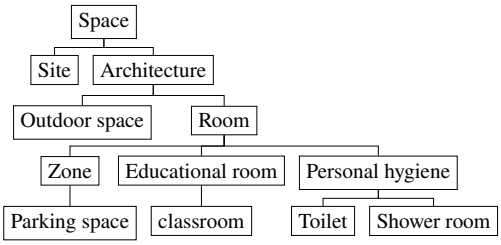


Figure 3. Part of REC ontology for space hierarchy.

To guide our methodology, we identify the RealEstateCore ontology (REC) as the most suitable framework due to its robust and comprehensive structure [15]. REC provides a detailed taxonomy that includes 240 sensor types (e.g., humidity, motion, temperature, lighting), as illustrated in Figure 2, and 92 spatial and architectural categories (e.g., back office, entrance, laboratory, warehouse), depicted in Figure 3. These categories are organized hierarchically to accommodate diverse use cases, enhancing the accuracy of KG construction and optimizing data retrieval processes. Moreover, REC employs the Label Property Graph (LPG) format, a data-centric model that facilitates real-time operations by enabling efficient storage and access of large datasets through graph databases [28]. Originally developed as an OWL ontology for RDF-based knowledge graphs, REC has been adapted into the Digital Twin Definition Language (DTDL)[3] to better align with user requirements and streamline deployment. DTDL, built on the JSON-LD format[4], ensures compatibility with the existing campus database, thereby eliminating the need

[2]https://www.estp.fr/en/dijon-campus

[3]https://learn.microsoft.com/en-us/azure/digital-twins/concepts-models

[4]https://json-ld.org/

for additional storage solutions while streamlining the integration process. This combination of hierarchical organization, real-time graph database performance, seamless integration, and compatibility with existing systems underscores REC's efficacy as the foundational framework for our approach.

4.2 Results and discussion

We present the results of applying the methodology to generate a KG at the Dijon Metropolitan Campus for a given day. The objective is to identify the pertinent classes within the REC ontology that align with the disparate points on the Dijon Metropolitan Campus and establish a relationship with the time series data. The resulting KG is exported to the Neo4J graph database[5] for storage and visualization. Figure 4 depicts a part of the resulting KG, in which the time-series data are linked to the standard semantics of the sensors and spaces of the Dijon Metropolitan Campus, encapsulated within the REC ontology model. Figure 5 illustrates an example of a node representing a cafeteria room that contains several attributes, including designation and display name, which have been retrieved from the metadata. This example demonstrates the hierarchy of the nodes, whereby the cafeteria room inherits from the food handling room, a room, and a space, corresponding to the REC ontology hierarchy and taxonomy. Figure 6 illustrates an example of a node representing a time series value connected to a node representing a temperature sensor. Node's properties include the temperature value and the timestamp. The objective is to demonstrate the enrichment of time-series data with standard semantics, facilitating its linkage to the sensors and equipment for efficient querying.

We evaluate the proposed methodology by presenting results from multiple perspectives, including processing execution time, repeatability, comparison of LLMs, concepts retrieval evaluation, and KG query evaluation.

4.2.1 Computational efficiency

The efficiency of the proposed methodology was assessed based on its execution time for generating the KG. The retriever identifies all relevant concepts within one minute, followed by KG construction in less than 20 seconds for a specific day's time-series data. Using a laptop equipped with an Intel Core i7 processor and 32 GB of RAM, the system generated 42, 728 JSON-LD documents (nodes) within this timeframe.

[5]https://neo4j.com/

4.2.2 Model consistency

The consistency of the retrieval methodology was assessed using three LLMs: Llama3.2[6], Mistral7B[7], and Gemma2[8]. Each model underwent 20 trials, with conservative parameter settings to minimize response variability. The results demonstrated that the retrieval process was reliable within individual models, with minimal fluctuations in performance due to the conservative configurations. However, notable discrepancies arose in how the models interpreted and retrieved concepts from the REC ontology. These differences highlight the distinct semantic inference strategies employed by each LLM, emphasizing the importance of carefully selecting the appropriate model for specific use cases and applications.

4.2.3 Model comparison and performance

The results shown in Table 1 illustrate pairwise comparisons that quantify the percentage of identical REC concepts retrieved by each LLM. Low values indicate that each LLM retrieved different concepts, while high values suggest greater homogeneity in the retrieved REC concepts across the models. The results generally show low to moderate values, indicating weak overlap in the retrieved concepts. This variation highlights differences in how each language model interprets and processes the semantics of the metadata and the REC ontology concepts, emphasizing the inherent differences in their inference mechanisms.

Table 1. Retrievability analysis across LLMs for 14 sensors (Se.) and 32 spaces (Sp.).

LLM	Llama		Mistral		Gemma	
	Se.	Sp.	Se.	Sp.	Se.	Sp.
Llama	-	-	28%	53%	21%	68%
Mistral	28%	53%	-	-	28%	50%
Gemma	21%	68%	28%	50%	-	-

4.2.4 Semantic retrieval accuracy

A manual qualitative assessment was conducted to evaluate the identification of 32 spaces and 14 sensor types from the REC ontology. Each retrieved concept was categorized as "correct," "ambiguous," or "incorrect." The ground truth for each sensor and space category was established through direct analysis of the data or reference to the BIM model. For instance, an incorrect retrieval was observed when the Gemma model classified a *meeting room* as a *workshop room*, while both the Mistral and Llama models correctly identified a *conference room*. In cases where ambiguity was present, all LLM models

[6]https://www.llama.com/
[7]https://docs.mistral.ai/
[8]https://github.com/google-deepmind/gemma

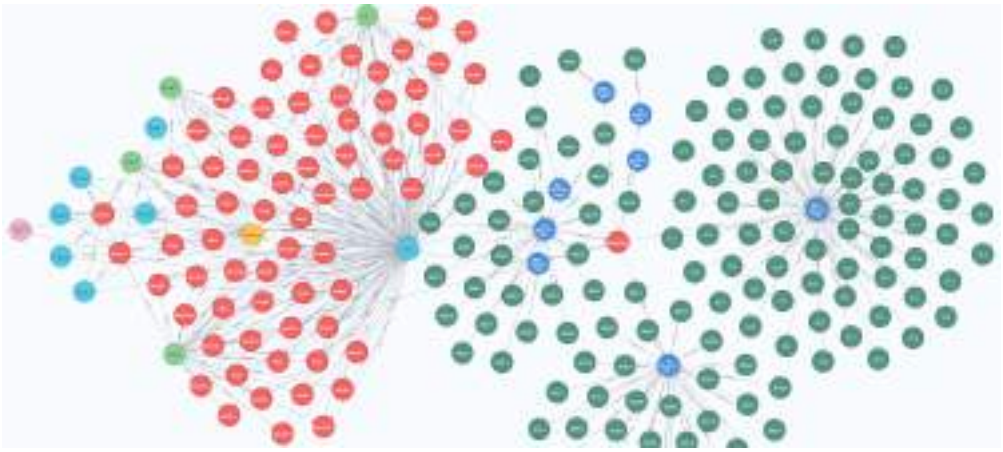


Figure 4. Part of the generated KG for the Dijon Metropolitan Campus: red nodes represent spaces, blue nodes represent sensors, and green nodes represent values.



Figure 5. Cafeteria room node with hierarchy.



Figure 6. Temperature observation node with values.

classified *toilets* as *personal hygiene rooms*, despite the explicit existence of the toilet concept in the REC ontology. An example of a correct retrieval is the occupancy sensor, which was accurately mapped to the corresponding concept within the REC ontology. The Dijon Metropolitan Campus employs energy sensors that monitor electricity consumption in kilowatt-hours (kWh) across various equipment, such as lighting, power outlets, and heating systems. The most appropriate concept for this data in the REC ontology is the *Electric Energy Sensor*. However, depending on the LLM, the retrieved concepts varied between Energy Usage Sensor and Energy Sensor. While both of these concepts are valid, they do not fully align with the ground truth, and as such, these responses were categorized as "ambiguous," leading to a loss of semantic precision regarding the type of electrical energy. Table 2 summarizes the performance of the models, revealing that the Llama model consistently outperformed the others, exhibiting high accuracy with minimal errors in both space and sensor retrieval tasks. The Gemma model demonstrated moderate performance, excelling in space retrieval but encountering more ambiguities in sensor retrieval. The Mistral model, in contrast, showed the weakest performance, particularly in space retrieval, with a significant number of incorrect responses.

However, ambiguous responses are not necessarily erroneous, though they do not invariably correspond with the ground truth. The models performed consistently across various scenarios, achieving an overall correct identification rate of 73% when ambiguous answers were considered accurate

Table 2. Retrievability evaluation of sensors (Se.) and spaces (Sp.) across LLMs: distribution of correct, ambiguous, and incorrect retrievals.

LLM	Correct		Ambiguous		Incorrect	
	Se.	Sp.	Se.	Sp.	Se.	Sp.
Llama	12	27	2	2	0	3
Mistral	5	17	3	2	6	13
Gemma	5	20	8	2	1	10

4.2.5 KG query evaluation

We leverage the semantics of metadata and time-series data structured within the REC ontology to assess the generated KG. The relevance of semantic relationships is explored through a series of Cypher queries[9], a declarative query language designed for use with the Neo4j graph database. The evaluation involved conducting automated queries on the KG and validating the results against a predefined ground truth. A total of 15 queries were formulated and categorized into three distinct types: **(1) room type query** counting, which examines the distribution of room types within the KG; **(2) sensor type query** counting, focusing on the frequency of occurrence of various sensor types; and **(3) sensor-to-room relationships**, aimed at evaluating the connections between sensors, their associated time-series data, and the specific spaces they monitor. For a complete list of queries, please refer to the file `queries_for_kg_evaluation.md` in the GitHub repository[10]. The results of these evaluations are summarized in Table 3, which presents the number of correct answers for each LLM. The ground truth was established by querying the original MongoDB database or through manual analysis of data from the BIM model. The Llama model demonstrated the most consistent performance, achieving near-perfect scores across all query types, thereby illustrating its robust ability to handle both sensor and spatial data. The Mistral model performed reasonably well in sensor-to-room mapping but exhibited weaker results in the room type and sensor type queries. The Gemma model, while excelling in sensor-related queries, faced challenges in accurately identifying room types. These findings highlight the effectiveness of using automated unit tests for evaluating and refining the performance of knowledge graph queries.

Table 3. Automatic query evaluation.

LLM	Room type	Sensor type	Sensor-to-room
Llama	4/5	5/5	4/5
Mistral	2/5	3/5	4/5
Gemma	1/5	5/5	5/5

5 Conclusion and future work

This paper presents an LLM-based framework that leverages an ontology to guide the automatic construction of a KG, facilitating the management of data in smart buildings. By integrating LLMs, the framework reduces reliance on human expertise while improving the accuracy of KG development. Experimental results demonstrate the effectiveness of the REC ontology in guiding KG creation from unstructured data. However, while the method accelerates the process, human verification is still required to ensure the correctness of identified concepts, introducing potential delays and biases. These preliminary findings suggest promising directions for future research, including integrating BIM assets into the KG [29] to enhance organization and interoperability. Expanding the REC ontology to address ambiguous or incorrect concepts could further refine the KG's semantic alignment. Additionally, exploring the use of alternative ontologies, such as Brick, would increase user autonomy and broaden the framework's applicability. Further testing with metadata from other smart buildings would also be beneficial. Once established, the KG can serve as a foundation for LLMs, enabling natural language queries that improve accessibility for diverse stakeholders. Moreover, applying advanced graph algorithms, such as graph neural networks, could offer valuable insights, supporting classification and clustering tasks that enhance decision-making in complex smart building scenarios.

[9] https://neo4j.com/docs/cypher-manual/current/introduction/

[10] https://tinyurl.com/3heep8pu

6 Acknowledgement

We thank Primatice, a subsidiary of Vinci Facilities, for their financial and material support. Primatice is developing TwinOps[11], a platform for managing digital twins of real estate assets to improve smart building operations and maintenance.

[11] https://www.twinops.com/

References

[1] IEA. IEA (2022), all countries targeted for zero-carbon-ready codes for new buildings by 2030, iea, paris, Accessed: 01/10/2024.

[2] J. Candanedo, D. Blum, Z. Chen, G. Fierro, V. Gori, H. Johra, H. Madsen, A. Marszal-Pomianowska, Z. O'Neill, O. Pradhan, et al. Data-driven smart buildings: State-of-the-art review. *Energy in Building and Communities Programme*, page 103, 2023.

[3] Y. Mousavi, Z. Gharineiat, A. A. Karimi, and S. Gonizzi Barsanti. Digital twin technology in built environment: A review of applications, capabilities and challenges. *Smart Cities*, 7(5):2594–2615, 2024.

[4] E. VanDerHorn and S. Mahadevan. Digital twin: Generalization, characterization and implementation. *Decision support systems*, 145:113524, 2021.

[5] M Mateev. Industry 4.0 and the digital twin for building industry. *Industry 4.0*, 5(1):29–32, 2020.

[6] M. Attaran and B. G. Celik. Digital twin: Benefits, use cases, challenges, and opportunities. *Decision Analytics Journal*, 6, 2023.

[7] G. Fierro and P. Pauwels. Survey of metadata schemas for datadriven smart buildings (annex 81). *Energy in Buildings and Communities Technology Collaboration Programme*, 2022.

[8] T. R. Gruber. A translation approach to portable ontology specifications. *Knowledge acquisition*, 5 (2):199–220, 1993.

[9] F. Lygerakis, N. Kampelis, and D. Kolokotsa. Knowledge graphs' ontologies and applications for energy efficiency in buildings: A review. *Energies*, 15(20): 7520, 2022.

[10] Z. Qiang, S. Hands, K. Taylor, S. Sethuvenkatraman, D. Hugo, P. G. Omran, M. Perera, and A. Haller. A systematic comparison and evaluation of building ontologies for deploying data-driven analytics in smart buildings. *Energy and Buildings*, 292, 2023.

[11] M. H. Rasmussen, M. Lefrançois, G. F. Schneider, and P. Pauwels. Bot: The building topology ontology of the w3c linked building data group. *Semantic Web*, 12(1):143–161, 2021.

[12] M. Compton, P. Barnaghi, L. Bermudez, R. Garcia-Castro, et al. The ssn ontology of the w3c semantic sensor network incubator group. *Journal of Web Semantics*, 17:25–32, 2012.

[13] B. Balaji, A. Bhattacharya, G. Fierro, J. Gao, J. Gluck, et al. Brick: Metadata schema for portable smart building applications. *Applied energy*, 226: 1273–1292, 2018.

[14] M. Poveda-Villalón and R. García-Castro. Extending the saref ontology for building devices and topology. In *LDAC*, pages 16–23, 2018.

[15] K. Hammar, E. O. Wallin, P. Karlberg, and D. Hälleberg. The realestatecore ontology. In *Proceedings of ISWC, Auckland, New Zealand, October 26–30, 2019*, pages 130–145. Springer, 2019.

[16] K. Berkoben, C. El Kaed, and T. Sodorff. A digital buildings ontology for google's real estate. In *ISWC (Demos/Industry)*, pages 392–394, 2020.

[17] L. Ehrlinger and W. Wöß. Towards a definition of knowledge graphs. *SEMANTiCS (Posters, Demos, SuCCESS)*, 48(1–4):2, 2016.

[18] N. Sahlab, S. Kamm, et al. Knowledge graphs as enhancers of intelligent digital twins. In *Proceedings of the 4th IEEE ICPS*, pages 19–24, 2021.

[19] V. Chaudhri, C. Baru, et al. Knowledge graphs: introduction, history and, perspectives. *AI Magazine*, 43(1):17–29, 2022.

[20] L. Chamari, J. van der Weijden, L. Boonstra, S. Hoekstra, E. Petrova, and P. Pauwels. Metadata schema generation for data-driven smart buildings. In *Proceedings of LDAC*, pages 139–150. CEUR-WS. org, 2023.

[21] D. Waterworth, S. Sethuvenkatraman, and Q. Z. Sheng. Building metadata inference using a transducer based language model. *arXiv preprint arXiv:2212.01964*, 2022.

[22] J. Koh, B. Balaji, D. Sengupta, J. McAuley, R. Gupta, and Y. Agarwal. Scrabble: transferrable semi-automated semantic metadata normalization using intermediate representation. In *Proceedings of the 5th Conference on Systems for Built Environments*, pages 11–20, 2018.

[23] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

[24] V. K. Kommineni, B. König-Ries, and S. Samuel. From human experts to machines: An llm supported approach to ontology and knowledge graph construction. *arXiv preprint arXiv:2403.08345*, 2024.

[25] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33:9459–9474, 2020.

[26] D. Mozolevskyi and W. AlShikh. Comparative analysis of retrieval systems in the real world. *arXiv preprint arXiv:2405.02048*, 2024.

[27] D. Edge, H. Trinh, N. Cheng, J. Bradley, A. Chao, A. Mody, S. Truitt, and J. Larson. From local to global: A graph rag approach to query-focused summarization. *arXiv preprint arXiv:2404.16130*, 2024.

[28] R. Angles. The property graph database model. In *Proceedings of the Workshop on Foundations of Data Management*, 2018.

[29] C. Ramonell, R. Chacón, and H. Posada. Knowledge graph-based data integration system for digital twins of built assets. *Automation in Construction*, 156: 105109, 2023.

Integration of Vision Platforms: Generating a Unified 3D Digital Map for Real-time Construction Site Monitoring

J. Shin[1], P. Kim[1], J. Choi[1], Y. Cho[1], G. Lee[2], S. Chi[1]

[1]Department of Civil and Environmental Engineering, Seoul National University, Republic of Korea
[2]Institute of Construction and Environmental Engineering (ICEE), Republic of Korea
wpals113@snu.ac.kr, leophilgu@snu.ac.kr, cjw5428@snu.ac.kr, cyh_i@snu.ac.kr, lgt0427@snu.ac.kr, shchi@snu.ac.kr

Abstract –

Various vision sensor-based platforms generate critical information for real-time construction site monitoring, but their independent applications limit the effective use of this data. This study proposes a framework that integrates three representative vision platforms—Unmanned Aerial Vehicle (UAV), Terrestrial Laser Scanner (TLS), and Closed-Circuit Television (CCTV)—to create a unified digital map. The framework operates as follows: UAVs capture aerial images for overall 3D digital map reconstruction, TLS refines vertical surfaces and structural details through map alignment, and CCTV cameras enable robust real-time 3D object tracking using ray marching and multi-view fusion, resulting in a dynamic digital map. Experimental results demonstrated that the framework enhances resolution of a digital map and improves dynamic object tracking compared to using individual platforms in isolation.

Keywords –

Unified Map, Registration, 3D Reconstruction, Object Tracking, UAV, TLS, CCTV

1 Introduction

In modern construction sites, various vision platforms such as Unmanned Aerial Vehicles (UAVs), Terrestrial Laser Scanners (TLS), and Closed-Circuit Television (CCTV) are actively utilized to enhance site monitoring by generating critical site information. However, Each platform has its strengths and limitations. UAVs, for example, offer high mobility and cost-effective coverage of large areas, making them widely used for 3D digital map generation in construction sites [1]. However, their high-altitude operations and the perpendicular orientation of cameras to the ground, commonly used in UAV surveys, result in low spatial resolution and a lack of information about vertical surfaces in their digital maps [2]. Furthermore, UAVs cannot capture the entire environment at once, which prevents representing real-time movements of dynamic objects on the digital map.

TLS, on the other hand, uses LiDAR and cameras for long-duration scans from the ground, producing high-resolution 3D digital models, particularly for vertical surfaces. While expensive(e.g., the FARO FOCUS S350 costs over $14,500), its benefits make it widely used in construction for applications such as detecting concrete cracks [3] and comparing as-is models with design plans [4]. Recent advancements in Scan-to-BIM technology [5] enable automatic BIM model generation from precise survey data, further driving TLS adoption [6]. However, its ground-level perspective limits the capture of horizontal surfaces like rooftops, and the extensive time and labor required for data collection make it impractical for large-scale mapping or dynamic object analysis [7].

CCTV systems offer a cost-effective solution for long-term, real-time monitoring, capturing sequential data with low operational costs [8]. Rising safety concerns over construction sites accidents have driven their widespread adoption [9]. They are particularly useful for tracking dynamic objects like heavy equipment productivity [10]. However, their limited field of view and issues with occlusion make it challenging to continuously track and monitor dynamic objects, such as materials and workers, in complex and active construction sites [11]. CCTVs, additionally, struggle to generate 3D spatial information about a scene because they are typically installed in scattered locations with fixed monocular cameras.

To summarize, while various vision platforms have been widely used in construction sites, their inherent limitations restrict the usability of the generated information when used independently. To produce comprehensive site information, it is essential to integrate the independent data from each platform into a unified system. UAVs can rapidly capture overall site data without being constrained by site complexity, while TLS can complement this by generating high-resolution models of critical structures and vertical surfaces that require greater detail. Meanwhile, CCTV systems can track dynamic objects, such as workers and equipment.

By integrating data from these platforms, a more complete and accurate digital models of the construction site can be achieved. To this end, this paper proposes a framework for the integration of information from each vision platform into a unified 3D digital map of the site, leading to more comprehensive construction site monitoring in real-time.

2 Literature Review

2.1 Integration of UAV and TLS Data

Many previous studies on aligning UAV and TLS data utilize the Iterative Closest Point (ICP). ICP is an algorithm that iteratively aligns two datasets by finding the closest point for each point and calculating the rotation and translation between them. However, it is highly sensitive to the quality of the initial alignment [12]. The previous studies often perform initial alignment using GPS data [13] or manual alignment via software [14], followed by fine alignment using ICP.

UAV-generated 3D digital maps, created from aerial imagery, and TLS-generated 3D digital models, created from ground-based LiDAR, differ significantly from the distribution and density characteristics of their points. These differences in point distribution can negatively impact alignment accuracy [15]. Moreover, construction sites are typically large-scale environments, resulting in digital maps with a high number of points. This makes it challenging to apply computationally intensive algorithms due to their high computational costs.

In response, this study aims to resolve the differences in point distribution and density characteristics, perform reliable automatic initial alignment, and maximize the performance of ICP fine alignment.

2.2 Camera-based 3D Object Tracking

Camera-based 3D object tracking can be broadly categorized into multi-view geometry-based methods and monocular methods. Multi-view geometry involves installing multiple cameras with overlapping fields of view and estimating relative spatial relationships through feature points to infer 3D coordinates. This approach, however, requires cameras to be installed with similar viewpoints and overlapping coverage, which is challenging to achieve in typical construction sites.

On the other hand, monocular methods primarily rely on homography or depth estimation models. Homography represents the transformation between points on two planes. Assuming the ground is a plane, a homography relationship can be established between the ground plane in the 3D real-world coordinate system and the 2D image plane. The relationship can be calculated if at least four corresponding point pairs are identified. However, since homography is limited to flat planes like the ground, accurately incorporating dynamic information captured by CCTV into a 3D digital map becomes challenging in more complex scenarios involving uneven or non-planer surfaces.

MiDaS [16], a state-of-the-art single-image depth estimation model, offers an alternative approach. MiDaS is trained to generate depth maps from 2D images labeled with depth information. However, monocular depth estimation models can experience significant performance degradation when applied to scenes that differ substantially from those seen during training.

3 Methodology

This study is conducted in three stages: overall 3D map reconstruction, map refinement with registration, and object tracking in a 3D map, as shown in Figure 1.

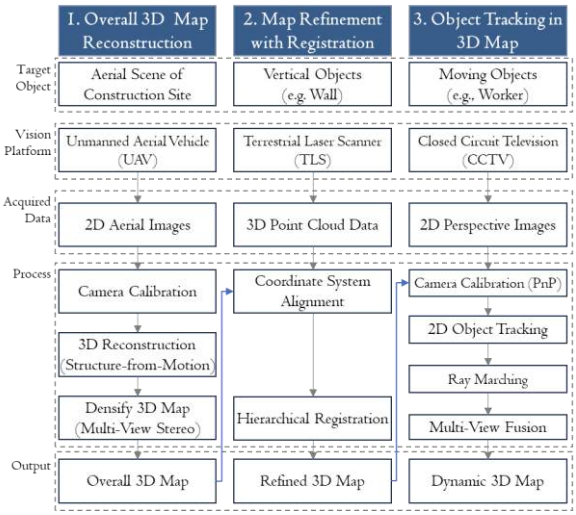


Figure 1. Framework of Vision Platforms Integration

The first stage involves generating an overall 3D map. A UAV is deployed to capture sequential 2D aerial images, and Structure-from-Motion (SfM) [17] is applied to reconstruct a sparse 3D map of the entire site. Subsequently, Patch-based Multi-View Stereo (PMVS) [18] creates a densified overall 3D map.

The second stage focuses on refining the generated 3D digital map. Aerial images alone have limitations in representing high-resolution details and generating 3D models of structures with vertical surfaces, such as walls. To address these limitations, high-resolution 3D point cloud data is collected using a TLS. The UAV-derived 3D digital map and the TLS-based 3D digital models are first coarsely aligned to a unified coordinate system. Subsequently, fine alignment is performed using hierarchical registration. The integration of UAV and TLS data enhances the level of detail and improves the representation of vertical surfaces in the initial map.

The third stage involves tracking objects and

incorporating this information into the 3D map. For CCTV footage, objects are tracked, and their positions in the 3D map are estimated using ray marching. Multi-view fusion is then applied to ensure robustness against tracking failures and maintain the consistency of object information across multiple CCTVs. By integrating these three stages, this study generates a unified 3D digital map that not only provides refined 3D representations but also incorporates dynamic object information.

3.1 Overall 3D Map Reconstruction

The process of generating an overall 3D digital map using UAV involves the following steps: camera calibration is first conducted to estimate the camera matrix, which defines the transformation between 2D image points and their corresponding 3D coordinates. The camera matrix is composed of intrinsic and extrinsic parameters. Intrinsic parameters, such as focal length and principal point, describe the internal characteristics of the camera. Extrinsic parameters define the camera's position and orientation relative to the world coordinate system. The intrinsic parameters are typically determined using a chessboard pattern, while the extrinsic parameters are estimated using Structure from Motion (SfM). SfM is a photogrammetry-based technique that iteratively computes camera poses and generates a sparse 3D map from sequential images, optimizing both pose estimation and mapping accuracy throughout the process. Subsequently, Patch-based Multi-View Stereo (PMVS) is employed to identify overlapping patches between consecutive images and reconstruct them in 3D, thereby densifying the sparse map.

3.2 Map Refinement with Registration

The TLS captures high-resolution 3D data of the surrounding area using LiDAR technology. The 3D data generated by UAV and TLS are represented in different coordinate systems, each with distinct scales and origins. To unify these coordinate systems, GPS data is used to restore the scale of each 3D map to real-world dimensions and align their coordinate system references. Since GPS inherently has an error margin of several meters, an additional hierarchical registration process is required to achieve accurate integration.

3.2.1 Coordinate System Alignment with GPS

Geo-registration converts the overall 3D digital map from the SfM coordinate system into the Earth-Centered Earth-Fixed (ECEF) coordinate system, a global 3D Cartesian coordinate system centered on the Earth. This transformation is essential for accurately positioning and orienting objects on a global scale. The process uses Random Sample Consensus (RANSAC) to determine the optimal similarity transformation between the two coordinate systems.

In contrast, TLS directly captures 3D points via LiDAR, generating maps based on GPS data. To unify digital maps, the TLS-generated 3D map is transformed into the ECEF coordinate system, ensuring coarse alignment with the UAV digital map.

3.2.2 Hierarchical Registration of 3D Digital Maps

Hierarchical registration is used to efficiently align large-scale UAV and TLS point clouds by breaking the process into two stages, as shown in figure 2. For initial alignment, a Fast Point Feature Histogram (FPFH) analyzes the relationships between neighboring points in the point cloud data to extract feature points that encapsulate local geometric information. These features are matched using a RANSAC algorithm to establish an initial transformation. While FPFH provides reliable initial alignment, it incurs high computational costs when applied to large point cloud datasets. To overcome this challenge, both 3D digital maps are down-sampled with voxelization, where the points within each cubic voxel are represented by their average position. This process reduces computational complexity while mitigating differences in point distribution.

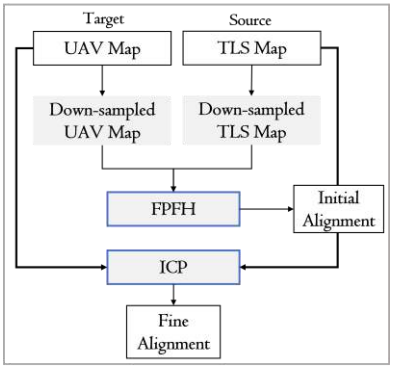


Figure 2. Hierarchical Registration Process

Following this, the ICP algorithm refines the alignment by identifying the closest corresponding points and updating the transformation matrix. Since finding the nearest points is a relatively simple operation and FPFH ensures a reliable initial alignment, ICP is performed on the entire point cloud. By combining FPFH for initial alignment and ICP for fine refinement, this hierarchical approach achieves efficient and accurate automatic alignment of UAV and TLS maps without succumbing to computational overhead or overfitting from differences in point distribution.

3.3 Object Tracking in a 3D Map

The proposed method for object tracking in a 3D map integrates information about dynamic objects into the map for enhanced situational awareness. The process

begins with camera calibration, during which the intrinsic and extrinsic parameters of each camera are estimated. This calibration establishes the relative position and orientation of the CCTV cameras within the registered refined digital map. Subsequently, YOLO-based 2D object detection is applied, identifying objects and determining their positions as 2D pixel coordinates x_{object}. These coordinates are then back-projected into the 3D space using ray marching to estimate the objects' 3D positions X_{object}. To improve the accuracy and consistency of tracking, multi-view fusion is employed, leveraging data from multiple perspectives.

3.3.1 Camera Calibration for CCTV Systems

For CCTV cameras, the extrinsic parameters are determined using the Perspective-n-Points (PnP) [19] algorithm. When four matching pairs between the 2D image and the 3D map are specified, the PnP algorithm calculates the camera's rotation and position. Further optimization is conducted using the Levenberg-Marquardt [20] to refine these estimates. Since the extrinsic parameters of a CCTV camera typically remain fixed after installation, calibration only needs to be performed once during the initial setup of the camera. This ensures that the camera's position and orientation are accurately aligned with the 3D digital map.

3.3.2 2D Object Tracking with Pose Tracker

2D object tracking is conducted using the BoT-SORT [21], a multi-object tracking method that performs object tracking without additional training. BoT-SORT can also extract key points of the detected objects. As illustrated in Figure 3, for a worker object, the 2D point of the object is defined as the center point (marked as a black cross) between the coordinates of the left foot and right foot (marked as red dots).



Figure 3. Definition of the worker's 2D point

3.3.3 Back Projection with Ray Marching

To represent the 2D point of an object into the 3D digital map, back projection is performed. Back projection is the process of reconstructing a 3D world point from a 2D image point.

$$\lambda P^{-1}\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (1)$$

In Equation 1, P is a 3×4 camera matrix, u and v are 2D image points, and X, Y, and Z are 3D world points. λ is a scale parameter that represents the depth from the 2D image point to the corresponding 3D world point. The camera matrix P is obtained through camera calibration, leaving the depth parameter λ as the only unknown variable. In essence, back projection is the process of determining this unknown depth.

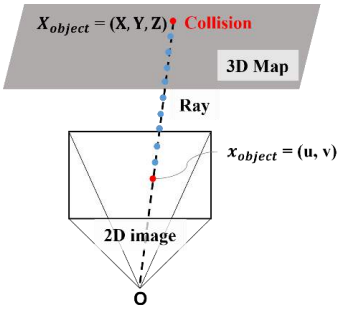


Figure 4. Ray Marching Method

Back projection can be achieved using ray marching. As shown in figure 4, ray marching is a technique that casts a ray from a 2D point x_{object} by sampling 3D points along the ray and searches for the point X_{object} where the ray collides with a surface in 3D space. Because this method directly interacts with the digital map, it requires a high-quality map that is noise-free, high-resolution , and free of voids. Fortunately, the registration of UAV and TLS data eliminates noise, enhancing both the resolution and completeness of the digital map. Ray marching involves sampling candidate 3D points along a ray extending from the camera's origin toward the corresponding 2D image point.

$$r(t) = O + tD \quad (2)$$

Equation 2 reformulates Equation 1 from the perspective of a ray, where O represents the camera origin and D is the direction vector of the ray. Both O and D are derived from the camera matrix P. The parameter t denotes a positive real number depth value, and by incrementing t, the 3D position $r(t)$ of a sample point along the ray is calculated. If the shortest distance between $r(t)$ and the points in the digital map falls below a predefined threshold, the corresponding point in the digital map is determined to be the object's position.

3.3.4 Multi-View Fusion

The 3D position obtained in the previous steps is

managed by assigning a unique track ID for each camera. A single camera, however, loses the ability to track the object if it moves out of the camera's field of view. When the object reappears, a new track ID is assigned, resulting in inconsistency between multiple cameras. Multi-view fusion overcomes this inconsistency in track IDs and enables a robust object tracking system by integrating data from multiple cameras into a unified 3D digital map.

The detailed algorithm for multi-view fusion is explained with an example in Figure 5. At T=3, when a new track ID appears in a camera, the system queries the database to check whether any object is being tracked near the 3D position of the newly detected object. If no such object exists, a new object ID is assigned, and the database begins tracking this object. Subsequently, at T=40, if the same object appears in another camera, track matching is performed to associate the track ID from the camera with the object ID in the database. Once matched, the object's position can be obtained from both CAM0 and CAM1. This ensures that even if a tracking failure occurs in one camera, such as CAM1 at T=40, the object can still be tracked using CAM0.

If, at T=50, all cameras fail to track the object, the system predicts the next position based on the average position change over the previous five timestamps stored in the database. This prediction enables continued tracking, even during temporary tracking failures. Finally, at T=109, if different track IDs are assigned to the same object in other cameras, track matching ensures that the same object ID is used. However, duplicated 3D positions estimated by distinct cameras can be different due to errors in object detection. To address this, the system calculates the tracking error for each 3D position and selects the point with the lower error, reducing redundancy and improving accuracy. The error in 2D object tracking propagates into the 3D position estimation error through back projection. Since the relationship between 2D and 3D points involves a nonlinear transformation, this propagation is described by the Jacobian matrix J.

$$cov(X_w) = J \times cov(x_{pixel}) \times J^T \quad (3)$$

In Equation 3, $cov(X_w)$ represents the covariance of the 3D point, indicating the 3D object tracking error, while $cov(x_{pixel})$ represents the covariance of the 2D point, characterizing the error in 2D object tracking. This systematic error propagation ensures that the system accounts for inaccuracies in both 2D and 3D tracking, leading to more precise object localization.

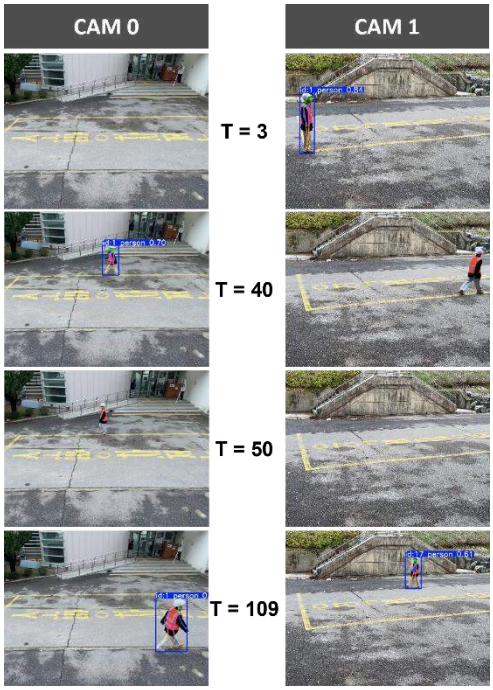


Figure 5. Example of Multi-view Fusion

4 Experiments

The experiments took place in a lane-marked area at Seoul National University. To simulate a multi-view CCTV system, two iPhone 12s were installed with overlapping fields of view, and their resolution was reduced to 360×640 to match typical CCTV settings. A DJI Mavic 3E UAV and a FARO Focus M Laser Scanner (TLS) were used for 3D map generation. The study consisted of two experiments.

- UAV and TLS Map Alignment

This experiment assessed the error in digital map alignment and the improvement in resolution for vertical surfaces. The evaluation metric for alignment was the Inlier Root Mean Square Error (Inlier RMSE) in meters. The metric was calculated by finding the nearest UAV-based neighbor for each transformed TLS point and calculating the RMSE. Points exceeding a set distance threshold were classified as outliers and excluded, ensuring a more precise alignment assessment. A voxelization size of 0.5m was used, with inlier thresholds of 0.5m, 1.0m, and 1.5m.

- Multi-View Fusion-based CCTV Object Tracking

This experiment evaluated the proposed ray marching-based object tracking method against baseline approaches using homography and the MiDaS depth estimation model. Since MiDaS generates only relative depth maps, four reference points were selected to obtain true depth values. A cubic spline interpolation was then

applied to convert the relative depth map into an absolute depth map for the analysis. Moreover, the effectiveness of multi-view fusion was validated. Object tracking performance was assessed using the mean and RMSE of the Absolute Trajectory Error (ATE). The mean ATE is a standard measure of position estimation accuracy, while RMSE is more sensitive to large errors, providing a comprehensive evaluation of tracking.

The ground truth trajectory for evaluation was defined as the yellow rectangle shown in Figure 6. The ATE was calculated as the distance between the estimated 3D points and the closest ground truth points.

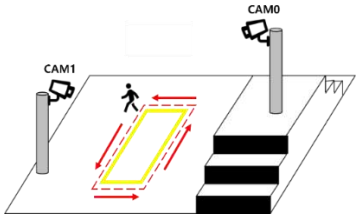


Figure 6. Testbed for Object Tracking

4.1 Evaluation of Detailed 3D Map

To validate the alignment accuracy, an ablation study, as shown in Table 1, compared the proposed hierarchical method against individual alignment algorithms. The best RMSE of 0.101m at a 0.5m inlier threshold was achieved using $FPFH_{down}$ with downsampling for initial alignment, followed by ICP. This highlights their complementary strengths: FPFH captures global geometric features for coarse alignment, while ICP refines precision during fine alignment.

The UAV-based digital map covers $150 \times 200m^2$ with 10.8 million points, while the TLS model spans $40 \times 76m^2$ with 74.2 million points. Table 1 analyzes runtime efficiency. Without downsampling, ICP successfully aligned in 4:36 minutes, whereas FPFH failed due to high computational costs. By applying 0.5m voxelization-based downsampling, FPFH completed in 3:12 minutes, approximately reducing UAV points to 207,000 (52 × reduction) and TLS points to 64,000 (1,188× reduction). Since FPFH iterates over all points, downsampling reduced computational costs by over 1,000×, confirming its necessity.

Table 1 Inlier RMSE with Varying Inlier Thresholds and Runtime of Registration

	0.5m	$1.0m$	$1.5m$	Time(m)
ICP	0.300	0.610	0.920	4:36
$FPFH_{down}$	0.124	0.187	0.253	3:12
Proposed	**0.101**	**0.172**	**0.238**	7:48

Figure 7 illustrates the benefits of the proposed method. In the first row, vertical surfaces missing in the UAV digital map were completed through alignment, showing the ability to enhance map completeness. In the second row, the resolution of 3D models for structures of interest was significantly improved, increasing their utility for applications requiring high-quality 3D digital models, such as construction error analysis. In representing the wall structure, the UAV-based digital map had an average density of $124\ pts/m^2$. In contrast, the aligned digital map achieved an average density of $11{,}066\ pts/m^2$, almost 100-fold density increase. This substantial improvement demonstrates the enhanced capability of the proposed method to generate refined 3D representations of structures.

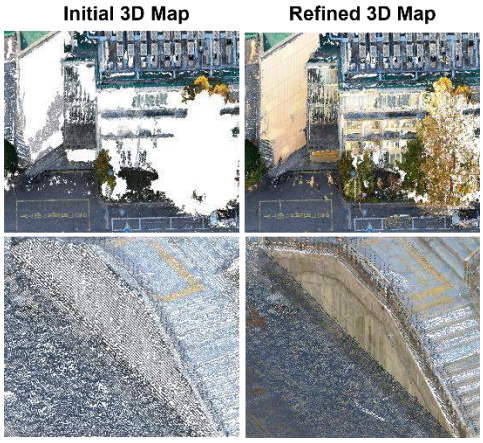


Figure 7. Qualitative Result of Registration

4.2 Evaluation of Object Tracking

Figure 8 illustrates the qualitative results of object tracking. The color gradient represents the progression of timestamps, with red indicating earlier timestamps and yellow indicating later ones. Ray marching exhibited a few errors near the bottom edge, while homography showed some errors along the top edge. Both the ray marching and homography methods, however, generally produced object trajectories that were closely aligned with the yellow rectangular ground truth path. In contrast, the depth estimation model demonstrated a trajectory that was relatively more deviated from the ground truth.

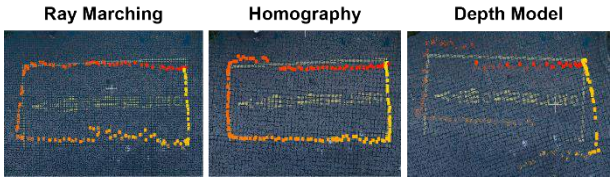


Figure 8. Qualitative Result of Object Tracking

As shown in Table 2, all methods achieved a mean ATE within 0.6m, indicating estimation errors in the tens-of-centimeters range. Among them, ray marching

demonstrated the highest accuracy. Homography showed larger errors, likely due to its assumption of a perfectly planar ground surface, which does not fully reflect real-world conditions. Ray marching, by directly interacting with the refined digital map, provides more precise estimation, leading to lower errors. Meanwhile, the depth model showed relatively higher estimation errors, highlighting the limitations of single-camera-based depth estimation models, particularly when no elaborate finetuning is performed for the specific site.

Table 2 ATE (meters) for 3D Object Tracking Results

	Ray Marching	Homography	Depth Model
Mean	**0.204**	0.367	0.595
RMSE	**0.285**	0.382	0.717

Figure 9 presents the results of multi-view fusion. Blue dots represent object positions from CAM0, red dots correspond to those from CAM1, and green dots indicate predicted positions due to tracking failure. Before applying multi-view fusion, overlapping areas contained redundant red and blue dots. After fusion, these redundancies were eliminated by selecting the point with the lower error. Additionally, green dots were appropriately placed to compensate tracking failures, ensuring continuity in the object's trajectory. In summary, multi-view fusion effectively mitigates camera tracking failures, enabling consistent and robust object tracking.

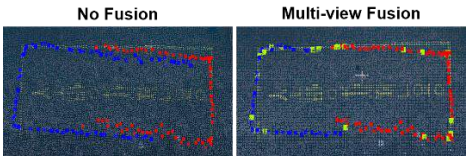


Figure 9. Validation Result of Multi-view Fusion

5 Limitations and Future Study

The experiment in this study was conducted in a controlled environment to accurately validate and compare the proposed methodology. However, real-world construction sites involve multiple objects and frequent occlusions, posing additional challenges in object tracking. As illustrated in Figure 10, when large objects obstruct the camera's view, ray marching may collide with obstacles before reaching the ground, resulting in inaccurate positioning. Furthermore, when multiple objects move in close proximity, mutual occlusion disrupts 2D tracking. While multi-view fusion helps mitigate tracking failures when objects are spaced apart, it becomes less effective in cases of close overlap, as shown in the right image of Figure 10, leading to track matching failures.

To address these challenges and enhance scalability, further enhancements are needed, such as integrating a depth estimation model for robust and seamless object tracking, along with further validation across diverse construction site scenarios.



Figure 10. Challenging Cases in Real-world Construction Sites

6 Conclusion

This study integrated commonly used vision platforms in construction sites—UAVs, TLS, and CCTV—to address their individual limitations. The proposed framework generated a unified 3D digital map, enabling seamless information sharing across platforms. A UAV provided an overall 3D digital map, while an alignment algorithm incorporating TLS enhanced resolution and vertical details. The refined map also served as a foundation for a robust CCTV-based 3D object tracking system, integrating dynamic object information into the 3D map.

Specifically, this study proposed a method to automatically align UAV and TLS maps using GPS, FPFH, and ICP in a step-by-step manner, effectively reducing computational costs for large-scale datasets and mitigating characteristic differences between heterogeneous data. In addition, the use of the ray marching technique and multi-view fusion enabled interaction with the refined 3D digital map, improving the robustness of CCTV-based 3D object tracking and ensuring consistency across cameras. The first experimental result demonstrated that the combination of FPFH and ICP significantly improved alignment accuracy compared to single alignment algorithms. The result of another experiment showed that the ray marching method achieved the highest accuracy in 3D object tracking.

In conclusion, this study highlights the significance of integrating vision platforms through a unified 3D digital map, improving the usability of site monitoring data. Nevertheless, challenges remain for the future research. A more robust object tracking system is needed to handle occlusions in complex construction environments, and further validation is required through real-world construction sites testing.

Acknowledgment

This research was conducted with the support of the

"National R&D Project for Smart Construction Technology (RS-2020-KA156875)" funded by the Korea Agency for Infrastructure Technology Advancement under the Ministry of Land Infrastructure and Transport. This work was also supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2023-00241758).

References

[1] Z. Ma and S. Liu. A review of 3D reconstruction techniques in civil engineering and their applications. Advanced Engineering Informatics, 37:163–174, 2018.

[2] B. Draeyer and C. Strecha. White paper: How accurate are UAV surveying methods. Pix4D White Paper, 4, 2014.

[3] M. Zhong, L. Sui, Z. Wang, and D. Hu. Pavement crack detection from mobile laser scanning point clouds using a time grid. Sensors (Switzerland), 20(15):1–20, 2020.

[4] E. B. Anil, P. Tang, B. Akinci, and D. Huber. Deviation analysis method for the assessment of the quality of the as-is building information models generated from point cloud data. Automation in Construction, 35:507–516, 2013.

[5] Q. Wang, J. Guo, and M.-K. Kim. An application-oriented Scan-to-BIM framework. Remote Sensing, 11(3):365, 2019.

[6] C. Wu, Y. Yuan, Y. Tang, and B. Tian. Application of terrestrial laser scanning (TLS) in the architecture, engineering and construction (AEC) industry. Sensors, 22(1):265, 2022.

[7] X. M. Garcia-Cruz, O. Y. Sergiyenko, V. Tyrsa, M. Rivas-Lopez, D. Hernandez-Balbuena, J. C. Rodriguez-Quiñonez, and P. Mercorelli. Optimization of 3D laser scanning speed by use of combined variable step. Optics and Lasers in Engineering, 54:141–151, 2014.

[8] X. Yang, et al. Computer-aided optimization of surveillance cameras placement on construction sites. Computer-Aided Civil and Infrastructure Engineering, 33(12):1110–1126, 2018.

[9] J. Seong, H.-S. Kim, and H.-J. Jung. The detection system for a danger state of a collision between construction equipment and workers using fixed CCTV on construction sites. Sensors, 23(20):8371, 2023.

[10] J. Kim and S. Chi. Action recognition of earthmoving excavators based on sequential pattern analysis of visual features and operation cycles. Automation in Construction, 104:255–264, 2019.

[11] Q. Wang, H. Liu, W. Peng, C. Tian, and C. Li. A vision-based approach for detecting occluded objects in construction sites. Neural Computing and Applications, 36(18):10825–10837, 2024.

[12] S. Rusinkiewicz and M. Levoy. Efficient variants of the ICP algorithm. Proceedings third international conference on 3-D digital imaging and modelling, 145-152, IEEE, 2001.

[13] X. Tong, X. Liu, P. Chen, S. Liu, K. Luan, L. Li, S. Liu, X. Liu, H. Xie, Y. Jin, and Z. Hong. Integration of UAV-based photogrammetry and terrestrial laser scanning for the three-dimensional mapping and monitoring of open-pit mine areas. Remote Sensing, 7(6), 6635–6662, 2015.

[14] I. Aicardi, P. Dabove, A. M. Lingua, and M. Piras. Integration between TLS and UAV photogrammetry techniques for forestry applications. IForest, 10(1), 41–47, 2017.

[15] J. Šašak, M. Gallay, J. Kaňuk, J. Hofierka, and J. Minár. Combined use of terrestrial laser scanning and UAV photogrammetry in mapping alpine terrain. Remote Sensing, 11(18), 2019.

[16] R. Ranftl, K. Lasinger, D. Hafner, K. Schindler, and V. Koltun. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(3):1623–1637, 2020.

[17] J. L. Schonberger and J. M. Frahm. Structure-from-motion revisited. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4104–4113, 2016.

[18] S. Yao, H. AliAkbarpour, G. Seetharaman, and K. Palaniappan. 3D patch-based multi-view stereo for high-resolution imagery. In Proceedings of Geospatial Informatics, Motion Imagery, and Network Analytics VIII, 2018.

[19] J. A. Hesch and S. I. Roumeliotis. A direct least-squares (DLS) method for PnP. In Proceedings of the International Conference on Computer Vision, pages 383–390, 2011.

[20] A. Ranganathan. The Levenberg-Marquardt algorithm. Tutorial on LM Algorithm, 11(1):101–110, 2004.

[21] N. Aharon, R. Orfaig, and B. Z. Bobrovsky. BoT-SORT: Robust associations multi-pedestrian tracking. arXiv Preprint arXiv:2206.14651, 2022.

Developing a Digital Twin-based Framework for Construction Fire Hazard Recognition Training

Kexin Liu[1], Mohamed Sabek[1], Gaang Lee[1], Max Kinateder[2], Vicente A. Gonzalez[1]

[1]Department of Civil & Environmental Engineering, University of Alberta, Canada
[2]Construction Research Centre, National Research Council Canada, Canada
kexin10@ualberta.ca, sabek@ualberta.ca, gaang@ualberta.ca, Max.Kinateder@nrc-cnrc.gc.ca, vagonzal@ualberta.ca

Abstract –

Construction sites are dynamic environments where fire hazards pose significant safety risks. Effective hazard recognition is the key step to prevent such hazards. While existing research has developed various training methods, limited studies focus on fire hazards in construction. Moreover, most approaches fail to adapt to evolving site conditions. To address these gaps, this study proposes a digital twin (DT)-based framework for dynamic fire hazard recognition training, with current validation progress presented. This framework integrates 360° imagery, 3D Gaussian Splatting, and Immersive Virtual Reality (IVR) to create an adaptive and interactive training environment. A user-centered training approach is designed to enhance personalized learning by incorporating individual trainee profiles and situation awareness (SA) assessments. Additionally, a cloud-based data system enables long-term tracking and scenario updates based on historical hazard data and trainee performance. This modular conceptual framework provides a foundation and guideline for future research on dynamic fire hazard recognition training. Further work will focus on framework validation through pilot studies in real construction settings to assess training effectiveness and usability.

Keywords –

Construction Fire; Hazard Recognition; Safety Training; Digital Twin; Situation Awareness

1 Introduction

Construction sites face continuous challenges in hazard prevention due to dynamic changing environments, diverse stakeholders, and complex interactions between people, equipment, and materials. Fire hazards, though less frequently than the "fatal four" hazards (i.e., falls, electrocutions, struck-by, and caught-in/between), often result in severe consequences, including risks to human life, significant property damage, and project delays. From 2016 to 2020, an annual average of 4,300 fires on construction sites were reported in the United States, causing five deaths, 62 injuries, and USD 376 million in property damage [1].

Fires can spread rapidly in construction sites with lots of combustibles, making early-stage hazard recognition critical [2]. With advancements in sensors and vision-based systems, such as computer vision and machine learning, automated detection systems are increasingly being developed to monitor, identify, and predict construction hazards [3]. While these systems provide valuable support, SA and human judgment remain indispensable [4]. SA involves three levels: perception (Level 1), perceiving the current state of the environment; comprehension (Level 2), understanding the meaning of the situation; and projection, anticipating potential future developments[5]. These abilities allow experts to make timely and accurate decisions and adapt to unexpected states [5]. Construction personnel (e.g., construction workers and managers) equipped with well-developed SA can adapt to unexpected situations, interpret subtle contextual signals, and make informed decisions that automated systems may not be able to address [6]. Unfortunately, individuals often struggle to identify hazards in dynamic and complex environments due to high cognitive demands and a lack of SA, stemming from the complexity of fully recognizing and predicting fire, a complicated process involving many factors, and the absence of structured training approaches [2]. Traditional training methods, such as classroom-based presentations, often adopt a one-size-fits-all approach, failing to engage workers and overlooking individual learning needs and weaknesses [7][8].

Immersive technologies, such as virtual reality (VR) and augmented reality (AR), have proven to be effective safety training tools, providing immersive and engaging learning experiences without exposing individuals to risks [9]. Despite their potential, existing studies on construction hazard training tend to focus on either broad categories of construction hazards or high-frequency hazards, such as slips and falls [10]. Research suggests

that workers often recall familiar hazards while overlooking less obvious but equally critical hazards, such as fire hazards [11]. Moreover, since each construction project is unique and evolves through different phases, the fire hazards workers encounter can vary significantly. For instance, the foundation phase involves fire hazards from flammable fuels for machinery, while the structure phase faces hazards from welding, cutting, and improper material storage. Additionally, tailoring training to the unique hazards of each construction site has proven to be more effective, as varying hazards impose different cognitive demands [12].

Collectively, there are gaps in (1) the lack of training systems specifically designed to improve SA in construction fire hazard recognition and (2) the absence of systems capable of dynamically adapting training focus to specific construction site conditions and evolving stages. To bridge these gaps and guide future research in this direction, this paper proposes a DT-based framework for fire hazard recognition training and presents the current progress in its validation.

2 Literature Review

2.1 Construction Hazard Recognition Training

The proposed training framework aims to improve construction personnel's SA ability to recognize and prevent fire hazards, a process influenced by various factors. Understanding these factors is vital for developing effective training programs. Environmental conditions, such as low visibility due to darkness or dust during construction, are one of the key factors [13], along with distractions prevalent in construction settings, which can hinder workers' focus [14]. Workers' physical or mental state (e.g., fatigue) also affects hazard recognition performance [6][15]. Furthermore, individual characteristics, such as age [16], personality traits [17], safety knowledge [18], and experience [19], affect how workers perceive, comprehend, and project potential risks. Additionally, cues provided during training, such as visual and question cues, affect trainees' performance [19]. Therefore, it is extremely important to consider the environmental variations trainees will encounter in the field, strategy, and trainees' individual features and needs when developing training programs.

Albert et al. [2] developed the first high-fidelity Building Information Modeling (BIM)-based augmented virtual environment. They categorized construction hazards into ten energy sources and incorporated cognitive retrieval mnemonics to improve long-term cue retention. Building on this work, Jeelani et al. [8] developed a personalized training strategy using eye-tracking data for individualized performance feedback and self-diagnosis correction. However, 3D BIM models face challenges in presenting real-world environments, as they offer an artificial, computer-generated representation of construction sites, potentially reducing training effectiveness [20]. Eiris et al. [21] observed that participants identified more hazards in BIM-based virtual environments than in 360 images of real construction sites due to clearer visuals. However, 360 images better reflected real-site conditions, raising concerns about the applicability of training methods that do not closely align with real-world environments. To address these limitations, researchers have adopted augmented 360 images, enhanced with cues (e.g., visual, audio) or instructions, as training environments, known as augmented virtuality, delivered by screen-based or VR head-mounted devices (HMDs) [22][23]. Given the page limitation, Table 1 summarizes key research in this field, detailing hazard type, training material and environment, cues, interaction, and assessment metrics. The literature review highlights a lack of focus on less frequent hazards, such as fire, as discussed in the Introduction section, as well as the insufficient integration of SA as training objectives to further improve hazard recognition abilities.

Table 1. Key studies on construction hazard recognition training

	Hazard Type	Training Material and Environment	Cues	Interaction	Assessment Metrics
[2]	Various types	BIM models augmented by experts and with construction images	Visual	Screen-based Devices	Hazard recognition index
[8]	Various types	Static real-world images annotated by experts	Visual	Screen-based Devices	Hazard-Recognition Level
[22]	Various types	Augmented stereo-panoramic videos and virtual models	Visual	Screen-based Devices	Hazard Recognition Score
[24]	Various types	BIM models enriched with avatars	Question	VR HMDs	The time used; the number of hazards identified
[23]	Fall hazard	Augmented 360-degree Panorama	Visual and audio	VR HMDs	Hazard Identification Index (HII)

[25]	Various types	construction site simulations with hazards	Visual and audio	VR HMDs	Eye-tracking metrics; hazard recognition accuracy
[26]	Fatal four	BIM model supplemented with regulations	Visual	VR HMDs	HII; Eye-tracking metrics

2.2 DT Frameworks for Construction Safety

In the architecture, engineering, and construction (AEC) industry, a digital twin is typically defined as a dynamic, real-time digital replica of physical assets, processes, and systems to enhance the management and optimization of construction and facility operations [27]. A DT system should include three key components: the physical artifact, the digital counterpart, and the bidirectional data exchange between them [28]. Unlike static models such as BIM, digital models, or digital shadows, which may only offer one-way updates, a true DT is characterized by its ability to synchronize with real-world changes in real time or near real time [27].

In the context of construction safety, DT applications are emerging but remain underdeveloped. In the area of hazard detection, DTs have been used to dynamically track workers' locations and detect risk events, such as falls, in real time [29]. For training, Vahdatikhaki et al. [30] developed a DT-based system for excavator training, allowing instructors to review trainee performance and provide targeted feedback dynamically. Harichandran et al. [32] further proposed a DT-based framework for dynamically updating VR safety training scenarios based on construction progress, enabling trainees to perceive and familiarize themselves with potential safety risks before entering the actual site.

A major advantage of DT-based training lies in its potential to improve SA [31]. In manufacturing, for instance, DT-enhanced training has been shown to improve operators' responsiveness and anticipation of risks by aligning VE with live system conditions [33]. DT can deliver higher-fidelity contextual information by dynamically reflecting environmental and operational changes, which is particularly valuable in fast-changing construction sites. Applying this concept in construction safety training, especially hazard recognition training, allows for site-specific interventions to support continuous hazard awareness and safer decision-making under realistic conditions. Despite its potential, research on integrating DT into construction safety training for enhancing SA remains limited and fragmented [31]. Furthermore, the application of DT in construction hazard recognition training has been largely unexplored.

3 System Architecture

To address the identified gaps, a conceptual framework was developed following a structured methodology for system architectures [34]. The primary objective of this framework is to provide context-specific training for construction personnel, aiming to improve their SA while strengthening their ability to effectively recognize and respond to fire hazards. Based on a comprehensive review of the literature, the system's key objectives and functions are outlined below, with detailed explanations in the following subsections.

- Objective 1: Enable trainees to perceive training environments realistically, ensuring their reactions and decisions reflect actual site conditions. Function 1: Capture real construction site environments with detailed representations of site objects and spatial arrangements.
- Objective 2: Provide context-specific training scenarios that adapt to evolving project phases. Function 2: Dynamically update virtual environments from real construction site data based on project checkpoints.
- Objective 3: Deliver personalized feedback and reinforcement training to foster targeted skill improvement. Function 3: Provide pre- and post-training assessments and a dynamic feedback mechanism to adapt training content (e.g., scenarios and strategies).
- Objective 4: Record identified construction site conditions for fire hazards and trainee performance metrics to support continuous improvement. Function 4: Provide an accessible database to store and manage historical data.

Accordingly, a system architecture has been proposed. As illustrated in Figure 1, it consists of four major components: (1) data acquisition and (2) data processing and integration, which together address objective 1, (3) user-centric training, which addresses objective 3, and (4) dynamic updating with cloud storage, which address objective 2 and 4. These components form a continuous loop, enabling the system to deliver effective and adaptive fire hazard recognition training. As this framework is based on existing DT frameworks, it can be integrated as a detailed subcomponent within broader DT-based construction safety frameworks. The following sections provide a detailed explanation of each function and current validation progress.

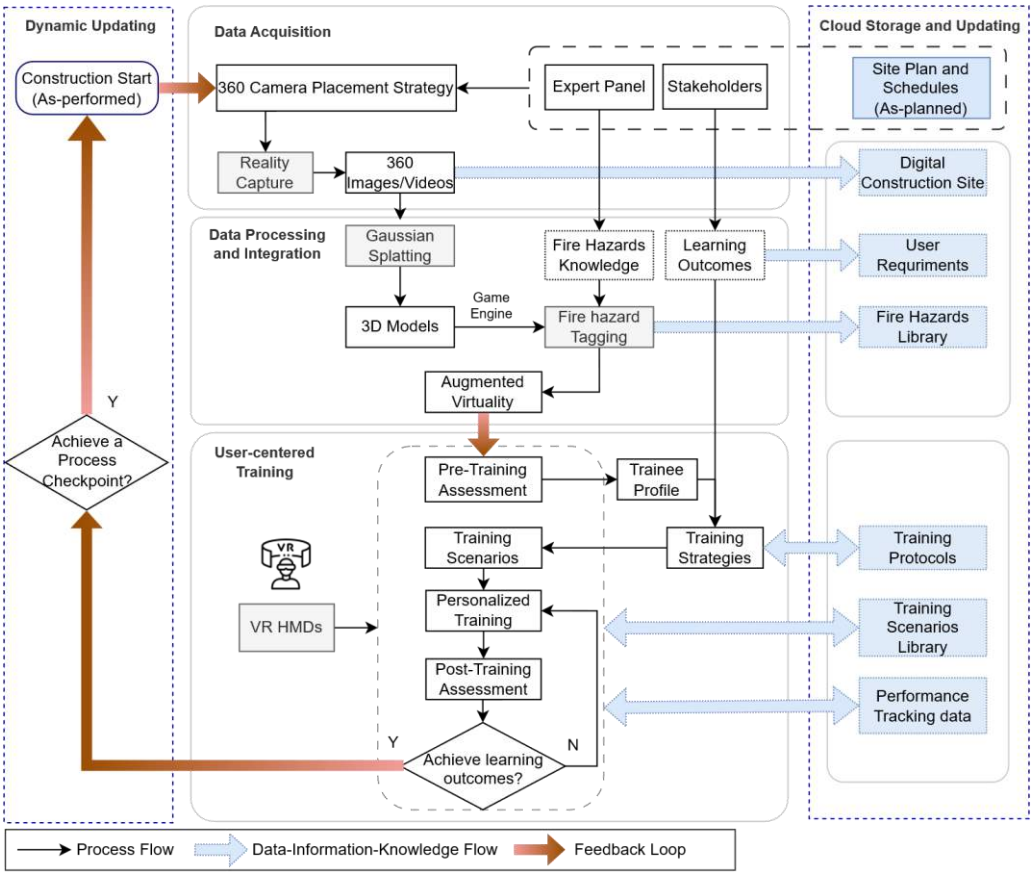


Figure 1. The system architecture of DT-based construction fire hazard recognition training system

3.1 Data Acquisition

Given the limited historical data (e.g., incident reports) and guidelines on construction fire hazards, obtaining input from stakeholders, such as safety managers and construction workers, and fire hazard experts, is the critical first step in building a reliable system (as shown in the dashed box in Figure 1). However, as more data becomes available, Artificial Intelligence-driven methods can be integrated to enhance autonomy over time. Regular with stakeholders, combined with analysis of site plans and schedules, help identify high fire risk zones, where fire hazards are likely to occur and relevant to training (e.g., areas requiring identification of hidden hazards such as electrical wiring). This guides the development of comprehensive camera placement plans, ensuring effective data collection aligned with real-world safety requirements and training needs. Additionally, fire hazards category and learning outcomes can be identified to further tailor training to specific objectives.

This paper proposes the use of 360° cameras to capture construction environments for training. Compared to synthetic, BIM model-based training scenarios, those captured from real construction site data, such as 360° imagery, are more effective in achieving training outcomes, as highlighted in the literature review section. However, recognizing the limitations of 360° images/videos in navigation and interactivity, this paper introduces 3D Gaussian Splatting (3DGS) [35] as a method to convert these images/videos into 3D models (detailed in Section 3.2). In a recent study, Rüter et al.[36] demonstrated the potential of 3DGS for construction site visualization, highlighting that it enables fast, high-quality 3D model generation while eliminating the need for time-consuming methods such as laser scanning. Given its efficiency, cost-effectiveness, and ease of deployment, the process of converting 360° images/videos into 3DGS was selected for this research.

To maximize environmental representation while avoiding obstructions and distortions caused by low ceilings or personnel, 360° cameras are proposed to be strategically placed across the construction site. Medium-height placements are prioritized to achieve optimal results. To ensure consistency in subsequent cube map conversions, the camera can be maintained in a vertical position. Frames can be extracted from recorded videos at regular intervals, producing a uniform and structured visual dataset. Using a dual-lens 360° camera, this method can generate equirectangular images, and rectangular projections of spherical data in a 2:1 aspect

ratio, ensuring full site coverage with minimal effort [37]. This approach reduces the need for precise camera framing while providing detailed, comprehensive captures of construction environments. Figure 2 shows an example of the generated data.

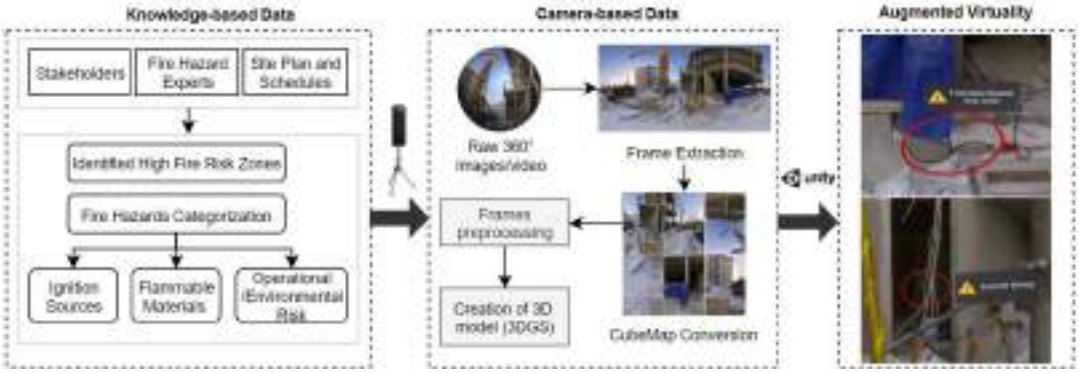


Figure 2. Illustration of data acquisition, processing and integration flow for the proposed system

3.2 Data Processing and Integration

In this phase, the collected raw 360 images or videos can be converted into 3D models, while input gathered from fire hazard experts is used to augment the generated 3D model in game engines (e.g., Unity 3D), to generate training scenarios. Figure 2 illustrates the data processing and integration process of knowledge-based and camera-based data, with an example provided to demonstrate current progress in the development effort. The acquired equirectangular images are processed for 3D model generation and training applications. First, to address the distortion inherent in spherical images captured with fisheye lenses, equirectangular images are converted into cube maps. This process projects the spherical data onto six cube faces, straightening perspective lines and removing distortions [35]. These cube maps, compatible with 3D modelling tools, enable further processing [35]. To optimize dataset quality and efficiency, only the left, right, and front views, which are central to the camera and environment, are retained [37]. Less relevant views, such as the top and bottom faces, are excluded to streamline the data and focus on critical environmental details. Furthermore, to ensure the usability of the curated cube map images, frames with poor lighting or obstructions are removed. Additional angles can be extracted using software (e.g., Insta360 Studio) that allows for frame extraction, stitching, and refinement of 360° images, ensuring dataset completeness and quality [37]. These processed images are then put into tools supporting gaussian splatting, such as Jawset Postshot and Polycam, to generate high-fidelity 3D point clouds.

To augment the generated 3D virtual environments for creating training scenarios, game engines are commonly used to enable interactivity and functionality [23]. In the development progress of this paper, Unity 3D, a widely used game engine known for its flexibility, was used for tagging fire hazards and enhancing interactivity. By using knowledge obtained from fire hazard experts and fire safety regulations, fire hazards in specific scenes can be identified and annotated. For instance, as shown on the right of Figure 2, exposed wiring and flammable materials near the cable that could potentially cause a fire are annotated with a red circle. The augmented scenarios form the foundation for further training scenario design. This structured and refined data acquisition, processing, and integration pipeline enables the system to deliver tailored, context-specific fire hazard recognition training.

3.3 User-centered Training System

The proposed system emphasizes a customizable and user-centered VR training framework, guided by four essential dimensions, learner specifics, context, representation, and pedagogy, adapted from established frameworks for immersive virtual training [38]. Learner specifics serve as the foundation, including trainees' roles and baseline SA. As shown in Figure 3(a), users can select their role, such as safety manager or construction worker, to tailor the training content based on their responsibilities. Baseline testing is typically conducted as a pre-training assessment, using a combination of questionnaires and virtual environments exploration [2]. During these explorations, trainees navigate scenarios with hazards but without any cues, allowing the system to evaluate their baseline hazard recognition abilities [2] [24]. Metrics such as the hazard recognition index and hazard identification index, summarized in Table 1, are commonly used to assess perception (Level 1 SA) by counting the number of hazards recognized. However, these metrics provide limited evaluations of comprehension (Level 2 SA) and projection (Level 3 SA). Given the visual nature of fire hazard recognition, data such as eye-tracking metrics could be incorporated to assess different levels of SA [4].

Context-specific scenarios developed in the data processing and integration phase form the foundation for

interactive design guided by pedagogical strategies. Representation of the training content is achieved through immersive VR simulations delivered via HMDs, recreating realistic and interactive construction site environments, ensuring that trainees exhibit behaviors and reactions approximately similar to those they would have in real-world scenarios [23]. For pedagogical strategies, the system incorporates scenario-based learning and interactive learning approaches widely adopted in existing research [22-26]. For example, as shown in Figure 3(d), hazards such as exposed wiring are annotated within the virtual environment. When the trainee points and clicks on the hotspot using a VR controller, the corresponding fire hazard label is displayed, as shown in Figure 3(e). This interaction helps develop fire hazard perception (Level 1 SA), which involves identifying fire hazard elements in the workplace. To further develop comprehension (Level 2), understanding that exposed wiring poses a fire hazard, and projection (Level 3), predicting the consequences of such a hazard and proposing interventions, detailed instructions are provided when the trainee clicks the "Instruction" button, as shown in Figure 3(f). However, more specific training strategies need to be developed to align with specific training objectives, and evaluating the effectiveness of different training strategies requires further well-designed experiments. Currently, there is limited research exploring how different training protocols affect trainees' performance [24]. Following each training session, a post-training assessment will be conducted using metrics in pre-training assessment, evaluating the trainee's performance on SA across all three levels. If it does not meet the learning outcomes, tailored reinforcement training will be conducted.



Figure 3. An Example of a VR Training Scenario in the Proposed System

3.4 Dynamic Training Scenarios Updating and Data Storage

The key role of DT in this system is to (1) dynamically update training scenarios, (2) track training performance over time, and (3) store historical data for continuous improvement. For scenario updating, the frequency depends on the specific project needs, as real-time data input is not required for training based on recognition site objects [32]. Instead, new training scenarios are updated when obvious changes occur in the construction layout or activities, such as the introduction of new materials or equipment. Milestones are typically a point when significant changes occur at the construction site. At each milestone, the system initiates the data acquisition phase to capture updated site environments. The triggering process can be initiated manually by construction managers who identify such changes or automatically through sensors integrated into the project that detect alterations in site conditions. Consequently, this data enters the data processing and integration procedure, identifying new fire hazards and ensuring training scenarios remain relevant and aligned with real-world requirements. The updating mechanism uses both the as-planned construction schedule and real-site conditions to determine when a milestone is reached. Regarding training performance, most existing construction training programs focus only on short-term learning without assessing how well trainees retain and apply knowledge over time [22]. By integrating DT, the proposed system stores trainees' performance data after every session. Over time, this data can be analyzed alongside construction phases and training frequencies to support skill development and reinforcement [23]. For historical data storage, the system uses a cloud storage platform (e.g., Amazon S3) to securely store identified fire hazards, established training scenarios, and training performance. Stored digital representations of construction sites can support not only fire but broader hazards, while also serving as training data for AI-based automated hazard detection to assist human detection. Notably, for smaller projects with limited data requirements, local or hybrid storage can be used as an alternative to cloud storage based on specific needs.

4 Conclusions and Future Work

This study presents a DT-based framework for construction fire hazard recognition training, designed to develop and improve trainees' SA. This framework contributes to the body of knowledge in construction fire safety by introducing a rapid workflow for creating augmented virtual construction environments, integrating SA into user-centered fire hazard recognition training, and integrating DT technology for dynamic training scenarios updating and long-term performance tracking. The proposed system has the potential to help trainees to better adapt to dynamic construction sites, develop and improve the SA of construction fire hazards, and ultimately reduce the risk of construction fires. Additionally, the modular structure of the framework allows for gradual adoption, reducing the need for a full-scale investment upfront. While the proposed system shows significant promise, several areas require further development. First, stakeholder engagement is needed to validate system requirements. Second, the framework's performance and scalability need to be validated through real-world implementation across various construction scenarios. Future work will focus on conducting pilot studies to evaluate training effectiveness, user experience, and SA improvements.

Acknowledgement

This research has been supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) via the Discovery Grant program (Grant No. RGPIN-2023-04239). The authors gratefully acknowledge NSERC's support for this research.

References

[1] Richard Campbell, Fires in structures under construction, 2022. Accessed: Nov. 21, 2023. [Online]. Available: www.nfpa.org/education-and-research/building-and-life-safety/construction-safety

[2] A. Albert, M. R. Hallowell, B. Kleiner, A. Chen, and M. Golparvar-Fard, Enhancing construction hazard recognition with high-fidelity augmented virtuality, *J Constr Eng Manag*, vol. 140, no. 7, Jul. 2014.

[3] S. Chen, F. Dong, and K. Demachi, Hybrid visual information analysis for on-site occupational hazards identification: a case study on stairway safety, *Saf Sci*, vol. 159, Mar. 2023.

[4] Y. Luo, Q. Yang, J. Seo, and S. Ahn, Theoretical framework for utilizing eye-tracking data to understand the cognitive mechanism of situational awareness in construction hazard recognition, *J Constr Eng Manag*, vol. 40. 2024.

[5] M. R. Endsley, Toward a theory of situation awareness in dynamic systems, *Human Factor*, 1995.

[6] A. Ibrahim, C. Nnaji, M. Namian, A. Koh, and U. Techera, Investigating the impact of physical fatigue on construction workers' situational awareness, *Saf Sci*, vol. 163, Jul. 2023.

[7] M. Wolf, J. Teizer, B. Wolf, S. Bükrü, and A. Solberg, Investigating hazard recognition in augmented virtuality for personalized feedback in construction safety education and training, *Advanced Engineering Informatics*, vol. 51. 2022.

[8] I. Jeelani, S. M. Asce, A. Albert, A. M. Asce; Roger Azevedo, and E. J. Jaselskis, Development and testing of a personalized hazard-recognition training intervention, *J Constr Eng Manag*, 2016.

[9] L. I. Domgue K *et al.*, Video see-through augmented reality fire safety training: a comparison with virtual reality and video training, *Saf Sci*, vol. 184, Apr. 2025.

[10] S. Hasanzadeh, B. Esmaeili, and M. D. Dodd, Examining the relationship between construction workers' visual attention and situation awareness under fall and tripping hazard conditions: using mobile eye tracking, *J Constr Eng Manag*, vol. 144, no. 7, Jul. 2018,

[11] A. Albert, M. R. Hallowell, B. Kleiner, A. Chen, and M. Golparvar-Fard, Enhancing construction hazard recognition with high-fidelity augmented virtuality, *J Constr Eng Manag*, vol. 140, no. 7, Jul. 2014.

[12] P. C. Liao, X. Sun, and D. Zhang, A multimodal study to measure the cognitive demands of hazard recognition in construction workplaces, *Saf Sci*, vol. 133, Jan. 2021.

[13] J. Hong, S. Song, H. Kang, J. Choi, T. Hong, and D. E. Lee, Influence of visual environments on struck-by hazards for construction equipment operators through virtual eye-tracking, *Autom Constr*, vol. 161, May 2024.

[14] M. Namian, A. Albert, and J. Feng, Effect of distraction on hazard recognition and safety risk perception, *J Constr Eng Manag*, vol. 144, no. 4, Apr. 2018.

[15] J. Li, H. Li, H. Wang, W. Umer, H. Fu, and X. Xing, Evaluating the impact of mental fatigue on construction equipment operators' ability to detect hazards using wearable eye-tracking technology, *Autom Constr*, vol. 105, Sep. 2019.

[16] J. Li, Y. Ouyang, and X. Luo, Impact of age on construction workers' preattentive and attentive visual processing for hazard detection, *Journal of Management in Engineering*, vol. 40, no. 3, 2024,

[17] S. Hasanzadeh, B. Dao, B. Esmaeili, and M. D. Dodd, Role of personality in construction safety: investigating the relationships between personality, attentional failure, and hazard identification under fall-hazard conditions, *J Constr Eng Manag*, vol. 145, no. 9, Sep. 2019,

[18] H. Fu, Y. Tan, Z. Xia, K. Feng, and X. Guo, Effects of construction workers' safety knowledge on hazard-identification performance via eye-movement modeling examples training, *Saf Sci*, vol. 180, Dec. 2024.

[19] H. Fu, Z. Xia, Y. Tan, and X. Guo, Influence of cues on the safety hazard recognition of construction workers during safety training: evidence from an eye-tracking experiment, *Journal of Civil Engineering Education*, vol. 150, no. 1, Jan. 2024.

[20] R. Eiris, M. Gheisari, and B. Esmaeili, Pars: using augmented 360-degree panoramas of reality for construction safety training, *Int J Environ Res Public Health*, vol. 15, no. 11. 2018.

[21] R. Eiris, M. Gheisari, and B. Esmaeili, Desktop-based safety training using 360-degree panorama and static virtual reality techniques: a comparative experimental study, *Autom Constr*, vol. 109, Jan. 2020.

[22] I. Jeelani, K. Han, and A. Albert, Development of virtual reality and stereo-panoramic environments for construction safety training, *Engineering, Construction and Architectural Management*, vol. 27, no. 8. 2020.

[23] R. Eiris, A. Jain, M. Gheisari, and A. Wehle, Safety immersive storytelling using narrated 360-degree panoramas: a fall hazard training within the electrical trade context, *Saf Sci*, vol. 127, Jul. 2020.

[24] A. F. E. Lucena and F. A. Saffaro, Guidelines for exploring construction sites in virtual reality environments for hazard identification, *International Journal of Occupational Safety and Ergonomics*, vol. 28, no. 1, pp. 86–95, 2022.

[25] M. Wolf, J. Teizer, B. Wolf, S. Bükrü, and A. Solberg, Investigating hazard recognition in augmented virtuality for personalized feedback in construction safety education and training, *Advanced Engineering Informatics*, vol. 51. 2022.

[26] D. Liu and C. Li, Construction hazard-recognition training for generation z students: a comparison study based on eye-movement data, *Engineering, Construction and Architectural Management*, 2024.

[27] Z. Fang, Q. Lu, X. Xie, A. K. Parlikad, J. Schooling, and M. Pitt, Definitions and principles of digital twins, in *Digital Twins in the Built Environment*, ch. Chapter 2, pp. 5–27.

[28] R. Sacks, I. Brilakis, E. Pikas, H. S. Xie, and M. Girolami, Construction with digital twin information systems, *Data-Centric Engineering*, vol. 1, no. 6, Nov. 2020.

[29] L. Messi, A. Corneli, M. Vaccarini, and A. Carbonari, Development of a twin model for real-time detection of fall hazards. In *Proceedings of the 37th International Symposium on Automation and Robotics in Construction*， 2020, 256–263.

[30] F. Vahdatikhaki, L. Olde Scholtenhuis, and A. Doree, Digital twin-based instructor support system for excavator training. In *Proceedings of the 41st International Symposium on Automation and Robotics in Construction*， 2024, 521–528.

[31] Z. Zhang *et al.*, Digital technology enhanced situation awareness for construction safety: systematic review and future research directions, *Saf Sci*, vol. 167, Nov. 2023.

[32] A. Harichandran, K. W. Johansen, E. L. Jacobsen, and J. Teizer, A conceptual framework for construction safety training using dynamic virtual reality games and digital twins, in *38th International Symposium on Automation and Robotics in Construction (ISARC 2021)*, 2021, pp. 621–628.

[33] G. P. Agnusdei, V. Elia, and M. G. Gnoni, Is digital twin technology supporting safety management? A bibliometric and systematic review, Mar. 02, 2021, *MDPI AG.*

[34] NASA, Systems engineering handbook, Rev 1. Washington, DC: National Aeronautics and Space Administration (NASA), 2007.

[35] B. Kerbl, G. Kopanas, T. Leimkuehler, and G. Drettakis, 3D gaussian splatting for real-time radiance field rendering, *ACM Trans Graph*, vol. 42, no. 4, Aug. 2023.

[36] S. Rüter, K. Waldow, N. Bartels, and A. Fuhrmann, 3D gaussian splatting for construction sites, in *24 th International Conference on Construction Applications of Virtual Reality (CONVR 2024)*, 2024.

[37] L. Barazzetti, M. Previtali, and F. Roncoroni, 3D modelling with the samsung gear 360, in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, International Society for Photogrammetry and Remote Sensing, Feb. 2017, pp. 85–90.

[38] Z. Feng *et al.*, Towards a customizable immersive virtual reality serious game for earthquake emergency training, *Advanced Engineering Informatics*, vol. 46, Oct. 2020.

DSCformer: A Dual-Branch Network Integrating Enhanced Dynamic Snake Convolution and SegFormer for Crack Segmentation

Kaiwei Yu[1], I-Ming Chen[2] and Jing Wu[1]

[1]Department of Mechanical and Energy Engineering, Southern University of Science and Technology, China
[2]School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore

12332394@mail.sustech.edu.cn, michen@ntu.edu.sg, wujing099@gmail.com

Abstract -

In construction quality monitoring, accurately detecting and segmenting cracks in concrete structures is paramount for safety and maintenance. Current convolutional neural networks (CNNs) have demonstrated strong performance in crack segmentation tasks, yet they often struggle with complex backgrounds and fail to fully capture fine-grained tubular structures. In contrast, Transformers excel at capturing global context, but lack precision in detailed feature extraction. We introduce DSCformer, a novel hybrid model that integrates an enhanced Dynamic Snake Convolution (DSConv) with a Transformer architecture for crack segmentation to address these challenges. Our key contributions include the enhanced DSConv through a pyramid kernel for adaptive offset computation and a simultaneous bidirectional learnable offset iteration, significantly improving the model's performance to capture intricate crack patterns. Additionally, we propose a Weighted Channel Attention Module (WCAM), which refines channel attention, allowing for more precise and adaptive feature attention. We evaluate DSCformer on the Crack3238 and FIND datasets, achieving IoUs of 59.22% and 87.24%, respectively. The experimental results suggest that our DSCformer outperforms state-of-the-art methods across different datasets.

Keywords -

Dynamic Snake Convolution; Crack segmentation;

1 Introduction

Monitoring the quality of concrete structures is crucial in industrial settings, as aging concrete buildings, such as pavements and bridges, often develop cracks that pose significant safety risks [1]. To ensure structural health, regular crack inspections are indispensable. While traditional inspection methods have been widely used, they often fall short in accuracy and efficiency. In contrast, deep learning models have emerged as a promising solution, demonstrating superior performance in crack segmentation tasks due to their accuracy and robustness.

While Convolutional Neural Networks (CNNs) excel at capturing local details due to their strong translational invariance and edge-detection capabilities, their limited receptive field can hinder performance in complex backgrounds [2, 3]. Conversely, Transformers leverage embeddings and self-attention mechanisms to model global dependencies effectively, which is advantageous for addressing complex backgrounds, but they often struggle to capture fine details [4]. Combining these two methods offers the potential to leverage both local detail extraction and global context modeling, addressing the respective limitations of each approach.

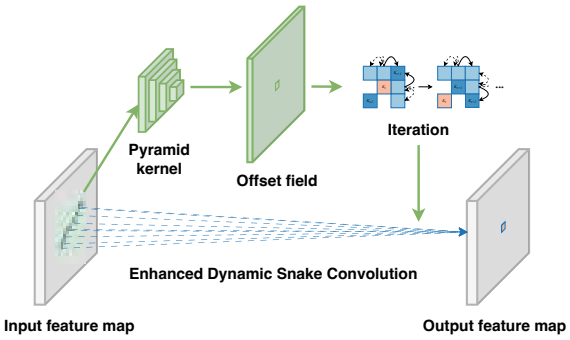


Figure 1. Overview of Enhanced Dynamic Snake Convolution.

The hybrid approach of combining CNNs and Transformers shows promise by leveraging both local detail extraction and global context modeling [5, 6, 7]. However, these methods often fail to account for the unique structure of crack segmentation, leading to suboptimal fusion. Additionally, combining CNNs and Transformers can introduce training fluctuations, making it challenging to fine-tune the model for both local and global features. In contrast, Dynamic Snake Convolution (DSConv) [8] is well-suited for crack segmentation due to its ability to focus on fine, curved local features, improving geometric perception. Despite its strengths, DSConv has limitations, such as fixed kernel sizes and limited adaptability in complex structures, suggesting room for improvement.

In this paper, we introduce DSCformer, a novel crack segmentation model that combines proposed enhanced DSConv with a Transformer-based architecture. Our approach introduces two key improvements to the original DSConv: a) a pyramid kernel for offset computation, consisting of convolutional kernels of varying sizes, and b) a method for learnable offset iteration that simultaneously operates in vertical and horizontal direction. Figure 1 shows the process for enhanced DSConv. These enhancements allow the model to capture fine crack structures more accurately. Additionally, we propose the Weighted Channel Attention Module (WCAM), which refines channel attention by utilizing separate Multi-Layer Perceptrons for average and max pooling. This leads to more adaptive and precise channel attention. The spatial attention module remains unaltered, contributing to spatial focus. By integrating these modifications into the DSConv encoder branch, which focuses on capturing fine local crack details, and combining it with a pre-trained SegFormer[9] encoder that specializes in capturing long-range dependencies, we fuse the extracted features and progressively upsample them to generate the final crack segmentation predictions. The principal contributions of this study are summarized as follows:

- We propose an enhanced DSConv with a pyramid kernel and bidirectional offset iteration, improving crack structure capture.
- We introduce WCAM, refining channel attention for greater adaptability and precision.
- We propose DSCformer, a hybrid model combining DSConv and SegFormer, outperforming state-of-the-art crack segmentation methods across two public datasets.

2 Related Works

2.1 Convolutional Networks for Crack Segmentation

Early works utilized CNN architectures like Fully Convolutional Networks (FCNs) [10] and U-Net [11] for segmentation. DeepCrack [2] follows an encoder-decoder structure. Among methods that alter convolution, Dilated Convolutions [12] expand the receptive field without sacrificing resolution, while Deformable Convolutions [13] improve feature adaptability by introducing learnable offsets to the kernels, enhancing object perception. Cracks, like blood vessels and roads, exhibit fine tubular structures, and enhancing their perception can significantly boost segmentation accuracy. Dynamic Snake Convolution (DSConv) [8] enhances geometric sensing for thin, curved structures by restricting deformable convolution sampling points, yielding improved performance. However, DSConv has limitations, such as a fixed offset kernel size and a one-directional learnable offset iteration, which may hinder adaptive topology perception.

2.2 Hybrid CNN-Transformer Models

Vision Transformers (ViT) [14] capture global image context through patch and position embeddings along with self-attention mechanisms. Swin Transformer [15] enhances computational efficiency with hierarchical shifted windows. Pyramid Vision Transformers (PVT) [16] combine multi-scale feature extraction with Transformer benefits, and SegFormer [9] introduces a hierarchical encoder without position embedding. However, pure Transformers struggle with fine detail capture.

Recent efforts combine CNNs and Transformers for crack segmentation. Crackformer [6] adds a convolution layer atop a Transformer encoder, while UTNet [5] and UCTransNet [17] integrate self-attention into U-Net structures. Crackmer [7] and DTrC-Net [18] employ parallel CNN-Transformer encoders. Despite these advances, more sophisticated hybrid models are needed for effective crack segmentation.

2.3 Attention Mechanisms for Feature Fusion

The Squeeze-and-Excitation (SE) network [19] recalibrates channel feature responses through average pooling, pioneering the concept of channel attention and enhancing the representational capacity of CNNs. Subsequently, the Convolutional Block Attention Module (CBAM) [20] extends the attention mechanism by sequentially applying a Channel Attention Module (CAM) and a Spatial Attention Module (SAM), utilizing both average and max pooling information to enable the network to more accurately focus on relevant features and their locations. In the context of crack segmentation, crackmer introduces a Complementary Fusion Module to more effectively aggregate multi-scale features, leveraging complementary information from different feature levels to improve segmentation accuracy. Nevertheless, the fusion of max and average poolings information is not sufficient.

3 DSCformer for Crack Segmentation

Figure 2 shows the DSCformer framework. Current hybrid crack segmentation methods combining CNN and Transformer mainly focus on feature fusion or sophisticated decoders. However, we believe the limitation lies in feature extraction, not fusion, as conventional convolutions struggle to capture crack features effectively. Dynamic Snake Convolution (DSConv) attempted to improve this but faced challenges with its offset calculation and iteration strategy. Inspired by this, we propose DSCformer, a dual-branch model combining enhanced DSConv and Transformer. The convolutional branch stacks DSC blocks with enhanced DSConv and Weighted Channel Attention

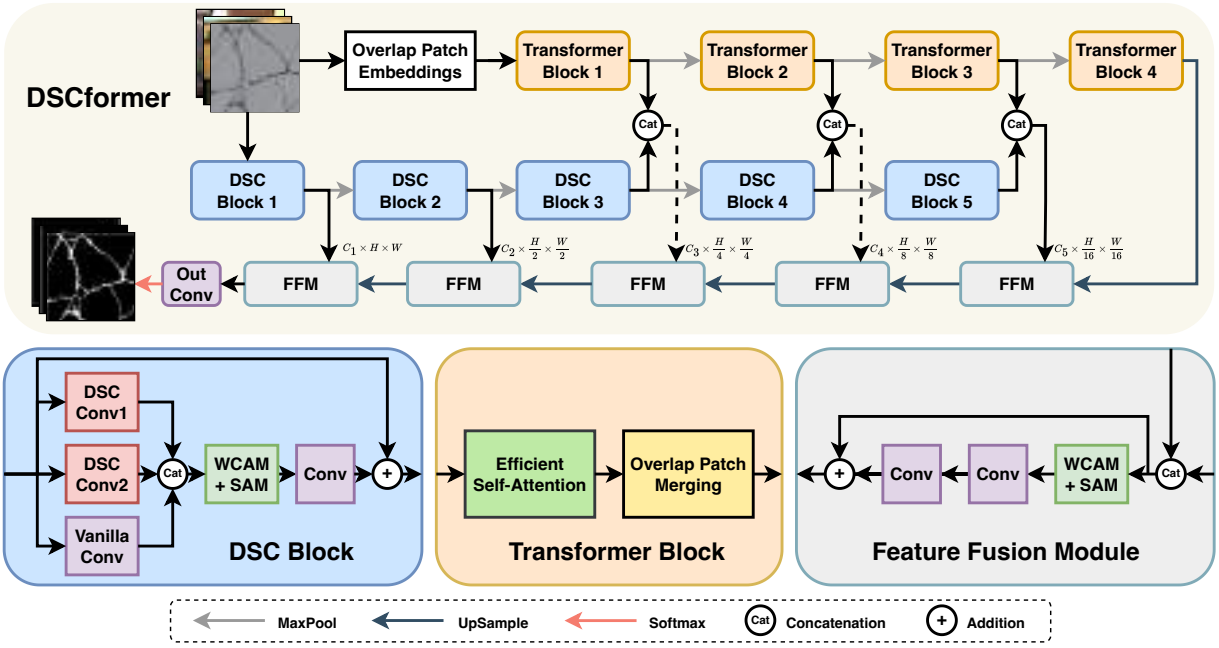


Figure 2. The overview architecture of the proposed DSCformer.

(WCAM), while the Transformer branch uses SegFormer. Crack segmentation masks are generated through progressive feature fusion and upsampling.

3.1 Enhanced DSConv encoder branch

The DSConv encoder branch is employed for local fine-grained structure feature extraction. We refine the computation strategy of DSConv, and propose a DSC block and WCAM. The DSC block incorporates WCAM to enhance feature fusion between vanilla convolution and enhanced DSConv. By stacking DSC blocks and adding max-pooling layers between DSC blocks, we perform progressive downsampling to obtain multi-scale feature maps at resolutions of 1/1, 1/2, 1/4, 1/8, and 1/16.

3.1.1 Enhanced Dynamic Snake Convolution

In this section, we introduce the computation process of Dynamic Snake Convolution (DSConv) and present our improvements. Given a standard $3 * 3$ convolution kernel with a dilation of 1 as K, the kernel samples information from nine points: the central point $K_i = (x_i, y_i)$ and its surrounding eight points, which is expressed as:

$$K = (x-1, y-1), (x-1, y), ..., (x+1, y+1) \quad (1)$$

To enable the convolution kernel to adapt to the tubular topological features around the current location, the goal is to allow the sampling points of the kernel to shift. Inspired by the concept of deformable convolutions, Qi et al. introduced deformation offsets Δ [8]. The offset convolution outputs offsets of 9 sampling points in the vertical and horizontal directions for a total of 18 channels. However, allowing the offsets to be completely learned freely can cause the receptive field to drift away from the target, which is unsuitable for capturing the characteristics of slender tubular structures. To address this, an iterative strategy is adopted to constrain the offsets, ensuring continuity in perception.

Unlike previous method that utilizes a $3 * 3$ convolution kernel to compute the sampling point offsets, we propose a pyramid-structured convolution kernel, which contains kernel sizes of $3 * 3$, $5 * 5$, $7 * 7$, $9 * 9$. Specifically, the $3 * 3$ kernel in our design outputs offsets for points near the center of the dynamic snake convolution, while the larger sizes of kernel outputs offsets for points further away, up to the $9 * 9$ kernel. This approach gradually perceives information from the outer layers, matching the receptive field for offset acquisition with the actual sampling points, thereby avoiding limited perception issues.

In the iterative offset strategy of DSConv, we consider a $9 * 9$ square, which is also its maximum receptive field. Each grid position in the kernel K is represented as $K_{t\pm c} = (x_{t\pm c}, y_{t\pm c})$, where $c = 0, 1, 2, 3, 4$. Unlike previous method, which fixes the offset as an integer in one direction while leaving the offset in the another direction unchanged, the enhanced DSConv do not fix the offset and iterate it in bi-direction simultaneously. So $t \pm c$ represents the iterative order of the convolution sampling points. Starting from the center point K_t,

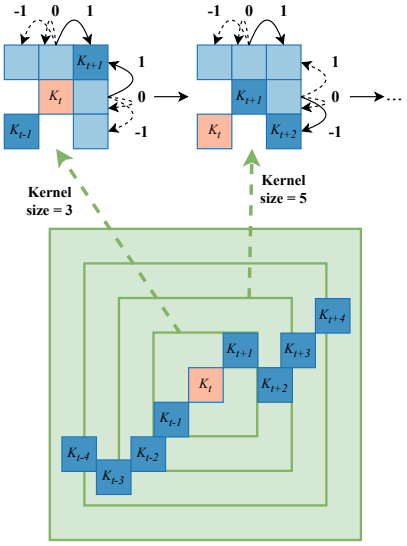


Figure 3. Offset iteration process. Iterating from the central point, each layer of the pyramid kernel represents its maximum receptive field.

$$K_{t+c} = (x_{t+c}, y_{t+c}) = (x_t + \sum_{t}^{t+c} \Delta x, y_t + \sum_{t}^{t+c} \Delta y) \quad (2)$$

$$K_{t-c} = (x_{t-c}, y_{t-c}) = (x_t - \sum_{t-c}^{t} \Delta x, y_t - \sum_{t-c}^{t} \Delta y) \quad (3)$$

After obtaining the coordinate of each sampling point after the offset iteration, the actual sampling point value is obtained by applying bilinear interpolation:

$$K = \sum_{K'} G(K', K) \cdot K' \quad (4)$$

Where K represents the position of the coordinates, K' represents the grid point near K, and G(·, ·) is the bilinear interpolation kernel. Note that G is two dimensional. It is separated into two one dimensional kernels as:

$$G(K, K') = g(K_x, K'_x) \cdot g(K_y, K'_y) \quad (5)$$

As shown in Figure 3, the enhanced DSConv improves the crack sensing ability of the snake chain away from the center point, and more flexibly captures the crack characteristics in the $9 * 9$ receiving field.

3.1.2 Weighted Channel Attention Module

In order to better focus on more relevant channel information, we propose WCAM, which is based on CAM, and makes better utilization of average pooling and maximum pooling information by introducing a few parameters. Reduces the deviation of attention due to accidental maximum values.

Mathematically, similarly, spatial information is first aggregated using max pooling and average pooling operations to obtain two distinct spatial identifiers: F^c_{avg}, F^c_{max}, each containing channel information. These two identifiers are then passed through independent MLPs, denoted as $M_{avg}(\cdot) = W^1_{avg}(W^0_{avg}(\cdot))$ and $M_{max}(\cdot) = W^1_{max}(W^0_{max}(\cdot))$ which, compared to a shared MLP, introduce only a small number of additional parameters with minimal impact on computational complexity. After applying the sigmoid activation function, the two output feature vectors are weighted, with the weights being learnable parameters. In summary, WCAM can be computed as follows:

$$M_c(F) = \sigma(w_{avg} M_{avg}(F^c_{avg}) + w_{max} M_{max}(F^c_{max})) \quad (6)$$

Here, σ denotes the sigmoid function, F contains F^c_{avg} and F^c_{max}, $W^0_{avg}, W^0_{max} \in \mathbb{R}^{C/r \times C}$, $W^1_{avg}, W^1_{max} \in \mathbb{R}^{C \times C/r}$, and $w_{avg}, w_{max} \in \mathbb{R}^C$. The ReLU activation function is followed by W^0_{avg}, W^0_{max}.

3.1.3 DSC block

In the DSConv branch, we propose the DSC block to extract features at different scales. Initially, the feature map is fed into three parallel convolutions: two enhanced DSConvs for extracting fine structural features especially and one standard convolution for capturing local features. The purpose of using two enhanced DSConvs is to better compare the original module (snake convolutions in both x- and y-directions), and the segmentation accuracy is virtually the same as when using only one enhanced DSConv. The outputs of the three convolutions are concatenated along the channel dimension and then processed through WCAM and SAM to adjust the feature map values, focusing on important channel information and spatial positions. This is followed by a standard convolution layer for feature fusion. Finally, a residual connection is introduced to mitigate the issue of gradient vanishing.

3.2 SegFormer encoder branch and decoder

To capture global information, often challenging for convolutional models, we integrate a Transformer branch using the SegFormer-b0 encoder, known for its efficient self-attention and Non-positional encodings, ideal for semantic segmentation. With only three million parameters, this encoder offers a lightweight yet robust solution. Fine-tuned on the ADE20k dataset [21], the pre-trained SegFormer-b0 enhances the DSConv branch by providing background context, improving crack segmentation. The encoder outputs multi-scale feature maps at 1/4, 1/8, 1/16, and 1/32 of the original resolution.

Table 1. Performance Comparison on Crack3238 and FIND Datasets with three test results.

Dataset	Methods	P. (M)	IoU (%) ↑	Pre. (%) ↑	Re. (%) ↑	F1 (%) ↑	HD (mm) ↓
Crack 3238	Unet	17.3	$53.43_{\pm 0.13}$	$65.74_{\pm 0.57}$	$67.67_{\pm 3.42}$	$66.30_{\pm 0.15}$	-
	DeepCrack	30.9	$54.22_{\pm 0.77}$	$65.86_{\pm 1.29}$	$72.06_{\pm 0.87}$	$67.25_{\pm 0.89}$	$56.28_{\pm 5.82}$
	DSCNet	2.1	$53.99_{\pm 0.16}$	$67.15_{\pm 1.69}$	$70.45_{\pm 1.98}$	$66.89_{\pm 0.19}$	$41.15_{\pm 0.90}$
	UCTransNet	66.5	$55.40_{\pm 0.31}$	$67.61_{\pm 0.83}$	$73.26_{\pm 0.38}$	$68.52_{\pm 0.43}$	$44.23_{\pm 1.82}$
	DcsNet	15.2	$56.85_{\pm 0.09}$	$69.83_{\pm 1.48}$	$73.73_{\pm 1.37}$	$70.05_{\pm 0.08}$	$37.17_{\pm 0.73}$
	DTrC-Net	41.8	$56.90_{\pm 0.33}$	$70.17_{\pm 1.42}$	$73.11_{\pm 1.51}$	$70.06_{\pm 0.39}$	-
	DSCformer(ours)	14.8	$\mathbf{58.74}_{\pm 0.47}$	$\mathbf{72.38}_{\pm 2.22}$	$\mathbf{73.99}_{\pm 1.33}$	$\mathbf{71.74}_{\pm 0.48}$	$\mathbf{33.35}_{\pm 0.97}$
FIND	Unet	17.3	$85.57_{\pm 0.18}$	$\mathbf{92.97}_{\pm 0.61}$	$91.65_{\pm 0.82}$	$92.00_{\pm 0.11}$	$13.21_{\pm 0.73}$
	DeepCrack	30.9	$84.25_{\pm 0.16}$	$91.40_{\pm 0.84}$	$91.65_{\pm 0.88}$	$91.21_{\pm 0.09}$	$14.12_{\pm 0.96}$
	DSCNet	2.1	$85.75_{\pm 0.07}$	$92.35_{\pm 0.70}$	$92.40_{\pm 0.67}$	$92.12_{\pm 0.04}$	$12.23_{\pm 0.20}$
	UCTransNet	66.5	$86.05_{\pm 0.26}$	$91.15_{\pm 0.29}$	$93.98_{\pm 0.48}$	$92.30_{\pm 0.16}$	$11.90_{\pm 0.61}$
	DcsNet	15.2	$85.02_{\pm 0.09}$	$91.97_{\pm 1.00}$	$91.98_{\pm 1.09}$	$91.64_{\pm 0.06}$	$14.67_{\pm 0.86}$
	DTrC-Net	41.8	$74.28_{\pm 0.07}$	$83.50_{\pm 0.54}$	$84.94_{\pm 1.59}$	$84.21_{\pm 0.02}$	$14.76_{\pm 0.09}$
	DSCformer(ours)	14.8	$\mathbf{87.31}_{\pm 0.12}$	$92.52_{\pm 0.35}$	$\mathbf{94.01}_{\pm 0.45}$	$\mathbf{93.04}_{\pm 0.09}$	$\mathbf{11.14}_{\pm 0.23}$

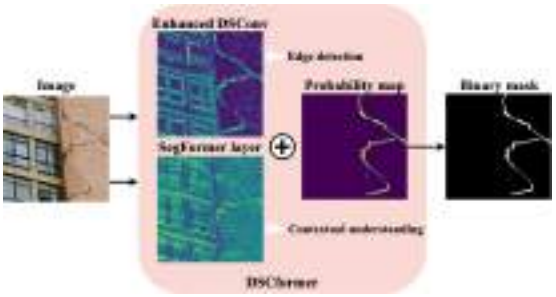


Figure 4. The difference between dual-branch encoder.

For the decoder, we fuse feature maps from the DSConv and SegFormer branches with upsampled lower-resolution features. These fused maps undergo WCAM and SAM processing to refine channel and spatial information. The refined features pass through two convolutional layers with residual connections to smooth the output and address vanishing gradients. As shown in Figure 4, enhanced DSConv excels at extracting edges and fine structures, while SegFormer is effective at global modeling. The final crack mask is obtained by fusing these features.

4 Experiments

4.1 Datasets

We use Crack3238 dataset [18], Fused Image dataset for convolutional neural Network-based crack Detection(FIND) dataset [22] to evaluate our method. Crack3238 dataset contains a total of 3,238 images, which consist of RGB images of cracks from various sources, featuring complex backgrounds, diverse crack patterns, and defective surfaces, significantly increasing the complexity of crack detection scenarios. The FIND dataset comprises 2,500 images, where fused image data is obtained by combining depth information with grayscale images. Both datasets encompass data from different sources and highly complex scenes, making them extremely challenging for crack segmentation tasks. We split the datasets, using 80% for training and 20% for testing.

4.2 Implementation Details

4.2.1 Training details.

Our model, implemented in PyTorch, was trained on a RTX 3090Ti GPU. To reduce overfitting, we applied data augmentation, including flips, rotations, and affine transformations. The SegFormer encoder uses pre-trained weights, while the DSConv encoder and decoder are randomly initialized. We used Adam (lr=0.0001, weight decay=10^{-4}) with cross-entropy and dice loss. Training ran for 100 epochs with a batch size of 8.

4.2.2 Performance metrics.

We evaluated segmentation performance using IoU, Recall, Precision, F1-score, and Hausdorff Distance. The formulations are as follows: IoU is given by $IoU = \frac{TP}{TP+FP+FN}$, Recall is given by $Recall = \frac{TP}{TP+FN}$, Precision is given by $Precision = \frac{TP}{TP+FP}$, and F1-score is given by $F1 = \frac{2\times Precision\times Recall}{Precision+Recall}$. Additionally, we employed the Hausdorff Distance, calculated as $d_H(A,B) = \max\left(\max_{a\in A}\min_{b\in B}|a-b|, \max_{b\in B}\min_{a\in A}|a-b|\right)$, to measure the maximum distance between the segmentation mask and the ground truth. All metrics for each image are calculated and averaged.

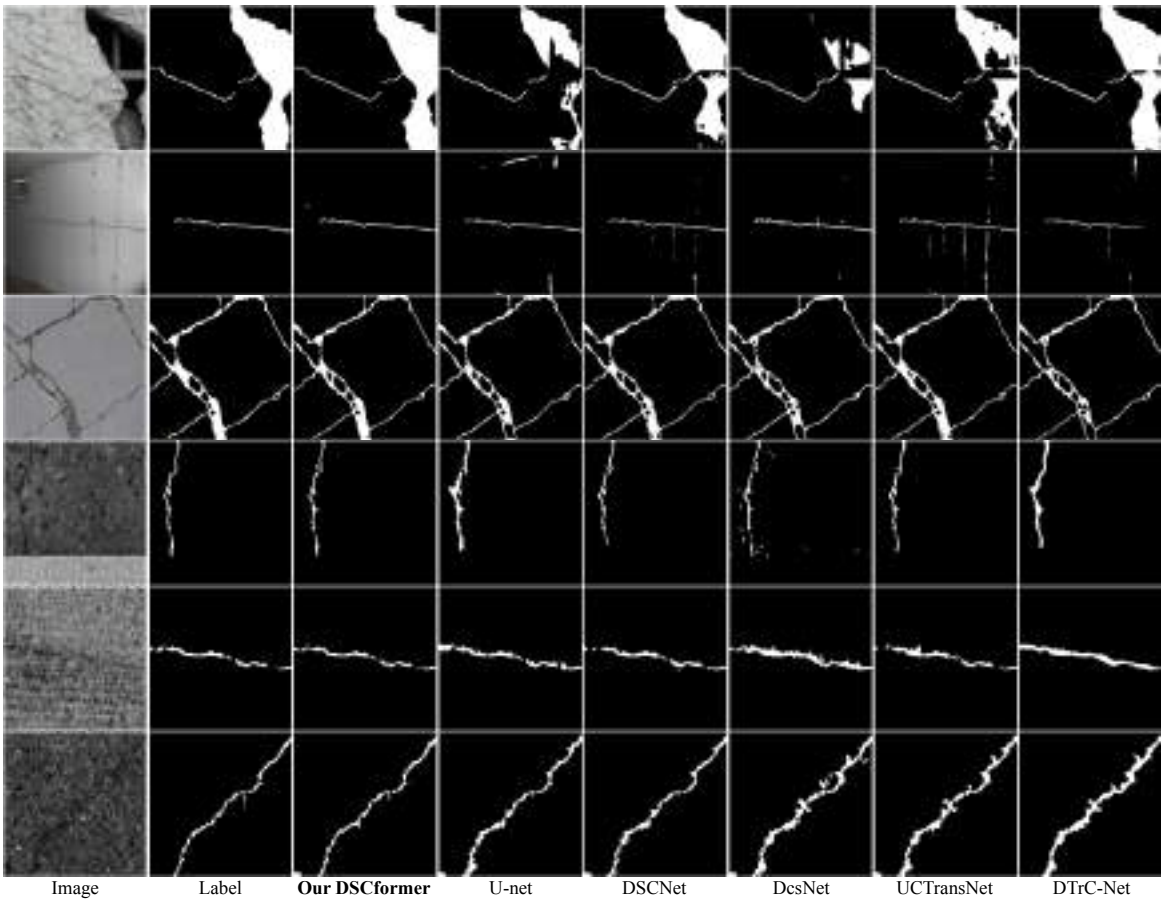


Figure 5. The qualitative comparison on the Crack3238 and FIND datasets.

4.3 Comparison with State-of-the-art Methods

We compare the proposed DSCformer with several state-of-the-art models, including U-Net, DeepCrack, DcsNet, UCTransNet, DTrCNet, and DSCNet. For fairness, we use their original code and settings, testing each model three times. Results are shown in Table 1, with metrics presented as mean ± standard deviation, highlighting the best results in bold. DSCformer outperforms all models, achieving an IoU of 58.74%, Precision of 72.38%, Recall of 73.99%, and F1-score of 71.74% on the Crack3238, and an IoU of 87.31% and F1-score of 93.04% on the FIND. It also balances Precision and Recall effectively, with Precision achieving second-best performance.

Additionally, a visual comparison in Figure 5 illustrates DSCformer's superior performance, particularly in handling noise and accurately capturing large defect areas. DSCformer also excels at delineating crack edges, avoiding the overestimation of crack width seen in other models, and ensuring more precise segmentation, especially at boundaries. This level of detail highlights DSCformer's strong generalization across varying defect types and sizes.

Table 2. Generalization performance on Crack3238 datasets.

Augmentation	DSCformer	DTrC-Net	DSCNet
None	**58.74%**	56.90%	53.99%
Geometric	**53.52%**	50.50%	49.94%
Appearance	**49.36%**	46.38%	48.04%
Combined	**42.40%**	37.50%	39.69%

We also conducted a generalization experiment to evaluate the performance of three models—DSCformer, DTrC-Net, and DSCNet—under various data augmentation conditions, including no augmentation, geometric augmentation, appearance augmentation, and combined augmentation. The results in Table 2 show that DSCformer consistently outperforms the other models across all augmentation types, though its performance declines with increased augmentation complexity, especially under combined augmentation (42.40%). In comparison, DTrC-Net and DSCNet exhibit more significant performance drops, particularly under appearance and combined augmentations. Al-

though affected by the complexity of data augmentation, DSCformer still shows strong generalization ability.

Additionally, in terms of computational efficiency, DSCformer achieves an inference speed of 30 FPS, which is comparable to the hybrid model UCTransNet, while other models range between 60-100 FPS. This shows that DSCformer balances accuracy and computational efficiency well without excessive resource consumption.

Overall, DSCformer's ability to manage noise, precisely segment large defects, and accurately delineate crack edges makes it a highly effective method for crack segmentation tasks, outperforming existing models both quantitatively and visually. Additionally, it demonstrates strong generalization ability, maintaining superior performance across different data augmentation conditions.

4.4 Ablation Studies

We conducted ablation experiments on the Crack3238 dataset to assess the impact of the proposed improvements on the performance of our model. As summarized in Table 3, the dual-branch architecture of DSCformer exhibits a significant advantage over the single-branch models, DSCNet and SegFormer. Specifically, DSCformer achieved an IoU improvement of 4.75% over DSCNet and 4.07% over SegFormer, highlighting the performance of integrating both branches for crack segmentation.

Table 3. Ablation experiments on Crack3238 datasets. 'CA' denotes the Channel Attention.

Methods	Conv	CA	IoU
DSCNet	DSConv	w/o	53.99%
SegFormer	w/o	w/o	54.67%
DSCformer	Vanilla Conv	WCAM	57.46%
	DSConv	WCAM	57.74%
	Enhanced DSConv	w/o	58.16%
	Enhanced DSConv	CAM	58.28%
	Enhanced DSConv	WCAM	58.74%

The enhanced DSConv also shows substantial benefits. By incorporating the enhanced DSConv into the model, we observed a notable 1.00% increase in IoU compared to the original DSConv, indicating that the improvements to the DSConv effectively contribute to more accurate crack boundary detection and segmentation. Additionally, the introduction of WCAM provided further gains in segmentation performance. When WCAM was applied within DSCformer, it outperformed the standard CAM, resulting in a 0.46% increase in IoU. This improvement reflects the value of our modifications to the attention mechanism. Overall, these results emphasize the significance of the hybrid model architecture and the specific enhancements to both the dynamic snake convolution and channel attention mechanisms. Each modification contributes to the overall effectiveness of DSCformer, leading to superior crack segmentation performance.

5 Conclusion

In this work, we introduced DSCformer, a hybrid model combining enhanced Dynamic Snake Convolution (DSConv) and Transformer architectures for crack segmentation. DSCformer addresses the limitations of CNNs and Transformers by capturing both fine local details and global context. The enhanced DSConv improves crack feature detection with a pyramid kernel and bidirectional offset iteration, while the Weighted Channel Attention Module (WCAM) refines channel attention for better feature fusion. Experiments show that DSCformer outperforms state-of-the-art methods on benchmark datasets, demonstrating its robustness and potential for broader applications in analyzing fine topological structures.

6 Acknowledgment

This material is based upon work supported by the National Science Foundation of China(NSFC)62203205 and U1913603. The views, opinions, findings and conclusions reflected in this publication are solely those of the authors and do not represent the official policy or position of the NFSC.

References

[1] Jay Lee and Hanqi Su. A unified industrial large knowledge model framework in industry 4.0 and smart manufacturing. *International Journal of AI for Materials and Design*, page 3681, 2024.

[2] Yahui Liu, Jian Yao, Xiaohu Lu, Renping Xie, and Li Li. Deepcrack: A deep hierarchical feature learning architecture for crack segmentation. *Neurocomputing*, 338:139–153, 2019.

[3] Wentao Zhang, Haoran Kang, Wenhui Wang, Yangtao Ge, Haiou Liao, Rui-Jun Yan, I-Ming Chen, Zhenzhong Jia, and Jing Wu. A high-accuracy crack defect detection based on fully convolutional network applied to building quality inspection robot. In *2022 International Conference on Advanced Robotics and Mechatronics (ICARM)*, pages 63–68. IEEE, 2022.

[4] Elyas Asadi Shamsabadi, Chang Xu, Aravinda S Rao, Tuan Nguyen, Tuan Ngo, and Daniel Dias-da Costa. Vision transformer-based autonomous crack detection on asphalt and concrete surfaces. *Automation in Construction*, 140:104316, 2022.

[5] Yunhe Gao, Mu Zhou, and Dimitris N Metaxas. Utnet: a hybrid transformer architecture for medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III 24*, pages 61–71. Springer, 2021.

[6] Huajun Liu, Xiangyu Miao, Christoph Mertz, Chengzhong Xu, and Hui Kong. Crackformer: Transformer network for fine-grained crack detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3783–3792, 2021.

[7] Jin Wang, Zhigao Zeng, Pradip Kumar Sharma, Osama Alfarraj, Amr Tolba, Jianming Zhang, and Lei Wang. Dual-path network combining cnn and transformer for pavement crack segmentation. *Automation in Construction*, 158:105217, 2024.

[8] Yaolei Qi, Yuting He, Xiaoming Qi, Yuan Zhang, and Guanyu Yang. Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6070–6079, 2023.

[9] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in neural information processing systems*, 34:12077–12090, 2021.

[10] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[11] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015. URL `https://arxiv.org/abs/1505.04597`.

[12] Fisher Yu, Vladlen Koltun, and Thomas Funkhouser. Dilated residual networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 472–480, 2017.

[13] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, pages 764–773, 2017.

[14] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[15] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021.

[16] Wenhai Wang, Enze Xie, Xiang Li, Deng-Ping Fan, Kaitao Song, Ding Liang, Tong Lu, Ping Luo, and Ling Shao. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 568–578, 2021.

[17] Haonan Wang, Peng Cao, Jiaqi Wang, and Osmar R Zaiane. Uctransnet: rethinking the skip connections in u-net from a channel-wise perspective with transformer. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pages 2441–2449, 2022.

[18] Chao Xiang, Jingjing Guo, Ran Cao, and Lu Deng. A crack-segmentation algorithm fusing transformers and convolutional neural networks for complex detection scenarios. *Automation in Construction*, 152: 104894, 2023.

[19] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.

[20] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19, 2018.

[21] Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Scene parsing through ade20k dataset. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 633–641, 2017.

[22] Shanglian Zhou, Carlos Canchila, and Wei Song. Fused image dataset for convolutional neural network-based crack detection (find), 2022. URL `https://doi.org/10.5281/zenodo.6383044`.

A framework for computer vision-based real-time asphalt pavement property monitoring

Dongying Wang[1], Qinshuo Shen[1], Faridaddin Vahdatikhaki[1], Seirgei Rosario Miller[1], and André Dorée[1]

[1]Department of Civil Engineering, University of Twente, Netherlands
d.wang-1@utwente.nl, q.shen@utwente.nl, f.vahdatikhaki@utwente.nl, s.r.miller@utwente.nl, and a.g.doree@utwente.nl

Abstract –

The monitoring of the road construction operation is critical to ensure high-quality roads. This is especially the case for the compaction of the asphalt surface layer since the operation is very time-sensitive and mostly irreversible. Historically, monitoring of the asphalt compaction is limited to temperature and compaction homogeneity. While useful, these pieces of information fail to generate direct insight into the resultant physical properties of the asphalt layer (e.g., density and roughness), which are the ultimate key performance indicators. Several sensor-based methods and approaches have been proposed but these methods are particularly suited for post-construction monitoring rather than real-time application. To address this gap, this research aims to conduct a preliminary feasibility study on the use of a computer vision-based method for the real-time monitoring of the asphalt properties during the compaction. Therefore, a framework is proposed to install cameras on the asphalt roller along with a Global Positioning System (GPS) kit and track the changes of the asphalt texture at the micro level. These changes are then correlated with the progression of the compaction effort. The results indicate that the Computer Vision (CV) method can very successfully identify micro variations in the asphalt texture. This is very promising because these variations can be potentially correlated with some physical properties of the asphalt layer. While this correlation study is the future work of this research, the preliminary results already contribute to highlighting the potential of CV methods to be incorporated into asphalt compaction monitoring system.

Keywords –

Asphalt construction; Computer Vision; Texture analysis

1 Introduction

The asphalt road is the backbone of the modern transportation system. According to the Ministry of Infrastructure and Water Management of the Netherlands, the total road traffic in 2017 was 140.8 billion vehicle Kilometers [1]. The quality of asphalt roads is of vital importance to ensure the safety of passengers and also the efficiency of transportation. However, the quality of asphalt roads is a complex parameter to control as it depends on multiple factors, including but not limited to asphalt material, production procedure, base layer condition, and the intensity of use. One, often less acknowledged, factor that significantly impacts the long-term performance of asphalt roads is the quality of the construction process. The consistent and homogenous compaction of the asphalt layer at the right temperature significantly influences the longevity and quality of roads. Considering the irreversible nature of asphalt pavement construction, it is critical to accurately inspect the asphalt pavement during the construction process to spot and fix the failures in the early stages.

In recent years, realizing that early detection of failures during the construction phase could contribute to the increased service life of road pavements while also reducing its life cycle cost by up to 37% [2], on-site monitoring during the compaction process has been taken more seriously. Makarov et al. [3] has proposed a framework for compaction guidance systems that can translate the temperature and compaction count data into clear suggestions for compaction strategies. Han et al. [4] developed a construction quality evaluation framework combining building information model (BIM) and geographic information systems (GIS), focusing on the temperature, speed, and rolling frequency during the compaction phase. Zeng et al. [5] presented a real-time asphalt pavement segregation identification algorithm based on digital imaging. However, the above-mentioned research primarily focuses on the monitoring of temperature and roller behaviours instead of the real-time monitoring of the resultant physical properties of the

asphalt layer (e.g., roughness and density). Or, in other words, are just indirect assessments of asphalt quality.

The real-time asphalt density and roughness monitoring during compaction has remained a bottleneck. Currently, the Ground-penetrating Radar (GPR) is considered to be an effective, feasible, and non-destructive tool for the real-time estimation of the density of asphalt concrete [6]. Wang et al. [7] proposed a signal-processing method for asphalt pavement layer thickness determination using GPR. However, the GPR-based method is strongly dependent on the asphalt material since the dielectric constant is essential for the AC density prediction [8]. For the real-time asphalt smoothness measurement, Chang et al. [9] introduced a laser-based system for the asphalt roughness identified from paver operation and roller compaction. Compared with existing as-build asphalt roughness measurements [10][11], this approach offers a solution for the real-time roughness measurement. However, due to the selection of the sensors, this approach is sensitive to typical events such as varying paving speeds(introducing vibration) and the environmental factors such as temperature, steam, and dust [9].

One possible method to address these limitations is to use computer vision techniques to directly assess the quality of the asphalt layer during construction. However, to the best of the authors' knowledge, the topic is not yet fully addressed. Therefore, this research aims to conduct a preliminary feasibility study of the vision-based real-time monitoring of asphalt properties. In this method, a geo-referenced image data collection system is designed. With the database built, image processing and data analysis algorithms will be introduced to correlate the change of asphalt surface texture during compaction with its corresponding physical properties. The remainder of the paper is structured as follows. In Section 2, the proposed computer vision-based asphalt monitoring framework is presented. Section 3 makes a brief explanation of a feasibility study taken in the real construction site will be introduced. Subsequently, the analysis of the results and discussions regarding the case study will be carried out. Section 4 encloses with a conclusion.

2 Methodology

A comprehensive geo-referenced data collection, image processing, and data analyzing framework is proposed in this section. The framework consists of 3 main stages, as illustrated in Figure 1. Schematic representation of the proposed framework.

In the initial stage, a data collection pipeline will be leveraged to collect geo-referenced asphalt pavement surface images. With the built dataset, an image processing algorithm will be developed, consisting of image pre-processing, segmentation, and result correction. In the last stage, the asphalt surface texture information generated from the image processing algorithm will be used as input for data analysis, where a machine learning method will be applied to make predictions on the corresponding physical properties of the asphalt pavement surface texture.

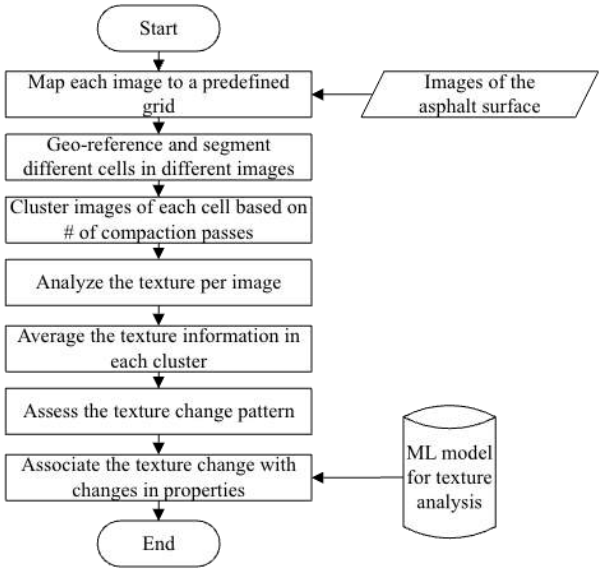


Figure 1. Schematic representation of the proposed framework

2.1 Data collection and pre-processing

The dynamic geo-referenced asphalt pavement surface texture data collection system is composed of a GPS signal receiver mounted on the top of the roller, and two cameras placed in the front of the roller and the back of the roller respectively, as illustrated in Figure 2. Schematic representation of the .

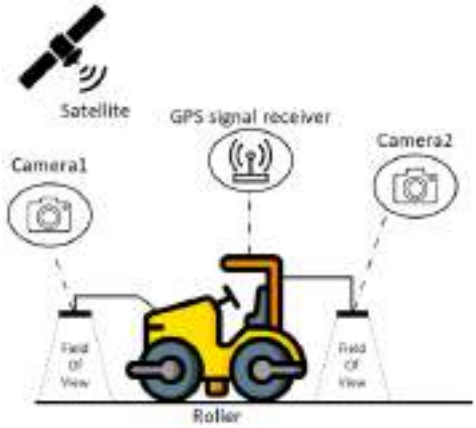


Figure 2. Schematic representation of the sensor setup

This system will take images of the asphalt surface as shown in Figure 3. The void area proportion generated

from different parameter sets . This dual-camera setup allows us to track asphalt pavement surface texture changes before and after compaction.

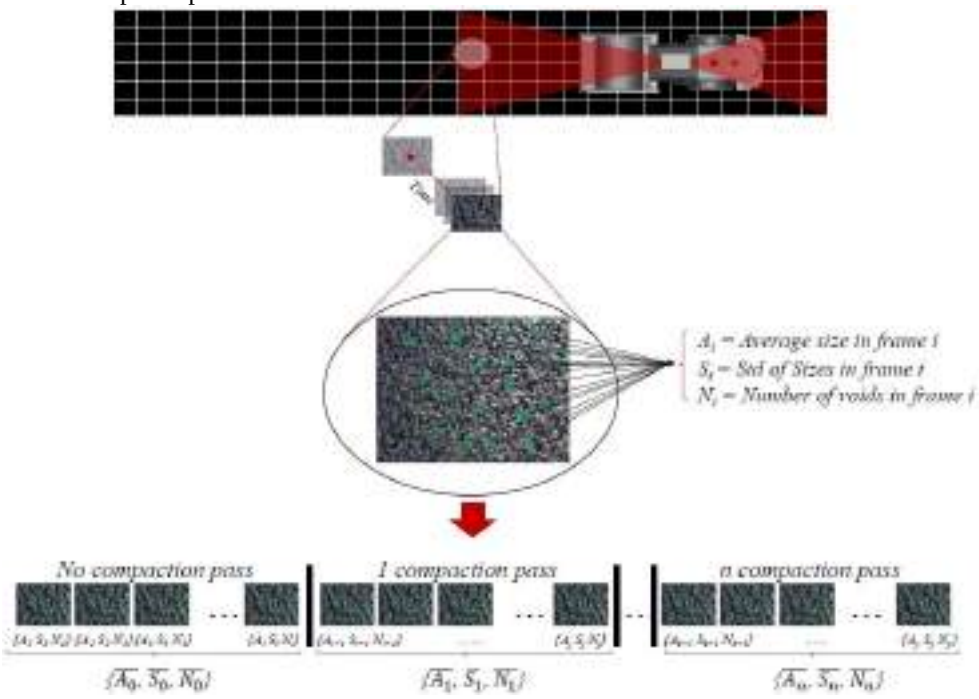


Figure 3. The void area proportion generated from different parameter sets

The raw images first need to be transformed to perspective images directly facing the surface and a compensation on the translation between the GPS receiver and the camera also need to be made. This is because the texture analysis, which is based on the shape and sizes of the voids between the aggregates, requires a uniform scale and perspective in all the images. By calibrating the camera [12], extrinsic and intrinsic parameters could be obtained. The image transformation is based on the relative orientation of the camera to the pavement surface and the translation between the camera and the GPS receiver. From the extrinsic parameters, namely the rotation matrix and translation vector, the camera's position and orientation relative to the world coordinate system could be obtained. In this phase, since the target is to convert the perspective of the camera and compensate the translation between camera and GPS signal, only the rotation matrix will be utilized and the translation vector will be replaced with a fixed translation vector between the GPS receiver and the camera. With the above-mentioned information, a homograph matrix can be computed, and with this matrix, the images captured could be projected into a perspective directly facing the surface and each image will be tagged with its rough geo-information.

Next, each raw image will be cropped into sub-images and each sub-image will be registered to a pre-defined physical grid, as shown in Figure 3. The void area proportion generated from different parameter sets. Each sub-image will be tagged with their fine geo-information. Notice that the pixel equivalent of each image is determined by (1) the installation height of the camera, (2) the orientation of the camera to the surface, and (3) the resolution of the camera[13], thus the cropped images may vary in pixel-wise sizes for different cameras.

Since the field of view (FOV) of each camera covers more than one grid, the same grid may be captured in several sequential images, and they will be clustered for result averaging. The determination of roller pass counts incorporates both image timestamps and the spatial coordinates of the roller. It is notable that sub-images that belong to the same grid from different frames will be clustered and the final output of this grid will be an average output value of all images in this cluster, as shown in Figure 3. The void area proportion generated from different parameter sets.

Sequentially, the segmentation algorithm will be conducted on the cropped images to segment the voids between the aggregates. The segmentation process is operated on each sub-image and it includes 3 steps.

(a) Brightness correction

As shown in Figure 4. Images in different lighting condition and the result of brightness correction, since

the images are captured in the natural environment, the lighting conditions could vary, and the algorithm is supposed to be sufficient to adapt to different scene illuminations.

With dynamic range compression, the images taken from different illuminated scenes are expected to produce good spectral tonal rendition and be tuned into similar brightness conditions. The Multiscale Retinex (MSR) [14] algorithm, one of the most popular and proven to be significant brightness correctness algorithms, is applied in this study. In this stage, the redundant information of the raw images such as dirt on the surface will be removed and the essential texture information (aggregate edges) will be preserved and enhanced. Figure 4 shows the result of brightness correction.

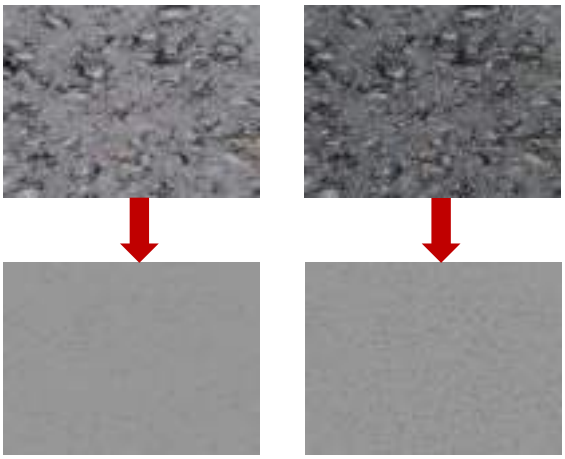


Figure 4. Images in different lighting condition and the result of brightness correction

(b) Segmentation

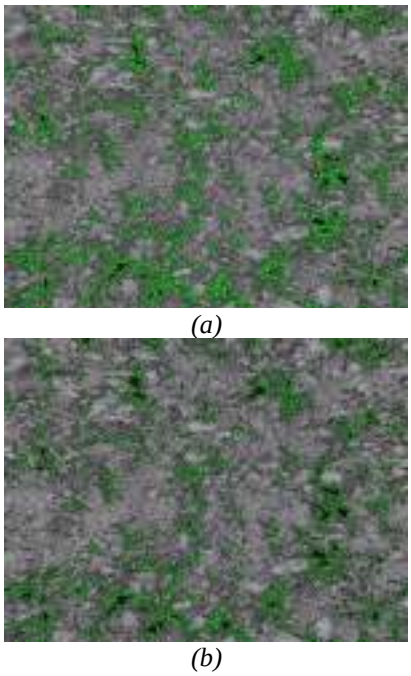
The aggregates and the voids between them exhibit different grayscale values in the images. The voids are dark areas in the images while the aggregate areas display larger grayscale values. Considering this grayscale distribution pattern, thresholding can be applied to segment the voids. Since the raw images are corrected to a similar brightness condition, a global threshold value could be set to traverse each pixel in the images and binarize the image. If the grayscale value of the pixel is greater than the threshold value, it is considered to be the aggregate and the value of this pixel will be set to 0, otherwise, it is considered as the target voids and will be set to 255. The threshold is calculated with Equation 1.

$$threshold = mean - a * deviation \quad (1)$$

Where $mean$ refers to the mean gray value of the image, $deviation$ refers to the deviation of the gray value in the image and a refers to an experience-based parameter, which will be explained in Section 3.

(c) Segmentation result correction

The segmentation result is based on the grayscale values and small connected components can be found in the segmented results. Ideally, the fine textures in the small connected components are expected to offer detailed information on the asphalt pavement surface. However, they negatively impact the performance of texture analysis since the fine texture structure of pavement image has a marginal effect on the physical properties of the asphalt pavement (e.g., roughness) [15] but increases the computation cost of analysis. Also, some of the small connected components represent noises that affect the accuracy of subsequent processing steps. To refine the analysis and emphasize representative asphalt surface texture patterns, small and isolated regions will be removed according to their size and morphological [16] connectivity. The comparison between the initial segmentation results and the results after correction is shown in Figure 5. The contours of the segmented voids are labeled with green polygons. It is worth mentioning that the "voids" in this work refer to the void area between the aggregates on the surface, which display dark regions on the images, and they could represent the surface texture of the asphalt layer.



(a)

(b)

Figure 5. An example of the result of segmentation (a) before and (b) after correction

2.2 Data analysis

The segregated void areas will be mapped into quantified physical properties, which will be considered as a regression problem. The input of the regression

algorithm includes handcrafted features such as the quantity of void areas and the total area size of the voids, and also the high-dimensioned features generated from a pre-trained convolutional neural network (CNN). With the input, the machine learning algorithm is expected to deliver the quantified physical properties.

2.2.1 Texture analysis

The output of the image processing step is a binary image, with the void areas set to 255 and the background set to 0. The mask image, representing the features of the original image, will be serialized into a vector as the input of the machine learning algorithm.

Note that the images from each grid are clustered and their feature vectors will be averaged and then forwarded to the next step as the input feature of the machine learning algorithm.

2.2.2 Correlating texture properties with physical properties

After getting a vector as the input of the machine learning algorithm, the output of the machine learning algorithm is determined to be the corresponding physical properties of the asphalt surface area. During the training and validation stage for the machine learning algorithm, this data will be collected on-site or in a lab test. The training and validation process will follow a regular machine learning algorithm development process.

3 Experimental results and discussion

Since the core assumption of this research is that there is a trackable change in the asphalt pavement texture when the compaction happens, the preliminary feasibility validation for this research focused on the texture change detection. Two case studies were conducted in two different road construction projects in cities of Almelo and Winterswijk, the Netherlands. On each project sites, we collected data on 4 different sampling stations. The data collection station for this preliminary verification consists of a stationary camera placed on the side of the road and capturing 1 image every 10 seconds, as shown in **Error! Reference source not found.**.

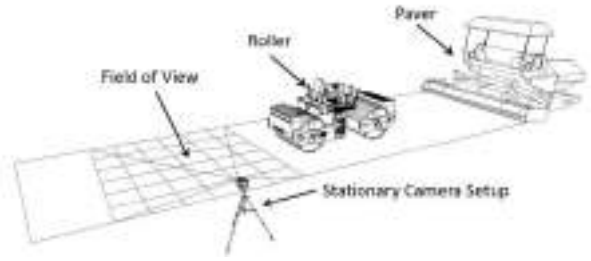


Figure 6. The experimental setup

The images captured within the compaction intervals are expected to display the same or similar pattern in texture, and images captured right before and after compaction are expected to display different texture patterns. Different models of cameras were used for each project, to assess the sensitivity to the camera model. The case studies aimed at (1) capturing the asphalt pavement surface images for the same area experiencing different rounds of roller pass, (2) extracting the surface texture, and (3) providing a quantified analysis of the texture indicating its change.

Following the steps introduced in Chapter 2.1, the total void area after each compaction was detected. As mentioned above, a few key parameters are crucial for the detection results. The sensitivity of the model to these parameters is shown in Figure 7.

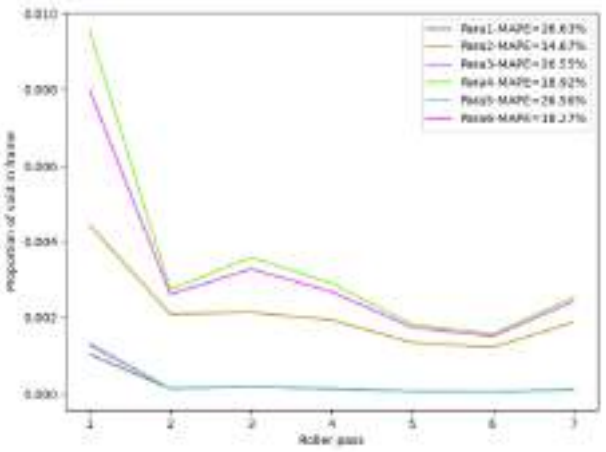


Figure 7. The void area proportion generated from different parameter sets

As shown in this figure, while all graphs indicate an overall negative exponential pattern, the magnitude and smoothness can vary. Given that the physical behaviour of asphalt conforms to a negative exponential trend, a negative exponential curve fitting was employed as the objective function to optimize the segmentation parameters. The mean absolute percentage error (MAPE), as indicated in Equation 2, was used for the evaluation of the fitting result.

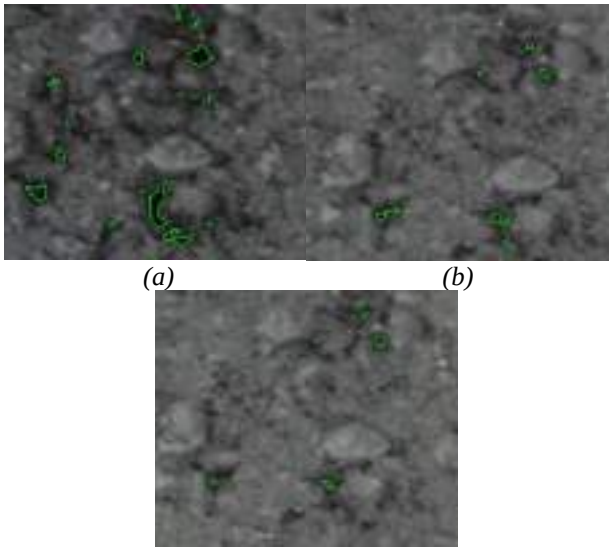
$$MAPE = \frac{\sum \frac{|A - P|}{A} * 100}{N} \quad (2)$$

Where N refers to the number of sampling points, A refers to the actual value and P refers to the predicted value.

The result shows that the MAPE of the designed algorithm could reach 14.67% with an optimal set of parameters, which indicates that the variation of the segmented void area to the roller pass could be well-fitted

to a negative exponential function. As the assumption is that the texture of the asphalt layer surface exhibits a similar trend as some of the mechanical or physical properties of the asphalt to the roller passes, which is known to be a negative exponential pattern, this result proved that the segmentation algorithm is sufficient.

An example of the detection results of a sampled area after different roller passes is shown in **Error! Reference source not found.**. The results below is generated from the optimal parameters.



(a) *(b)*

(c)

Figure 8. An example of the segmentation results (a) right after paving, (b) after 5 rounds of compaction, and (c) after 10 rounds of compaction

Figure 9 presents the variation of the total void area in relation to the compaction passes for different test locations. As shown in this figure, there is a clear degrading pattern of the total void areas with respect to compaction passes. This can be construed as a strong indication that changes in asphalt textures are tracible and can be captured by computer vision. This clearly points to the potential that the texture analysis could be correlated with the development of different physical properties.

It is worth noting that the ranges of the total void area percentage in each frame are expected to be more or less the same, while it varies in our test. One potential reason for this variance is that the tests are done on different construction sites where different materials are used and each material has different properties. Another possible reason for this variance is that the distance between the camera and the surface is not precisely adjusted and therefore different from one station to another. At a larger distance, due to the lower resolution and the filtering strategies, some of the small voids could be undetected. Since the feasibility analysis is based on each camera in a time series, this factor does not affect the result of this experiment. However, it needs to be considered in future research when images from mobile camera (i.e., mounted on equipment) are used for analysis.

Also, during the validation, we found some factors may interfere with the image quality and lead to an unsatisfying result if the cameras are installed on the roller. The factors are (1) the water on the rollers will result in the steam and it will render the images blurry, and (2) the vibration of the roller may result in blurry images. Thus, the above-mentioned factors need to be considered when installing the camera on the rollers and extra mechanical designs such as stabilizer need to be added to the system.

4 Conclusion

This research proposed a framework for the vision-based monitoring of the asphalt properties during construction. As a preliminary validation of the idea, the asphalt pavement texture analysis has been done with a stationary camera set. The preliminary results exhibit a negative exponential pattern on the total void area of the asphalt surface as the roller passes increase. The negative exponential trend follows the theoretical trend of changes in many of physical and mechanical properties of the asphalt layer. This indicates that this method can be leveraged to correlate these properties with the changes in the texture of the asphalt during the construction. The application could be a real-time system where operators can be informed about the condition of the asphalt layer during the operation to adjust their operational strategies.

It should be highlighted that while vision-based monitoring of the asphalt condition in the operation phase (e.g., for detecting cracks and potholes) has been extensively studied, to the best of the authors' knowledge, the use of texture analysis during the construction process and the correlation of the texture change with the progression of the physical properties of the asphalt layer has not been studied.

This research will be further pursued by (1) installing the cameras on the rollers and considering the environmental interferes this may have on the images, and (2) correlating different physical properties of the asphalt layer with variation of the asphalt texture during the construction.

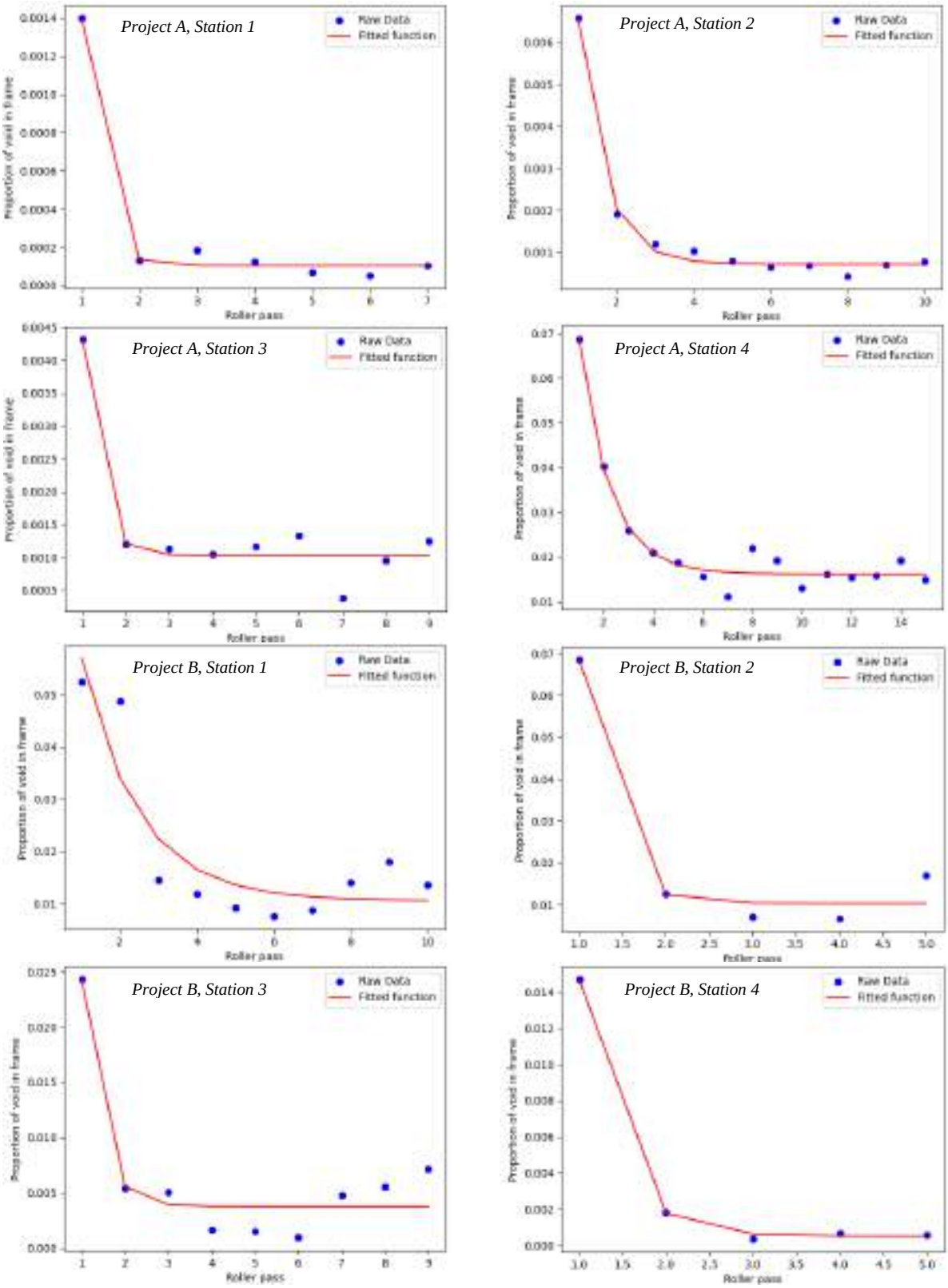


Figure 9. Fitting result on the void area after compaction

References

[1] Key Transport figures 2018. On-line: https://english.kimnet.nl/mobility-report/key-transport-figures-2018, Accessed: 09/12/2024.

[2] S. Park, S. Bang, H. Kim and H. Kim. Patch-Based Crack Detection in Black Box Images Using Convolutional Neural Networks. Journal of Computing in Civil Engineering, 33(9): 04019017, 2019.

[3] D. Makarov, F. Vahdatikhaki, S. Miller, A. Jamshidi, and A. Dorée. A framework for real-time compaction guidance system based on compaction priority mapping. *Automation in Construction*, 129: 103818, 2021.

[4] C. Han, F. Tang, T. Ma, L. Gu, and Z. Tong. Construction quality evaluation of asphalt pavement based on BIM and GIS. *Automation in Construction*, 141: 104398, 2022.

[5] S. Zeng, C. Zhang, K. Zeng, N. Liang, and S. Song. Real-time identification of asphalt mixture segregation during paving process using digital imaging technique and Four-side static moment. *Construction and Building Materials*, 397: 132436, 2023.

[6] S. Zhao, and I. L. Al-Qadi. Algorithm development for real-time thin asphalt concrete overlay compaction monitoring using ground-penetrating radar. *NDT & E International*, 104: 114-123, 2019.

[7] S. Wang, Z. Leng, Z. Zhang, and X. Sui. Automatic asphalt layer interface detection and thickness determination from ground-penetrating radar data. Construction and Building Materials, 357: 129434, 2022.

[8] S. Wang, I. L. Al-Qadi, and Q. Cao. Factors Impacting Monitoring Asphalt Pavement Density by Ground Penetrating Radar. *NDT & E International*, 115: 102296, 2020.

[9] G. K. Chang, D. Merritt, A. Gilliland, T. Mansell, E. Dukatz, and M. S. Oman. *Asphalt Real-Time Smoothness (ARTS) for Asphalt Paving*. No. NRRA202302. Minnesota. Department of Transportation. Office of Research & Innovation, 2023.

[10] S. Xu, Q. Liu, Y. Bo, Z. Chen, and C. Wang. Estimating the International Roughness Index of asphalt concrete pavement by response-based testing equipment and intelligent algorithms. 433: 136659, 2024.

[11] Y. Sang, Q. Yu, Y. Fang, V. Vo and R. Wix, Smartphone-Based IRI Estimation for Pavement Roughness Monitoring: A Data-Driven Approach. IEEE Internet of Things Journal, 11(11): 19708-19720, 2024.

[12] Z. Zhang. A Flexible New Technique for Camera Calibration. Pattern Analysis and Machine Intelligence, 22: 1330-1334, 2000.

[13] D. McCready. On size, distance, and visual angle perception. *Perception & Psychophysics*, 37: 323-334, 1985.

[14] D. J. Jobson, Z. Rahman and G. A. Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image Processing*, 6(7): 965-976, 1997.

[15] G. Qian, Z. Wang, H. Yu, C. Shi, C. Zhang, J. Ge, and W. Dai. Research on surface texture and skid resistance of asphalt pavement considering abrasion effect, *Case Studies in Construction Materials*, 20: e02949, 2024.

[16] R.C. Gonzales, and P.Wintz. Digital image processing, edition 2. Addison-Wesley Longman Publishing Co., Inc., USA, 1987.

Vision-based Segmentation, Measurement, and Pose Estimation in Modular Integrated Construction

Yuanyang Qi[1], Xiao Li[2], and Ruiqi Jiang[3]

[1]Department of Civil Engineering, The University of Hong Kong, China
[2]Department of Civil Engineering, The University of Hong Kong, China
[3]Department of Civil Engineering, The University of Hong Kong, China

yuanyang.qi@connect.hku.hk, shell.x.li@hku.hk, richj233@connect.hku.hk

Abstract -

Modular Integrated Construction (MIC) currently depends on manual assembly processes, which are not only inefficient but also have safety risks. To address this issue, we propose vision-based technologies for automatic segmentation, measurement, and position estimation to aid in the assembly process in construction. Specifically, we utilize the Yolov8-seg model, which is an advanced instance segmentation tool, to segment the surfaces of the corners of the modules. These segmented surfaces are then fitted with quadrilaterals to accurately localize the four corner points. The 3D coordinates of these corners are determined by using the corresponding depth map, thus enabling precise measurements. Additionally, using the 2D coordinates of the colour map of the four corners and the actual 3D coordinates created with the center of the quadrilateral as the origin, we employ the Perspective-n-Point (PnP) algorithm for precise position estimation. The experiments show the effectiveness of the proposed methods. By integrating these vision-based techniques with construction robotics, we can significantly enhance the efficiency and safety of MiC assembly, paving the way towards full construction automation.

Keywords -

Computer Vision; Segmentation; Measurement; Pose Estimation; Modular Integrated Construction

1 Introduction

Modular Integrated Construction (MIC) involves the design of standardized modules, which are pre-fabricated in off-site manufacturing facilities and subsequently assembled on-site [1]. This method simplified the building process, saving project time and costs. However, the module installation process remains arduous and time-consuming. Lifting large modules requires cranes and hoists, and relies heavily on worker experience to achieve precise lifting; if workers are inexperienced, installation errors, delays, and safety accidents may occur, affecting project time and costs [2]. Due to the increasing demand for faster, safer and more efficient installations [3], the integration of new technologies such as computer vision is becoming important to improve accuracy, speed, and safety in module assembly.

Vision-based techniques have great potential to solve problems in the field of construction, offering a range of advantages that improve the overall construction process, including quality management, progress and schedule monitoring, safety enhancement, and cost reduction [4]. These vision-based methods involve advanced image processing, deep learning, 3D reconstruction, and other techniques to extract valuable visual perceptual information. Furthermore, when integrated with construction robots, these techniques—such as segmentation, measurement, and pose estimation—can significantly improve the robots' perception, enabling more efficient and automated construction.

Segmentation is extracting useful regions from an image or video frame for analysis [5]. In recent years, deep learning applications such as convolutional neural networks (CNN) have greatly improved the accuracy and efficiency of image segmentation [6]. Generally, segmentation can be divided into semantic segmentation and instance segmentation, each provides solutions for different kinds of tasks. Semantic segmentation is the categorisation of each pixel into a specific class, without distinguishing between specific objects in each class. One of the pioneering works of semantic segmentation is Fully Convolutional Networks (FCN) [7], which achieves end-to-end image segmentation by grouping each pixel into specific classes using convolutional layers rather than fully connected layers. U-Net [8], on the other hand, uses an encoder-decoder structure to be able to capture both local and global features and skip connections to fuse deep and shallow information. DeepLab [9] uses dilated convolution to increase the sensory field, which facilitates the segmentation of objects at different scales. Instance segmentation not only focuses on which class each pixel belongs to but also focuses on distinguishing different objects in the same class, which is more challenging than semantic segmentation. Mask R-CNN [10] is one of the most widely used instance segmentation models, which improves on Faster R-CNN [11]

by adding pixel-level segmentation to achieve detection and segmentation at the same time. Leanne Attard et al. [12] employed Mask R-CNN to detect and localize cracks on concrete surfaces, thereby overcoming the limitations of manual detection and significantly reducing both time and cost.

Measurement technology plays a crucial role in Modular Integrated Construction (MIC), particularly for the precise measurement of prefabricated modules. Traditional measurement methods, which rely on manual techniques, are time-consuming, labour-intensive, and subjective [13]. In contrast, vision-based methods such as binocular vision, Time of Flight (ToF), and structured light offer more accurate, objective, and reliable measurement solutions. Binocular vision simulates the human eye by using two cameras to take pictures of the same object at different angles, and by calculating the disparity of the same object in the pictures of the two cameras, thus achieving the depth measurement of the object [14]. ToF obtains depth information by measuring the time it takes for the light signal to travel from the emitting source to the target object and back to the sensor [15]. The structured light system achieves 3d reconstruction of the object by projecting a known pattern onto the target object, and the camera achieves this by capturing changes in the pattern [16]. Liu [17] uses binocular vision to measure the size of cracks in walls, preventing and treating wall defects in advance, thus enabling structural and environmental monitoring of buildings. Yu et al. [18] combined structured light with an industrial robot to stitch together 3d data, simplifying the calibration process and improving the measurement accuracy, thus enabling accurate measurement of the 3D shape of large objects.

Pose estimation is widely used in augmented reality, construction robotics, and industrial automation to enable three-dimensional reconstruction, robotic assembly, and automated construction. Pose estimation aims at obtaining the position and attitude of an object in 3D space, which is crucial for the accurate placement and installation of MIC modules. Attitude estimation methods are classified into feature-based methods and deep learning-based methods etc. Feature-based methods extract feature points such as corner points, edges, key points, etc. from images or video frames and use feature matching to estimate position and orientation. For example, the Perspective-n-Point (PnP) algorithm uses a projection model to estimate pose by extracting feature points from 2d images and matching them with actual 3D coordinates [19]. Mi et al. [20] used the PnP algorithm for 3D position measurement of the pit to achieve automated monitoring of pit displacement. Deep learning-based algorithms use CNN to predict the pose directly end to end. PoseNet [21] based on CNN directly estimates the position and orientation of the camera in 3D space from a single image without feature matching.OpenPose can extract the keypoints from a single image and supports multi-person detection. Prabesh Paudel et al. [22] used PoseNet to detect the worker's position and assess the risk based on the worker's posture to ensure safety. Deep learning-based methods require a large amount of training data to learn effective features with high computational complexity, whereas feature-based pose estimation algorithms require less data, have low computational complexity and good real-time performance.

There are also several studies that apply computer vision techniques to MIC to address challenges in the manual assembly process. For instance, Roshan Panahi et al. [23] proposed an automated assembly progress monitoring system for modular construction factories using computer vision-based instance segmentation. Zhenjie Zheng et al. [24] used Mask R-CNN in the construction process, thus enabling accurate localisation and segmentation of MIC modules, which is conducive to automated process monitoring. However, while these studies focus on segmentation and monitoring, there remains a gap in fully automating not just the detection but also the precise measurement and pose estimation of construction modules.

In this study, we aim to develop a computer vision-based framework for automated segmentation, precise measurement, and position estimation of construction components, with a focus on MIC corner surfaces. The specific objectives of this study are: (1) to achieve accurate instance segmentation of MIC module corner surfaces using the Yolov8-seg model, (2) to extract and fit segmented surfaces into quadrilaterals for identifying precise corner points and measuring the MIC corner, and (3) to implement the PnP algorithm for pose estimation by using the relationship between 2D and 3D corner coordinates.

The contributions of this study include: (1) the integration of state-of-the-art computer vision techniques to automate traditionally manual tasks in the construction industry, (2) a novel application of instance segmentation and 3D depth-based analysis to achieve precise segmentation and measurement MIC components, and (3) the pose estimation of construction components through the combined use of depth data and 2D-3D point correspondence.

The following sections of this paper is structured as follows: Section 2 details the proposed methodology, including segmentation, measurement, and the PnP-based pose estimation process. Section 3 presents experimental results and evaluates the performance of the proposed framework. Finally, Section 4 provides a conclusion, summarizing the key findings.

2 Methodology

Figure 1 shows a typical MIC module. It is notable that there is a distinctive rectangular fitting at each of the four

corners of the module. This fitting has unique geometric features and functions that not only provide additional support for the module but also facilitate quick and precise assembly. Therefore this fitting is an important feature of the MIC module. By analyzing the corner features, we can achieve the detection and positioning of the corners of the MIC module, which enables the robot to automatically identify and manipulate the MIC module and improves the automation level of the construction.



Figure 1. Typical MIC module.

To further investigate the corner feature of the MIC module, a 1:1 model of the corner shown in Figure 2 was 3D printed to investigate automated vision-based processing methods, including segmentation, measurement, and pose estimation. The model allows us to simulate and evaluate the performance of the vision system accurately under laboratory conditions.



Figure 2. 3D printed 1:1 model of the MIC module corner.

2.1 Segmentation

Considering the trade-off between accuracy and processing speed, we have selected the Yolov8-seg model for image segmentation. Yolov8 introduces several enhancements to optimize performance, including an anchor-free detection method, the C2f module, a decoupled header, and a modified loss function [25]. These improvements collectively enhance the model's speed, accuracy, and generalization capabilities. Yolov8-seg, a variant derived from Yolov8, is specifically engineered for segmentation tasks. It inherits the high accuracy and speed of the YOLO series and extends these capabilities to achieve pixel-level segmentation. The architecture of Yolov8-seg is illustrated in Figure 3.

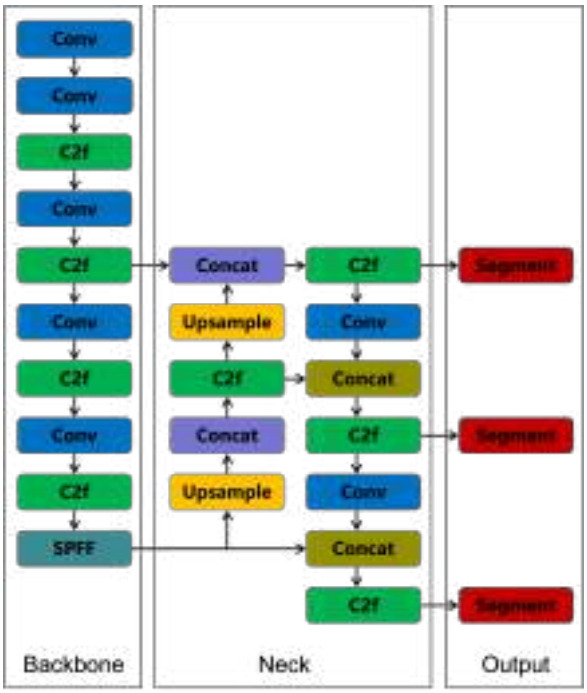


Figure 3. The architecture of Yolov8-seg.

Accurate image segmentation is essential for precise subsequent analysis. In our study, we employed the Yolov8-seg model to segment the surface of the corner model, which allows us to detect and localize the specific positions of the corner model accurately. After segmenting the image, we fit it into a quadrilateral to extract the four corner points. In this way, we can know the exact position of the four corner points in the 2D image and get the 2D coordinates of the four corner points. The segmented image and the coordinates of the four corner points provide a robust foundation for further geometric and dimensional analyses. These will subsequently be utilized for in-depth measurements and accurate pose estimation.

2.2 Measurement

The process of mapping points from a three-dimensional coordinate system to a two-dimensional image plane can be described using the pinhole camera model, which is shown in Figure 4.

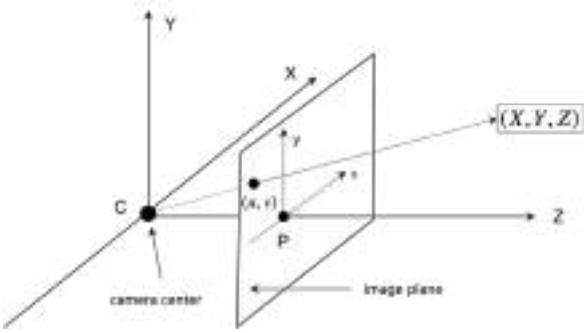


Figure 4. Pinhole camera model.

According to the pinhole camera model, the projection of a point (X, Y, Z) in three-dimensional space onto a point (u, v) on the two-dimensional image plane is described by the following equations:

$$\begin{aligned} u &= f_x \frac{X}{Z} + c_x, \\ v &= f_y \frac{Y}{Z} + c_y, \end{aligned} \tag{1}$$

where f_x and f_y are the focal lengths along the x and y axes, respectively, and c_x and c_y are the coordinates of the principal point on the image plane.

This can be further expressed in matrix form:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \begin{bmatrix} \frac{X}{Z} \\ \frac{Y}{Z} \\ 1 \end{bmatrix} \tag{2}$$

where K is the camera's intrinsic matrix and defined as:

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

To accurately measure the size of the corner model, we use a depth camera. The depth camera captures both a colour map and a corresponding depth map, with the depth map providing the depth value (i.e., the z-coordinate) for the 3D points. Using the formula:

$$\begin{aligned} X &= (u - C_x) \times \frac{Z}{f_x} \\ Y &= \left(v - C_y\right) \times \frac{Z}{f_y} \\ Z &= \text{ depth value} \end{aligned} \tag{4}$$

we can accurately derive the 3D coordinates of any point from its depth map. Using the 2d image coordinates of the four corner points obtained from the segmentation result in the previous step, combined with the depth map, the 3d coordinates of the four corner points can be obtained. The actual length and width of the corner model can be obtained by calculating the distance between two neighbouring points from the 3d coordinates of the four corner points.

2.3 Pose Estimation

Since we have previously determined the actual length and width of the corner model, we can construct a 3D coordinate system for the corner model based on the actual length and width. By setting the center point of the model as the coordinate origin, we can locate the four corners relative to this origin to obtain their actual 3D coordinates. Given that we have both the actual 3D coordinates of the four corner points and their corresponding 2D image coordinates, we can employ the PnP algorithm to accurately estimate the pose of the corner model.

The PnP algorithm is widely used in computer vision to estimate the position and orientation of the camera relative to the object based on a set of 2D-3D points. PnP is an optimization problem that aims to determine the optimal rotation and translation matrices that minimize the reprojection error. This error is calculated by measuring the difference between the predicted positions of the 3D points when re-projected onto the image plane and their actual positions. The reprojection error is mathematically expressed as:

$$e_i = x_i - K[R \mid t]X_i \tag{5}$$

where x_i represents a 2D point on the image plane, X_i is the corresponding 3D point, R and t denote the camera's rotation matrix and translation vector, respectively, and K is the camera's intrinsic parameter matrix.

The goal of PnP algorithm is to minimize the reprojection error, that is:

$$\min_{R,t} \sum_i \|e_i\|^2 \tag{6}$$

This optimization problem can be solved by iterative methods, where R and t are adjusted step by step through several iterations, so as to find the minimum value of the reprojection error and the optimal R and t.

3 Experiments and Results

3.1 Dataset

We used the Intel RealSense D455 depth camera, commonly used in construction, to collect data. The D455 depth camera has two high-resolution colour and depth sensors that capture depth information through the principle of stereo vision. The intrinsic parameters of the camera are shown in Table 1 and Tabel 2 including the focal lengths (f_x, f_y), the principal point coordinates (c_x, c_y), and the distortion coefficients.

Table 1. Parameters of the colour camera

Color camera	Intrinsic parameters
f_x, f_y	391.942, 391.942
c_x, c_y	324.305, 239.027
Distortion	[0, 0, 0, 0, 0]

Table 2. Parameters of the depth camera

Depth camera	Intrinsic parameters
f_x, f_y	387.781, 387.351
c_x, c_y	325.001, 247.284
Distortion	[-0.055, 0.066, -0.800, 8.003, -0.022]

The coordinate systems of the depth camera and the colour camera are aligned through a transformation that includes rotation and translation. This alignment ensures that the data from both cameras are accurately matched. This transformation is defined by the extrinsic parameters of the cameras, represented by the rotation matrix R and the translation vector t, as follows:

$$R = \begin{bmatrix} 0.99998343 & -0.00549165 & -0.00175197 \\ 0.00548666 & 0.99998093 & -0.00283836 \\ 0.00176752 & 0.00282870 & 0.99999446 \end{bmatrix}$$

$$t = \begin{bmatrix} -0.05912773311138153 \\ 0.0003899137955158949 \\ 0.00034507890813983977 \end{bmatrix}$$

We captured a total of 1018 images under different illumination conditions and camera heights, as shown in Figure 5. These 1018 images were randomly shuffled and split into two parts: 80% as the training set to train the model and 20% as the test set to evaluate the model.

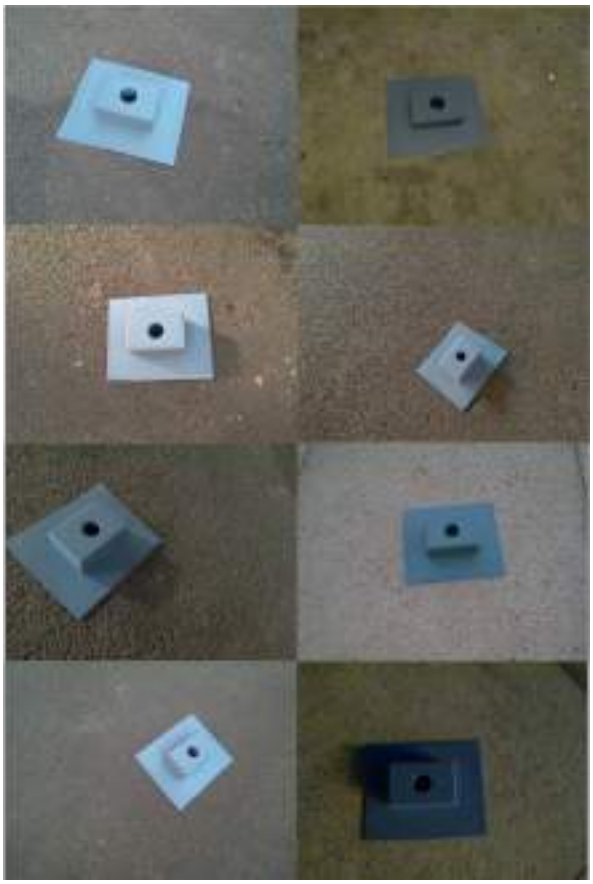


Figure 5. Image examples under different illumination conditions and camera heights.

3.2 Segmentation

We used the Yolov8-seg model for image segmentation, which enables instance segmentation by combining precise object detection and localization capabilities. The experiments were carried out using the NVIDIA GeForce RTX 4090 GPU. The input images were resized to a resolution of 640x640 pixels to standardize the input data format. We set the training with a batch size of 6 and the model was trained for 350 epochs.

To assess the effectiveness and accuracy of the model, we use a comprehensive set of evaluation metrics on the test set, including precision, recall, AP50 and AP50-95. These metrics help evaluate the model's ability to accurately detect and segment objects, providing a detailed analysis of its performance for segmentation.

After segmentation with Yolov8-seg, for each object, we get two key outputs: a bounding box and a segmentation mask. The bounding box marks the position of the object in the image with a rectangular box, which allows us to quickly detect the presence and approximate position of the object in the image. The segmentation mask provides more fine-grained output, including pixel-level categorization, indicating which pixels in the bounding box belong to the object and delimiting the object's boundaries.

Table 3 presents the results of the Yolov8-seg on the test set, indicating robust performance in the corner segmentation task. Furthermore, Figure 6 displays the segmentation result for the corner model using Yolov8-seg, clearly showing that precise and accurate segmentation was achieved.

Table 3. Experimental results of Yolov8-seg.

Evaluation metrics	Box	Mask
Precision	1	1
Recall	1	1
AP50	0.995	0.995
AP50-95	0.993	0.97

3.3 Measurement

After segmenting the surface of the corner model using the Yolov8-seg, we fit a quadrilateral to the segmented surface. The four vertices of the quadrilateral represent the

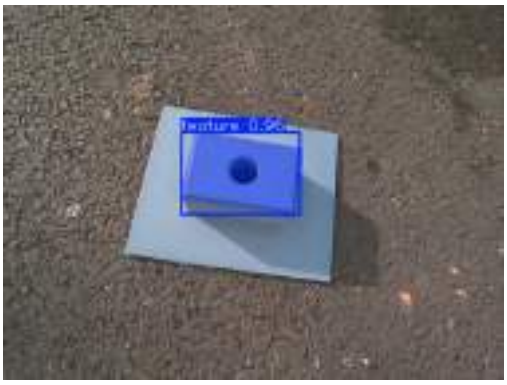


Figure 6. The segmentation result for the corner model using Yolov8-seg.

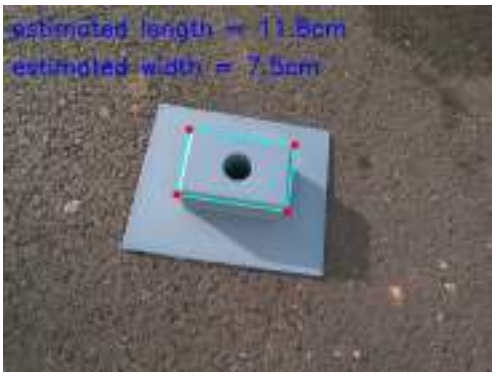


Figure 7. The measurement result for the corner model.

corners. By mapping these vertices to the corresponding depth map, we can obtain their actual three-dimensional coordinates. This enables us to measure the actual distances between the corners, thus determining the length and width of the corner model. Figure 7 shows the estimated dimensions of the corner model, with the length measured at 11.8 cm and the width at 7.5 cm. These estimations are compared to the actual dimensions, which are 12 cm in length and 8 cm in width. The test set, consisting of 204 images, was used to further validate the measurement accuracy, with the results showing that the measurement error is within 9 mm.

The measurement error can be attributed to several factors. One factor is the low resolution of the colour images, which limits the precision of segmentation and corner detection. Additionally, inaccuracies in the depth map, caused by sensor noise or resolution limitations, contribute significantly to errors in calculating the 3D coordinates of the corners. Segmentation errors from the Yolov8-seg model also play a role, as minor deviations in boundary detection can lead to slight inaccuracies in identifying the exact corner points.

3.4 Pose estimation

Once we have determined the exact coordinates and relative distances of the four corner points, we establish a coordinate system with the origin situated at the center of the quadrilateral formed by these points, which are at the upper left, upper right, lower left, and lower right, respectively. Following the segmentation process, we also acquire the 2D coordinates of these corner points within the colour map. Consequently, we can utilize the PnP algorithm to estimate the camera's position relative to the corner model surface. This is achieved by aligning the coordinate system of the corner model with the camera's coordinate system, allowing for precise calculation of the camera's orientation and position in relation to the corner model surface.

Figure 8 illustrates the results of estimating the pose using the PnP algorithm. The position of the camera relative to the surface of the corner model is described by a translation vector that quantifies the displacement of the camera to the model. The orientation of the camera relative to the surface of the angular model is represented by a quaternion. Quaternions provide a compact representation of orientations and do not suffer from the gimbal-lock problem that may result from orientation representations such as Euler angles.

To address occlusion, we consider introducing additional positioning tags, such as the Apriltag marker. When the corner features are not occluded, we simultaneously estimate the poses of both the Apriltag marker and the corner model, and calculate the relative pose deviation between them. When the corner model is occluded, we detect the pose of the Apriltag marker and use the previously obtained relative pose deviation between the Apriltag marker and the corner model to estimate the pose of the occluded corner model. Figure 9 shows pose estimation results in two scenarios. The upper image shows pose estimation when the corner features are not occluded, while the lower image shows pose estimation of occluded corner features using the Apriltag marker.

4 Conclusion

In conclusion, this study applied computer vision technology to MIC to enhance the efficiency, accuracy, and safety of the construction process. Unlike traditional construction methods, which rely heavily on manual labour and are susceptible to human error, our vision-based approach facilitates automated segmentation, measurement, and pose estimation. Specifically, we first use the Yolov8-seg model to segment the corner surfaces of modules. The

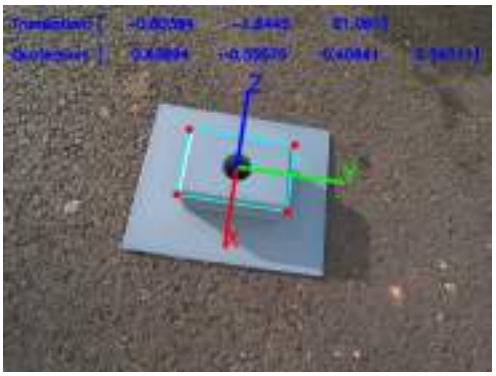


Figure 8. The pose estimation result for the corner model using PnP algorithm.

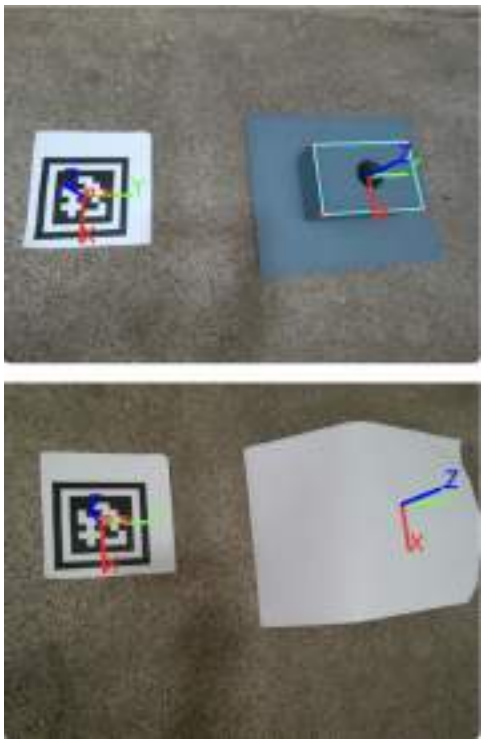


Figure 9. Pose estimation with the visible corner model (upper) and the occluded corner model using the Apriltag marker (lower).

segmented surfaces are fitted with quadrilaterals to position the four corners. By integrating the corresponding depth map, we accurately determine their 3D coordinates. Furthermore, using the 2D coordinates from the colour map and the 3D coordinates with the quadrilateral's center as the origin, we utilize the PnP algorithm for precise pose estimation. However, the 9mm measurement error is relatively high compared to the required precision in MIC. Future work will focus on algorithm optimization, addressing depth map resolution and sensor noise, and extending testing to real-world construction environments to evaluate performance under complex conditions.

References

[1] Sherif Abdelmageed and Tarek Zayed. A study of literature in modular integrated construction-critical review and future directions. *Journal of Cleaner Production*, 277:124044, 2020. doi:10.1016/j.jclepro.2020.124044.

[2] S Marzieh Bagheri, Hosein Taghaddos, and Ulrich Hermann. Automated safety and practicality enhancement of lift plans in modular construction. *Automation in Construction*, 168:105731, 2024. doi:10.1016/j.autcon.2024.105731.

[3] Kamyab Aghajamali, Ala Nekouvaght Tak, Hosein Taghaddos, Ali Mousaei, Saeed Behzadipour, and Ulrich Hermann. Planning of mobile crane walking operations in congested industrial construction sites. *Journal of Construction Engineering and Management*, 149(7):04023047, 2023. doi:10.1061/JCEMD4.COENG-13109.

[4] Shuyuan Xu, Jun Wang, Wenchi Shou, Tuan Ngo, Abdul-Manan Sadick, and Xiangyu Wang. Computer vision techniques in construction: a critical review. *Archives of Computational Methods in Engineering*, 28:3383–3397, 2021. doi:10.1007/s11831-020-09504-3.

[5] Siddharth Singh Chouhan, Ajay Kaul, and Uday Pratap Singh. Image segmentation using computational intelligence techniques. *Archives of Computational Methods in Engineering*, 26:533–596, 2019. doi:10.1007/s11831-018-9257-4.

[6] Farhana Sultana, Abu Sufian, and Paramartha Dutta. Evolution of image segmentation using deep convolutional neural network: A survey. *Knowledge-Based Systems*, 201:106062, 2020. doi:10.1016/j.knosys.2020.106062.

[7] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015. doi:10.1109/CVPR.2015.7298965.

[8] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*,

pages 234–241. Springer, 2015. doi:10.1007/978-3-319-24574-4_28.

[9] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017. doi:10.1109/TPAMI.2017.2699184.

[10] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017. doi:10.1109/ICCV.2017.322.

[11] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence*, 39(6):1137–1149, 2016. doi:10.1109/TPAMI.2016.2577031.

[12] Leanne Attard, Carl James Debono, Gianluca Valentino, Mario Di Castro, Alessandro Masi, and Luigi Scibile. Automatic crack detection using mask r-cnn. In *2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA)*, pages 152–157, 2019. doi:10.1109/ISPA.2019.8868619.

[13] S Dilaksha, KATO Ranadewa, D Weerasooriya, Agana Parameswaran, and Panchali Weerakoon. Comparative analysis of challenges in manual and automated construction progress monitoring in sri lanka. In *Proceedings The 12th World Construction Symposium— August*, page 380, 2024. doi:10.31705/WCS.2024.30.

[14] Randolph Blake and Hugh Wilson. Binocular vision. *Vision research*, 51(7):754–770, 2011. doi:10.1016/j.visres.2010.10.009.

[15] Radu Horaud, Miles Hansard, Georgios Evangelidis, and Clément Ménier. An overview of depth cameras and range scanners based on time-of-flight technologies. *Machine vision and applications*, 27(7):1005–1020, 2016. doi:10.1007/s00138-016-0784-4.

[16] Song Zhang. High-speed 3d shape measurement with structured light methods: A review. *Optics and lasers in engineering*, 106:119–131, 2018. doi:10.1016/j.optlaseng.2018.02.017.

[17] Bing Liu. Long-distance recognition of crack width in building wall based on binocular vision. In *2021 3rd International Conference on Artificial Intelligence and Advanced Manufacture*, page 1908–1912, New York, NY, USA, 2022. doi:10.1145/3495018.3495513.

[18] Haotian Yu, Yu Huang, Dongliang Zheng, Lianfa Bai, and Jing Han. Three-dimensional shape measurement technique for large-scale objects based on line structured light combined with industrial robot. *Optik*, 202:163656, 2020. doi:10.1016/j.ijleo.2019.163656.

[19] Xiao Xin Lu. A review of solutions for perspective-n-point problem in camera pose estimation. In *Journal of Physics: Conference Series*, volume 1087, page 052009. IOP Publishing, 2018. doi:10.1088/1742-6596/1087/5/052009.

[20] Chao Mi, Yi Liu, Yujie Zhang, Jiaqi Wang, Yufei Feng, and Zhiwei Zhang. A vision-based displacement measurement system for foundation pit. *IEEE Transactions on Instrumentation and Measurement*, 72:1–15, 2023. doi:10.1109/TIM.2023.3311069.

[21] Alex Kendall, Matthew Grimes, and Roberto Cipolla. Posenet: A convolutional network for real-time 6-dof camera relocalization. In *Proceedings of the IEEE international conference on computer vision*, pages 2938–2946, 2015. doi:10.1109/ICCV.2015.336.

[22] Prabesh Paudel and Kyoung-Ho Choi. A deep-learning based worker's pose estimation. In *Frontiers of Computer Vision: 26th International Workshop, IW-FCV 2020, Ibusuki, Kagoshima, Japan, February 20–22, 2020, Revised Selected Papers 26*, pages 122–135. Springer, 2020. doi:10.1007/978-981-15-4818-5_10.

[23] Roshan Panahi, Joseph Louis, Ankur Podder, Colby Swanson, and Shanti Pless. Automated assembly progress monitoring in modular construction factories using computer vision-based instance segmentation. In *Computing in Civil Engineering 2023*, pages 290–297. 2024. doi:10.1061/9780784485224.036.

[24] Zhenjie Zheng, Zhiqian Zhang, and Wei Pan. Virtual prototyping-and transfer learning-enabled module detection for modular integrated construction. *Automation in Construction*, 120:103387, 2020. doi:10.1016/j.autcon.2020.103387.

[25] Juan Terven, Diana-Margarita Córdova-Esparza, and Julio-Alejandro Romero-González. A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas. *Machine Learning and Knowledge Extraction*, 5(4): 1680–1716, 2023. doi:10.3390/make5040083.

Less is Better: A Comparative Analysis of Guidance Methods for Long-Context Window-Based Legal Interpretation

Jaekun Lee[1] **and Ghang Lee**[1,2*]

[1]Department of Architecture and Architectural Engineering, Yonsei University, Seoul, Republic of Korea
[2]Institute for Advanced Studies, Technical University of Munich, Munich, Germany
*Corresponding Author

jake102518@yonsei.ac.kr, glee@yonsei.ac.kr

Abstract -

More than 500 statutes and rules are interlinked with building codes and regulations, with 24 amendments introduced in 2023. This complex network of references, combined with frequent revisions, poses significant challenges for both legal experts and laypeople in interpreting building codes. To address this, prior research has explored the application of retrieval-augmented generation (RAG) techniques to large language models (LLMs). However, these efforts have struggled to retain contextual information due to data segmentation. This study proposes and evaluates six variants of zero-shot Chain-of-Thought (CoT) prompts, leveraging long-context window (LCW) LLMs for retrieving relevant legal texts. Among these, the prompt requiring minimal human guidance achieved 63.16% accuracy on a dataset of 171 legal interpretative question-answer (LIQA) pairs, outperforming other variants by up to 5.7%p. Furthermore, incorporating a high-level community summary improved the model's robustness to prompt variations. This approach reduced the need for extensive document hops when locating relevant provisions, narrowing the performance gap to 1.17%p. Lastly, the LCW LLM consistently referenced definition provisions, ensuring greater consistency in legal interpretation. The less the guidance, the better the performance. Less is Better!

Keywords -

Legal interpretation; Building code; Long context window; Multi-hop reasoning

1 Introduction

Legal interpretation involves clarifying the meaning and intent of legal texts, such as building codes and regulations, to ensure their accurate and proper application to specific cases. Understanding these codes is challenging, even for professionals, due to their complexity and frequent updates. In South Korea, more than 500 statutes and rules are intricately linked to building codes and regulations [1], with 24 amendments introduced in 2023 [2]. This complexity often requires reviewing over 400 provisions to process a single legal action such as issuing building permits [3]. The scale of this challenge is evident in the over 900,000 legal interpretation petitions related to housing and construction filed annually in South Korea [4]. Notably, a study conducted in 2017 revealed that 96.7% of handwritten petitions focused on legal interpretation issues [5].

Recent studies have leveraged large language models (LLMs) to minimize the manual effort and cost involved in legal interpretation. To enhance the capabilities of vanilla LLMs, three techniques have been commonly investigated: fine-tuning, retrieval-augmented generation (RAG), and prompt engineering. Among the three methods, RAG-based approaches are widely employed to deliver up-to-date factual information and mitigate hallucinations [1, 6, 7]. However, their performance depends significantly on the structure of the underlying database, emphasizing the need for a database designed to optimize efficient information retrieval [1]. Despite these efforts, data must still be segmented into chunks, resulting in contextual loss. This limitation can lead to the retrieval of documents that contain relevant keywords but present them within unrelated or misleading contexts [7].

To mitigate RAG's context loss, long context window (LCW) LLMs have recently gained attention. By accommodating longer input sequences, LCW LLMs demonstrate advantages in capturing relationships between pieces of information spread across extensive text [8]. However, these models still face challenges, including limitations in context length and the "lost in the middle" problem, where they tend to prioritize information at the beginning and end of the input sequence [9].

To overcome the limitations of existing methods, we propose and compare six variants of zero-shot Chain-of-Thought (CoT) prompts to identify effective strategies for utilizing a long context window in complex tasks such as legal interpretation. These variants differ in their approaches to retrieving relevant information: unguided, exhaustive, and structured search. Exhaustive search allows the model to perform unlimited recursive searches within and across statutory documents, while structured search guides the model to retrieve information based on the hi-

erarchical structure of the legal texts. Additionally, we assess the impact of providing a high-level community summary (HLCS) [10], which includes the titles of each legal provision, for all prompt variants.

For validation, we constructed the *legal interpretative question-answer* (LIQA) dataset using the Building Statutes Question Answering (QA) document published by the Ministry of Government Legislation (MOLEG) [11]. This document contains 171 QA cases requiring interpretation of legal provisions, as answers are not explicitly stated within the legal texts. The LLM identifies relevant provisions using each prompt variant to generate answers, which are then evaluated for accuracy to assess prompt effectiveness.

This paper addresses various terms related to natural language processing (NLP). *Multi-hop search* refers to retrieving information needed to answer a user inquiry through logical steps applied to the provided documents [12]. Specifically, *intra-document hop* refers to hops occurring within a single document, while *cross-document hop* involves hops between multiple documents [13]. *Contextual reasoning* is a type of reasoning task that involves interpreting a small set of factual information specific to the given problem [14]. *High-level community summary* is a summary of a group of data to understand the global structure and semantics of the dataset [10]. For Korean building codes and regulations, we adhere to the English terms provided by the official English-Korean legal terminology system [15, 16]. Additionally, the term *provision* is used throughout this paper to collectively refer to both attached forms and articles, including their sub-hierarchies.

2 Relevant Study

2.1 Retrieval-Augmented Generation

To accurately address a legal inquiry, it is essential to first retrieve the relevant provisions from a large legal corpus. Zhong [17] noted that while legal experts tend to resolve legal tasks by building well-defined rule-based or symbolic systems, NLP researchers prioritize legal information retrieval to support data-driven interpretation. Recently, RAG has gained significant attention as a method to efficiently supply task-specific information to vanilla LLMs [6].

Savelka et al. [7] assessed GPT-4's ability to explain legal terms using RAG by providing court case data that illustrated how statutory terms had been previously interpreted. Their findings demonstrated that incorporating legal information retrieval improves the quality of LLM-generated interpretations, particularly in terms of factual accuracy, clarity, relevance, information richness, and precision. Similarly, Cho and Kim [1] developed the Semantic Processing for Architecture Regulation Compliance (SPARC) model, a RAG-based framework. Their evaluations showed that the system attained high accuracy rates for general inquiries (81.8%) and moderate accuracy for complex legal interpretation cases (56.1%).

2.2 Prompt Engineering and Long Context Window

Since language models have limitations in their input sequence length, prompt engineering is aimed at utilizing the available context window as efficiently as possible. Blair-Stanek et al. [18] tested GPT-3's ability to answer statutory reasoning questions about U.S. tax codes using zero-shot, few-shot, and CoT prompting. Their findings revealed that GPT-3 exhibited incorrect prior knowledge of the U.S. tax codes. However, two-shot prompting improved accuracy by an average of 20.9%p compared to zero-shot result. Fuchs et al. [19] evaluated GPT-3.5's ability to generate a structured XML format of building regulations under four conditions: few-shot, detailed description, CoT, and self-consistency prompting. The study demonstrated that LLMs can translate regulations into structured formats without fine-tuning.

Although prompt engineering optimizes the use of a given sequence length, ensuring a large context window is crucial for long documents [20]. Previous studies have focused on developing LLMs with an expanded context window [8] or proposing frameworks to increase the amount of information that can be processed within the existing limits [20, 21]. In February 2024, Google introduced Gemini 1.5 pro, an LCW LLM capable of handling up to 10 million tokens. According to the official report [8], the model achieved over 99% recall in multimodal information retrieval tasks.

2.3 Research Gaps and Questions

While RAG enhances information retrieval in LLMs, its application in legal interpretation has several limitations. First, retrieval methods such as similarity search often fail to capture essential information due to their inability to recognize contextual relationships between predivided data chunks. For example, a retrieved sentence might contain relevant legal terminology but appear in an irrelevant or misleading context [7]. Second, addressing these retrieval failures in complex tasks requires extensive pre-processing, including the creation of a well-structured database, which is both time-intensive and costly [1].

Moreover, despite their strong performance in information retrieval, LLMs with LCW also remain imperfect. One notable issue is the "lost-in-the-middle" problem, where the model disproportionately focuses on the beginning and end of a prompt, even when the context window has not been fully utilized [22]. Additionally, vanilla LCW LLMs often perform poorly on domain-specific tasks, such as legal interpretation.

To enhance the interpretation of building codes, which demand both comprehension and the ability to parse relevant provisions from extremely long contexts, this study evaluates various types of Chain-of-Thought (CoT) prompts using LCW LLMs. The following sections present the proposed prompt designs, the implementation methods, and the experimental evaluation process.

3 Proposed Method

We propose six variants of zero-shot prompts to enable the retrieval of provisions without the need for frameworks like RAG or fine-tuning. These retrieval prompts are grouped into three categories: unguided search, exhaustive search, and structured search. Each category includes two versions—one with a high-level community summary and one without—yielding a total of six variants.

3.1 General Prompt Structure

All prompts are created based on the structure outlined in *prompting guide 101* by Google Gemini Team [23]. The guide recommends incorporating four components in a prompt: persona, task, context, and format, though including all of them is optional. The persona defines the role or perspective, the task specifies the objective, the context provides situational details, and the format outlines the desired response style. The guide also highlights the importance of using a clear verb or directive to define the task. In this paper, we propose a basic prompt structure, as illustrated in Basic Prompt, where braces denote variable or optional elements specific to the three prompt variants.

Basic Prompt "You are designed to assist users in locating legal provisions related to the Korean Building Code as they apply to inquiries. {**HLCS**} {**Context**} {**Format**}"

3.2 Prompt Variants

3.2.1 Unguided Search

The prompt used in unguided search does not include detailed CoT instructions. Instead, it provides only a command in the {Context} field. Additionally, the response format is defined as tuples within a list, as described below.

{**Context**}: "From the given files and inquiry, find all articles and attached forms related to the inquiry."
{**Format**}: "Return the law name and article or attached form number that you found as tuples in a list format. For example: [(Building Act, Article 3-2), (Enforcement Rule of the Building Act, Article 12), (Enforcement Rule of the Building Act, Attached Form 1)]. If it is not an attached form, number must include the word 'Article'. Do not directly answer the inquiry."

3.2.2 Exhaustive search

In exhaustive search, LCW LLM is guided to perform a recursive provision search, hopping between all interreferencing provisions until it encounters provisions unrelated to the inquiry. Before returning the final list, it evaluates the relevance of each provision and includes only the necessary ones. To track the multi-hop search path, results are returned in a dictionary format, where the key represents the found provision and the value represents its parent provision.

{**Context**}: "Using the given files and inquiry: 1. Identify the articles and attached forms related to the inquiry. 2. Locate other provisions that reference the provisions identified in Step 1. 3. Recursively find additional provisions that reference those identified in the previous step until provisions unrelated to the inquiry are reached. In subordinate codes, references may appear explicitly, such as 'Article 12 Paragraph 1 of the Act'. 4. From the provisions gathered, filter and return only those necessary to address the inquiry."
{**Format**}: "5. Return the results in a dictionary format, where each key is a provision, and its value is the provision that mentioned the key (its parent). Each element must be formatted as (law name, article or form number). If the provision is related to the inquiry itself, its value should be 'Inquiry'. For example: '(Building Act, Article 3-2)': 'Inquiry', '(Enforcement Rule of the Building Act, Article 28)': '(Building Act, Article 3-2)', '(Enforcement Rule of the Building Act, Attached Form 1)': '(Building Act, Article 3-2)'. If the provision is not an attached form, the number must include the word 'Article'. Do not directly answer the inquiry."

3.2.3 Structured search

In structured search, the context part is organized according to the hierarchical structure of the Korean building codes. For each Act, specific provisions necessary for its enforcement are determined by the Decree, while provisions are only incorporated into the Rule if they pertain to technical or minor matters requiring frequent updates [24]. Consequently, the search process is structured to prioritize provisions in the order of Act, Decree, and Rule. However, since some inquiry topics are not explicitly addressed in the Act, the model is allowed to initiate searches from either the Act or the Rule. The response format is the same as that used in unguided search, as described in Section 3.2.1.

{**Context**}: "Using the given files and inquiry: 1. Identify the articles and attached forms in the Act or Rule related to the inquiry. 2. Locate other provisions in Rule that reference the provisions identified in Step 1. 3. Find additional provisions in Decree that reference those identified in the previous steps. In subordinate codes, references may appear explicitly, such as 'Article 12 Paragraph 1 of the Act'. 4. From the provisions gathered, filter and return only those necessary to address the inquiry."

3.3 High-level Community Summary

For each prompt variant, two versions—one with HLCS and one without—are tested. In this study, the titles of provisions are used as summaries, as they are designed to concisely convey what each provision regulates, facilitating convenient search [24]. The HLCS prompt is provided below.

{**HLCS**}: Each article's title is provided in a text file named after the law, followed by '(Provision Title)', which can help determine relevance.

4 Experimental Design

4.1 Datasets

The dataset used to validate our prompts is derived from the Building Statutes Question Answering (QA) document, which includes 171 statutory building code interpretation cases across 21 categories, selected by MOLEG [11]. This document is well-suited for use as ground truth in evaluating the prompts because MOLEG provides standardized legal interpretation guidelines in South Korea to ensure consistency across administrative agencies' interpretations [25]. Additionally, the cases in this document require interpreting the meaning and relationships between provisions, as the answers are not explicitly stated in the legal texts.

Each case in the document is structured into four parts: the title, an inquiry, a response, and the reasoning. As illustrated in Figure 1, we developed the LIQA dataset by using the inquiry summaries as questions and compiling the corresponding ground truth answers as follows: the title served as the inquiry summary, the response as the comprehensive conclusion, and the reasoning as the basis of judgment.

4.2 Experimental Setup

Figure 1 also illustrates the experimental flow for evaluating the six prompt variants. Each variant searches and generates a list of relevant provisions for each question in the LIQA dataset. The content of the provisions in each list is then passed to the answer-generating LLM to produce the final answer. The prompt for the answer-generating LLM is shown in Response Prompt.

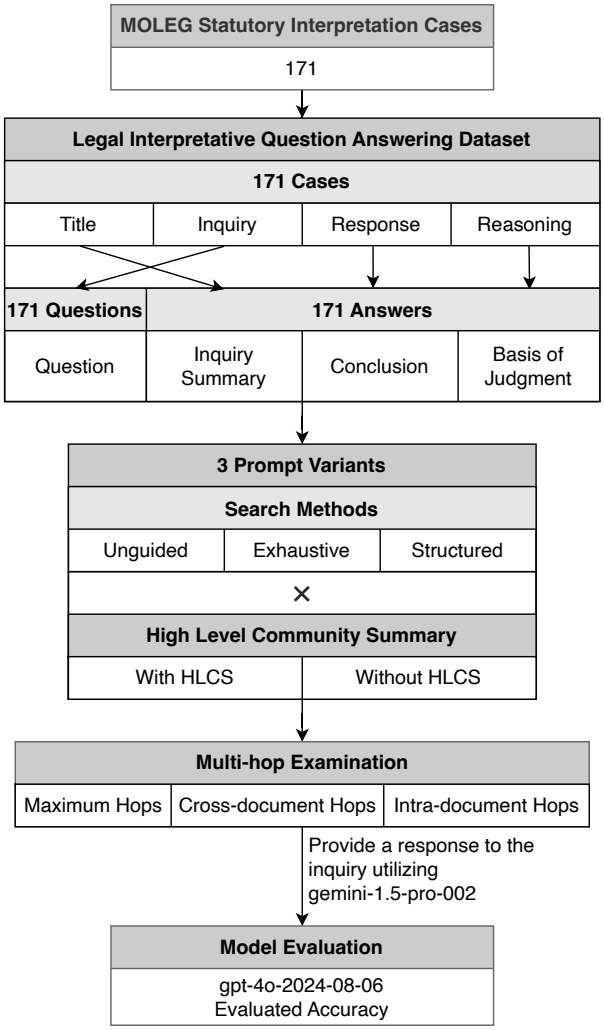


Figure 1. Flow chart of the experiment

Response Prompt You are designed to assist users in interpreting legal provisions related to the Korean Architecture Code. Your primary role is to provide interpretations of the laws as they apply to inquiries in Korean. You must analyze the legal texts in the provided documents or prompt only. The answer must consist of 3 parts: 'Inquiry Summary', 'Conclusion', 'Basis of Judgment'. The first part, 'Inquiry Summary', describes the core content of the inquiry. The second part, 'Conclusion', describes the answer or conclusion of the question. The last part, 'Basis of Judgment', contains the legal basis of the answer described in 'Conclusion'.

The proposed framework can accommodate any LCW LLM with a sufficiently large context window to handle the input data. For this experiment, we used the gemini-1.5-pro-002 model as the LCW LLM for CoT prompt processing and answer generation. To stay within context limits, provision search was restricted to the Building Act and its subordinate regulations. For hyperparameters, we followed Google AI Studio's defaults: *temperature* = 1, *top_p* = 0.95, *top_k* = 40, and *max_output_tokens* = 8192. Safety settings were disabled "threshold": "BLOCK_NONE") to prevent unintended response blocking. During answer generation, *temperature* was increased to 1.2 for enhanced creativity [26].

During answer generation, the content of the statutes is passed to the LLM as a string, retrieved via the Law Open Data Service through an Open API (hereafter referred to as LawAPI) provided by the Korean Law Information Center under MOLEG. For attached forms, however, the official PDF file of the form is provided to ensure the tabular format is preserved accurately.

The experiment was conducted on a MacBook Pro 14 (Model MacBookPro18,3) equipped with an Apple M1 Pro chip, 16 GB RAM, 10 CPU cores (8 performance cores and 2 efficiency cores), 12 GPU cores, and running macOS Sonoma 15.2. The application software used was Visual Studio Code version 1.96.2, with Python 3.11.5.

4.3 Accuracy Evaluation

The accuracy of each prompt variant was evaluated by comparing the final answer generated by the model to the ground truth answer. To ensure an efficient and fair evaluation, we employed a gpt-4o-2024-08-06 based assessment agent using the Evaluation Prompt. Q, R_H, and R_S represent a question, human response, and system response, braces indicating their variability by inquiry and prompt.

Evaluation Prompt Compare human_response and system_response for original_question. Return True if both responses reach the same conclusion or agree, otherwise return False.
original_question: {Q}
human_response: {R_H}
system_response: {R_S}

4.4 Multi-hop Examination

To trace the sequence of hops taken by the model during an exhaustive search, we instructed the LLM to return its results in a dictionary format, where the key represents the child (destination) and the value represents the parent (origin), as described in Section 3.2.2. The number of hops performed was then analyzed by converting the results into a tree format using Python code.

5 Results

5.1 Accuracy

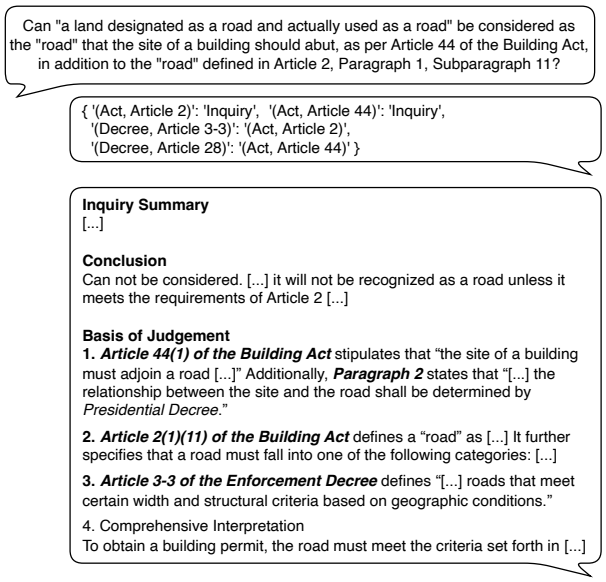


Figure 2. Example result of a Exhaustive Search + HLCS prompt.

Figure 2 depicts an example result when the exhaustive search prompt with HLCS is applied. Figure 3 presents the accuracy distribution for six prompt variants across different inquiry categories, with average accuracies represented by diamond marks and detailed in Table 1. Without HLCS, unguided search achieved the highest average accuracy of 63.16%, followed by exhaustive search at 60.82% and structured search at 57.89%. With HLCS, the accuracy gap between the prompt types narrowed, with structured search leading at 62.57%, followed closely by exhaustive (61.99%) and unguided (61.40%) search.

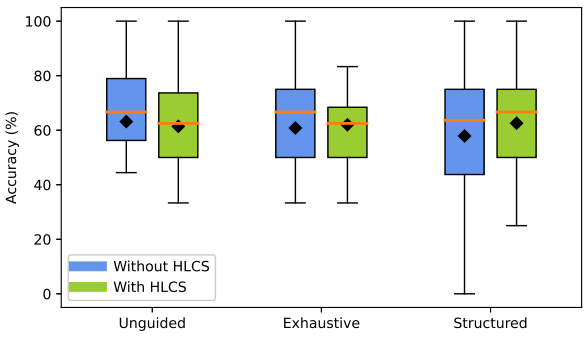


Figure 3. Comparison of six prompt variants on accuracy for each inquiry subject. Diamond marks indicate the weighted accuracy of each variant.

Table 1. Weighted average accuracy(%) of each prompt variant with and without HLCS. Ung, Exh, and Str stand for unguided, exhaustive, and structured search, respectively.

	Ung	Exh	Str
Without HLCS	**63.16**	60.82	57.89
With HLCS	61.40	61.99	62.57

5.2 Multi-hop Examination

Table 2 presents the top 5 most commonly mentioned provisions in the inquiries and the top 5 most frequently referenced provisions by the two exhaustive search prompts. Comparing the results reveals that Article 11 of the Building Act and Attached Form 1 of its Enforcement Rule are frequently both mentioned and referenced. Notably, the search prompts often referenced Article 2 of both the Act and the Rule, even though these were not explicitly mentioned in the inquiries. Article 2 in both cases contains definitions of the legal terms used in the content.

Table 2. List and count of the provisions commonly mentioned and searched. AF stands for Attached form. The number within parentheses represents frequencies (number of inquiries)

Idx	Mentioned	Exhaust / Exhaust+HLCS
1st	Act, 11 (25)	**Act, 2 (101)** / Rule, 2 (61)
2nd	Rule, AF 1 (20)	Rule, 2 (89) / Act, 2 (59)
3rd	Rule, 119 (15)	Rule, AF 1 (53 / 39)
4th	Rule, 86 (14)	Act, 11 (38) / Act, 11 (36)
5th	Act, 80 (10)	Decree, 6 (36) / Rule, 28 (30)

Figure 4 displays the percentile rank of the number of provisions referenced during exhaustive search, with and without HLCS. The plain exhaustive search referenced more provisions beyond the 97th percentile, with an average of 10.70 provisions compared to an average of 7.62 provisions for the HLCS-equipped search. Moreover, the HLCS-equipped LCW LLM performed fewer hops compared to the model without HLCS (Figure 5). The average length of the longest sequence of hops per inquiry was 2.91 for Exhaust+HLCS, which is 0.44 fewer than that of the model without HLCS (3.35). Specifically, it averaged 0.23 fewer cross-document hops (CDH) (1.62 vs 1.85) and 0.11 fewer intra-document hops (IDH) (1.22 vs 1.33) per inquiry.

Table 3 presents the average execution time of the six prompt variants for generating the list of relevant provisions. Exhaustive search took the longest time, both with and without HLCS, followed by structured search and unguided search. Meanwhile, the average prompt token count was similar across models—546K without HLCS and 556K with HLCS—since the majority of tokens came from the provided Building Act and its subordinate regulations.

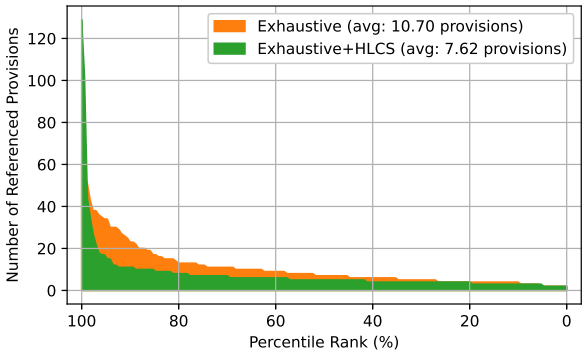


Figure 4. Comparison of the number of referenced provisions with and without HLCS across percentile ranks

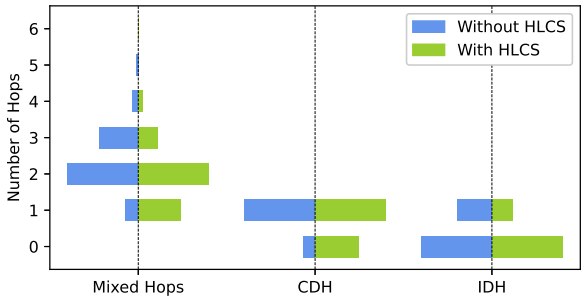


Figure 5. Horizontal bar plot illustrating the relative frequency of the number of hops taken by the exhaustive search.

Table 3. Average execution time(sec) of each prompt variant with and without HLCS.

	Ung	Exh	Str
Without HLCS	44.08	67.10	54.83
With HLCS	44.55	**83.06**	76.23

6 Discussion

When analyzing the trend in average response accuracy across six prompt variants, a counterintuitive pattern emerged: less guidance in the prompt for locating relevant provisions led to higher accuracy. Furthermore, incorporating HLCS improved the accuracy. Without HLCS, the accuracy gap between prompts reached up to 5.27 percentage points, whereas with HLCS, this gap narrowed to just 1.17 percentage points—a reduction of approximately 4.5 times. As detailed in Section 3.3, we provided provision titles as HLCS, and the results confirm their effectiveness

in legal document retrieval. However, while HLCS enhanced stability, it did not yield the highest performance, indicating that its primary effect is to ensure consistency in search results rather than directly enhancing accuracy.

HLCS's stability enhancement is further examined by the findings in Section 4.4. When the same prompt was used, the inclusion of HLCS reduced the average number of referenced legal provisions per inquiry by 3.08. Figure 5 shows HLCS reduced the length of multi-hop searches, making one- or two-hop searches predominant. Notably, the average number of cross-document hops decreased by 0.23, a more substantial reduction than the 0.11 decrease in intra-document hops. These results suggest that HLCS helps minimize unnecessary navigation and exploration across different documents. However, as shown in Table 3, using HLCS significantly increased execution time—rising by a factor of 1.24 in the case of exhaustive search—even though token usage was only 1.02 times higher. This may be attributed to the LLM repeatedly referencing the entire HLCS for each hop.

Table 2 showed that Article 2 of the Building Act and its Enforcement Decree, which are both definition provisions, were the most referenced. Exhaustive search prompt referred definition provisions in 59.1% of the inquiries. As noted by MOLEG [16], definition provisions are designed to clarify the meanings of terms within laws and regulations, ensuring their interpretation and application remain unambiguous. The frequent reference to these provisions suggests that the LCW LLM has the ability to provide consistent legal interpretations even when presented with new inquiries. Since we did not fine-tune or apply RAG with dataset-specific information, our findings may extend to other legal systems or domains.

7 Conclusion

This study designed and evaluated six prompt variants for LCW LLM-based legal interpretation, focusing on the unique challenges posed by the cross-referencing and hierarchical structure of legal texts. Notably, the exhaustive prompt aligned with the cross-referencing structure of laws, while the structured search prompt targeted hierarchical legal frameworks. Additionally, we investigated the influence of HLCS on model performance.

For evaluation, we developed the LIQA dataset, comprising 171 building code interpretation cases curated by MOLEG. Our findings revealed that providing less guidance to the model reduced the number of referenced provisions, improving response accuracy by up to 5.7%p. While HLCS helped stabilize model performance by narrowing the performance gap to 1.17%p, it did not maximize accuracy. Furthermore, the LCW LLM frequently referenced definition provisions, up to 59.1% of total inquiries, which contributed to consistency in legal interpretation.

Our research offers three key contributions. First, it demonstrates that minimizing human guidance enhances model performance when interpreting extensive and intricately cross-referenced legal data. Second, it highlights how the HLCS enhances stability across various prompt designs, reducing result variability and thereby lowering the manual effort and costs associated with LLM-based legal interpretation. Lastly, the LCW LLM consistently referenced definition provisions, contributing to more consistent and reliable interpretation results.

Future work could explore alternative LLM techniques, such as fine-tuning or RAG, to further optimize performance. Additionally, since the model typically requires fewer than three hops to locate relevant provisions, predefining the top three related provisions could help reduce token costs while maintaining high accuracy in legal interpretation tasks. Lastly, the designed guidance methods should be evaluated across various LLMs and multiple legal QA datasets to assess their generalizability, along with a qualitative comparison of the generated answers.

Acknowledgments

This work is supported in 2025 by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant RS-2021-KA163269 and RS-2024-00407028) and the Hans Fischer Senior Fellowship program at the Technical University of Munich - Institute for Advanced Studies (TUM-IAS) in Germany.

References

[1] S.-K. Cho and S.-S. Kim. *Study on the Efficient Response to Architectural Civil Complaints Using Large Language Models (LLM)*, volume 1. Architecture Urban Research Institute, Sejong, 2023.

[2] Y.-B. Lee. *Geonchug Beobje Donghyang 2023 [Building Legislation Trends 2023]*, volume 1. Architecture Urban Research Institute, Sejong, 2023.

[3] MOLIT. Korean building code, attached form 1. On-line: `https://www.law.go.kr/œ/œμœ/(2024-206,20240419)`, Accessed: 01/13/2025.

[4] Y.-G. Lee, J.-G. Lee, M.-J. Kim, and Y.-E. Hong. *Hannun-e Boneun Geonchugminwon Bigdata 2021 [Construction Civil Complaints Big Data 2021 at a glance]*, volume 1. Architecture Urban Research Institute, Sejong, 2022.

[5] S.-Y. Ryu, M.-J. Kim, Y.-J. Cho, and K.-H. Yu. A study on the analysis of civil appeals to building regulations. *Journal of the architectural insti-*

tute of Korea planning design, 33:13–21, 2017. doi:10.5659/JAIK_PD.2017.33.2.13.

[6] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W. Yih, T. Rocktäschel, S. Riedel, and D. Kiela. Retrieval-augmented generation for knowledge-intensive nlp tasks. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020. doi:10.48550/arXiv.2005.11401.

[7] K.D. Gray M.A. Westermann H. Xu H. Savelka, J. Ashley. Explaining legal concepts with augmented large language models (gpt-4). In *arXiv*, Preprint. doi:10.48550/ARXIV.2306.09525.

[8] Gemini Team. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. In *arXiv*, Preprint. doi:10.48550/arXiv.2403.05530.

[9] K. Hewitt J. Paranjape A. Bevilacqua M. Petroni F. Liang P. Liu, N.F. Lin. Lost in the middle: How language models use long contexts. *Transactions of the Association for Computational Linguistics*, 12: 157–173, 2024. doi:10.1162/tacl_a_00638.

[10] Darren Edge, Ha Trinh, Newman Cheng, Joshua Bradley, Alex Chao, Apurva Mody, Steven Truitt, and Jonathan Larson. From local to global: A graph rag approach to query-focused summarization. In *arXiv*, Preprint. doi:10.48550/arXiv.2404.16130.

[11] Korea Institute of Registered Architects. *GeonChug-BeobLyeong Jil-UiHoeSinJib [Building provision inquiry collection]*, volume 1. MOLIT and KIRA, Seoul, 2023.

[12] W. Xiong, X. Li, S. Iyer, J. Du, P. Lewis, W.Y. Wang, Y. Mehdad, S. Yih, S. Riedel, D. Kiela, and B. Oguz. Answering complex open-domain questions with multi-hop dense retrieval. In *Proceedings of the 2021 International Conference on Learning Representations (ICLR)*, pages 2590–2602, Hong Kong, China, 2024.

[13] K. Lu, I.-H. Hsu, W. Zhou, M.D. Ma, and M. Chen. Multi-hop evidence retrieval for cross-document relation extraction. In *Findings of the Association for Computational Linguistics (ACL)*, pages 10336–10351, Toronto, Canada, 2024.

[14] F. Giunchiglia. Contextual reasoning. *Epistemologia*, 16:345–364, 1993.

[15] KLRI. Glossary, english law system. On-line: `https://elaw.klri.re.kr/eng_service/lawsystem.do`, Accessed: 10/06/2024.

[16] MOLEG. Beoblyeong ib-ansimsa gijun [standards for drafting and reviewing laws]. On-line: `https://www.lawmaking.go.kr/lmKnlg/jdgStd/list`, Accessed: 10/06/2024.

[17] H. Zhong, C. Xiao, C. Tu, T. Zhang, Z. Liu, and M. Sun. How does nlp benefit legal system: A summary of legal artificial intelligence. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020. doi:10.18653/v1/2020.acl-main.466.

[18] Blair-Stanek, A. Holzenberger, and B. N. Van Durme. Can gpt-3 perform statutory reasoning? In *Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law*, pages 22–31, New York, USA, 2023.

[19] S. Fuchs, M. Witbrock, J. Dimyadi, and R. Amor. Using large language models for the interpretation of building regulations. In *arXiv*, Preprint. doi:10.48550/arXiv.2407.21060.

[20] S. Chen, S. Wong, L. Chen, and Y. Tian. Extending context window of large language models via positional interpolation. In *arXiv*, Preprint. doi:10.48550/arXiv.2306.15595.

[21] W. Wang, L. Dong, H. Cheng, X. Liu, X. Yan, J. Gao, and F. Wei. Augmenting language models with long-term memory. In *Advances in Neural Information Processing Systems (NeuralPS)*, 2023. doi:10.48550/arXiv.2306.07174.

[22] N.F. Liu, K. Lin, J. Hewitt, A. Paranjape, M. Bevilacqua, F. Petroni, and P. Liang. Lost in the middle: How language models use long contexts. *Transactions of the Association for Computational Linguistics*, 12: 157–173, 2024. doi:10.1162/tacl_a_00638.

[23] Gemini Team. *Prompting guide 101*, volume 2024 edition. Gemini for Google Workspace, California, USA, 2024.

[24] MOLEG. *Beoblyeong Ib-anSimsa Gijun [Standards for Drafting and Reviewing Laws]*. MOLEG, Sejong, South Korea, 2023.

[25] MOLEG. Beoblyeong haeseog annae [about legal interpretation], legislative work information (2024). On-line: `https://www.moleg.go.kr/menu.es?mid=a10106010000`, Accessed: 01/13/2025.

[26] Gemini Team. About generative models. On-line: `https://ai.google.dev/gemini-api/docs/models/generative-models`, Accessed: 02/27/2025.

Automating WMA Construction Planning: A Physics-aware Framework

Q. Shen[1], F. Vahdatikhaki[1], S.R. Miller[1], and A.G. Dorée[1]

[1,] Department of Construction Management and Engineering, University of Twente, the Netherlands
q.shen@utwente.nl, f.vahdatikhaki@utwente.nl, s.r.miller@utwente.nl, a.g.doree@utwente.nl

Abstract –

The Dutch road construction industry is transitioning from hot mix asphalt (HMA) to more eco-friendly warm mix asphalt (WMA). While WMA offers clear environmental benefits, its distinct material properties, particularly the compactability, demand different construction practices. However, limited industry knowledge hinders WMA construction planning, driving the need for automated tools for more efficient WMA construction strategic planning. This research proposed a physics-aware surrogate modelling framework to improve strategy evaluation and optimization, and enable the automative exploration of optimal construction strategies. A preliminary feasibility study demonstrated the possibility of developing a surrogate model using Random Forest to efficiently capture the temperature-dependent compactability of a WMA mixture. The industrial adoption and potential advancement of this tool, including the integration with other sensing technologies and automatic construction, to ensure high-quality pavement construction were also discussed.

Keywords –

Asphalt construction; Warm mix asphalt; Physics-based simulation; Surrogate model

1 Introduction

In the Netherlands, in response to ever-growing environmental concerns about hot mix asphalt (HMA), the road construction industry is compelled to adopt more sustainable practices, e.g., exploring alternative materials. This created a strong ambition to transition from HMA to warm mix asphalt (WMA). For instance, the Dutch road construction industry is determined to complete this shift by January 2025. Compared to HMA, WMA is produced at lower temperatures (100–140 ℃) using synthetic or organic additives. It offers notable environmental benefits, including reduced emissions, lower fossil fuel consumption, and increased use of reclaimed asphalt pavement (RAP) without compromising mechanical performance [1]. Despite the change in the material, the contractors' primary goal remains the same, i.e., ensuring the asphalt construction process quality to achieve the desirable product quality, such as optimal density. Specifically, the asphalt must be compacted within an ideal temperature window since compaction at high temperatures can cause permanent deformation and compaction at low temperatures can hinder the densification process. However, asphalt construction is widely recognised as a dynamic environment where uncertainties abound. These uncertainties, coupled with the time-sensitive nature of road construction, make achieving optimal compaction efficiency highly challenging. This often results in significant variability during construction, leading to substandard road quality.

These challenges are not unique to HMA construction but are amplified in WMA due to the substantial difference in material characteristics (particularly the different compactability) and limited knowledge about its construction. This makes it more difficult to plan explicit operational strategies (e.g., required quantity of pavers and rollers, paving speed, compaction patterns and trajectories, etc. [2]). Addressing this issue is critical, as a key incentive for contractors to adopt WMA lies in building a solid knowledge base for a stable and sustainable construction process. This would help them meet quality requirements, avoid penalties under extended guarantee periods, and support the push for sustainable road construction. To stay competitive, contractors are motivated to adopt planning tools that can optimize WMA construction strategies by embedding a deeper understanding of WMA's material behaviour into automation techniques, to enable efficient and reliable assessment and adjustments in the decision-making process of WMA construction strategies.

On this premise, this paper presents a framework of a WMA construction planning tool that can provide contractors with an efficient and explicit assessment of pavement quality regarding different construction strategies, thus advancing automation in construction strategic planning. This study revolves around a central question of how a WMA construction planning tool can be developed with the embedment of WMA's unique material features and its structural responses towards different construction strategies. A framework that applies physical modelling and physics-aware surrogate modelling was then proposed. Unlike conventional

models, it better captures WMA's compaction behaviour for more precise decision-making. The results of a preliminary feasibility analysis are presented in this paper.

The remainder of the paper is structured as follows. First, the research problem is investigated to derive the scope of this WMA construction planning tool. This is followed by a demonstration of the framework of developing such a planning tool, with a case study indicating the feasibility of the proposed framework. The paper ends with a conclusion.

2 Problem Investigation

Construction planning involves identifying explicit on-site activities and evaluating their implications [3]. Nonetheless, asphalt construction comprises a multitude of interconnected activities, resulting in considerable complexity for the effective planning of construction strategies.

Previous studies explored the development of asphalt construction planning tools, with particular attention to using simulation techniques, as simulations effectively capture construction processes and test scenarios, enabling more efficient and informed decision-making [4,5]. For instance, Dalence et al. proposed a framework integrating temporal (e.g., equipment, speed, roller passes) and spatial (e.g., trajectory, asphalt cooling rate) characteristics, where data on key factors influencing asphalt compaction is collected to generate decision variables [4]. These variables are fed into simulation models, which provide outputs such as equipment behaviours and operational feedback, facilitating strategy assessment and improvement.

Nevertheless, existing simulation-based asphalt construction planning tools predominantly rely on indirect assessment of the process quality (e.g., paver and roller output, compaction efficiency, and process consistency), where the actual structural response of asphalt mixture (which will be reflected in the pavement product quality, such as density) during the construction process are not explicitly considered. Although, intuitively, good construction process quality leads to good pavement product quality, not being able to precisely quantify the exact impact of construction strategies on the resultant structural responses of the pavement during the simulation can considerably affect the accuracy/reliability of simulation-based assessment of different construction strategies.

Therefore, understanding the interplay between mechanical and thermal characteristics during asphalt compaction is vital for precise construction planning and decision-making. However, the current body of knowledge lacks a clear understanding of the quantifiable correlation between construction quality and pavement properties. In essence, this would make the planning of WMA construction performance-based, where the most suitable on-site operations will be selected based on the resultant pavement quality and performance in other indicators (e.g., cost and productivity). In other fields, the similarity can be found in strategic planning of earth moving activities. Traditional earthwork planning focused on minimizing material transport, relying heavily on workers' experience. However, construction automation has enhanced efficiency and productivity by reducing these constraints, by optimizing operations, particularly coordinating excavators and dump trucks to maximize efficiency [6,7].

To address this limitation, advanced physical modelling is needed to capture the structural and thermal behaviour of WMA during its interaction with equipment and the environment. Two widely applied approaches are Finite Element Method (FEM) [8] and Discrete Element Method (DEM) [9]. FEM can accurately model asphalt's structural response under dynamic loads but lacks granular-level accuracy. In contrast, DEM effectively simulates contacts within granular systems and micro-mechanics. Nevertheless, DEM greatly increases the computational time. To address this, Komaragiri et al. [10], developed a simulation model using a physics engine (i.e., Bullet Physics [11]), which significantly improves computational efficiency without compromising performance. However, existing physics-based simulation models calibrated for HMA are unsuitable for WMA due to differences in material characteristics. Thus, developing physics-based models tailored to WMA is crucial for capturing its unique behaviour during compaction. Additionally, the high computational cost of physics-based models limits their use in process simulation, where analysis of multiple simulation scenarios for optimised processes is always required. One potential solution is to use data-driven techniques, such as machine learning (ML), to quantify the correlation between asphalt construction process indicators and resultant product quality indicators. In the research of Shen [12], on-site data was collected, including real-time machinery movement and asphalt cooling patterns, to capture key asphalt construction process characteristics. The data was then paired with density measurements from pavement samples to develop a dataset for machine learning model training. However, this study also highlighted critical issues, such as limited and inconsistent data availability, ineffective data management practices, and variability in data collection methods, which greatly influenced the accuracy and generalizability of data-driven approaches.

To tackle these difficulties, surrogate modelling offers a solution by approximating physics-based models with data-driven techniques, such as machine learning (ML), allowing for faster predictions without sacrificing accuracy [13]. By generating datasets from physics-

based simulations under varying input conditions, ML-based surrogate models can be developed to predict WMA's structural responses efficiently, enabling the exploration of optimal construction strategies without excessive computational expense.

3 Framework for Physics-aware WMA Construction Planning Tool

3.1 The Scope of the WMA Construction Planning Tool

To identify the scope of the WMA construction planning tool for this research context, a workshop was held with Dutch public clients and contractors in the asphalt construction sector. The workshop explored: (1) the impact of material transitions on WMA compaction, (2) factors influencing strategy formulation, and (3) the structure of the tool. The workshop identified the tool's primary functional requirement: to assess construction strategies accurately and to offer intuitive feedback on compaction quality (primarily density consistency) considering specific project characteristics.

Specifically, the proposed planning tool should begin by extracting pre-construction information, such as design data, ambient conditions, and project constraints, to define the scope of construction strategy exploration. It must evaluate strategies by assessing the resultant pavement quality. This calls for the development and integration of physics-aware models to comprehensively explore the complex system's behaviour, characteristics, and dynamics in WMA construction. The tool's output will focus on two aspects: resource allocation (e.g., equipment, materials, staffing, budgeting) and operational strategies (e.g., trajectory planning, sequencing, and equipment coordination), ensuring timely and cost-effective project execution.

Therefore, the framework of the physics-aware WMA construction planning tool can be determined, as shown in Figure 1. The rest of this chapter will provide a detailed description of each component.

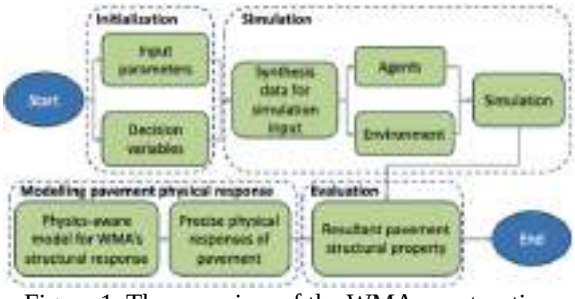


Figure 1. The overview of the WMA construction planning tool

3.2 Initialization

The initialization of the planning tool gathers input parameters and decision variables. Input parameters, which reflect fixed project characteristics, include design and engineering details (e.g., road geometry, and asphalt mixture properties) and project-specific constraints (e.g., equipment configurations), as listed in Table 1.

Identified decision variables, adjustable by decision-makers, focus on resource allocation and operational strategies, with particular emphasis on the compaction strategy in asphalt construction planning, due to its vital role in meeting asphalt pavement's functional requirements. These decision variables allow the exploration of various WMA construction strategies, as demonstrated in Table 2. The compaction strategy involves defining each roller's trajectory, determined by the length and width of the compaction section, and the number of and distance between compaction lanes (Figure 2). Additionally, the mobility of rollers is also affected by the compaction patterns, which regulate the number of compaction corridors and other rules that each roller needs to follow during compaction, especially when multiple rollers are compacting simultaneously. These compaction patterns will be pre-defined by the decision-makers as the input to the planning tool.

Table 1. Identified input parameters of the WMA construction planning tool framework

Data category	Data	Unit
Road geometry	Road width	m
Asphalt	Mixture design	-
	Production temperature	°C
	Initial density	kg/m^3
	Cooling rate	°C/min
Paver	Screed width	m
Truck	Truck capacity	m^3
	Dumping rate	m^3/min
Roller	Drum width	m
	Wheel length	m
	Compaction force	kN

Table 2. Identified decision variables of the WMA construction planning tool framework

Data category	Data	Unit
Resource allocation	Paver quantity	#
	Roller quantity	#
	Truck quantity	#
Operational strategies	Paver speed	m/min
	Roller speed	m/min
	Length of compaction section	m
	Width of compaction section	m
	Number of compaction lanes	#
	Hauling duration	hours
	Returning duration	hours
	Compaction pattern	-

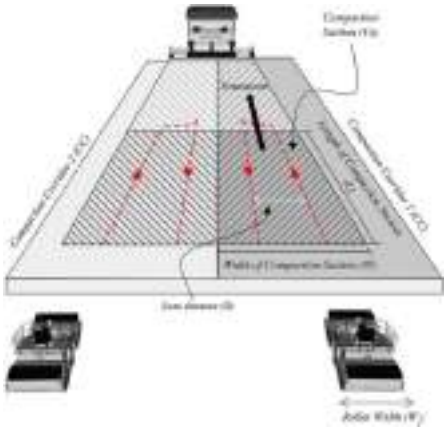


Figure 1. Asphalt compaction strategy [2]

3.3 Simulation

The collected data from the previous layer will then be fed into a simulation model to assess the corresponding impact on the product quality. Specifically, the simulation model will be informed by the previous work of Dalence et al. [4], where an agent-based simulation (ABS) model was proposed to create a context-realistic simulation of the HMA construction process. It is worth noting that despite differences in material properties, WMA and HMA share similar construction logic and workflow, making the ABS simulation model's scope a valuable reference for the WMA. This simulation model aligns closely with the functional requirements of the planning tool and operates within the same contextual framework. Given its demonstrated applicability and relevance, re-developing a new simulation model is deemed unnecessary.

The simulation, using inputs from Tables 1 and 2, defines agents, interactions, and conditions. Figure 3 shows key agents: trucks deliver asphalt, coordinating with the paver for continuous transfer. The paver operates at a set speed, signaling trucks as needed, while rollers compact within the optimal temperature window. Predefined compaction strategies are used to maintain continuity and efficiency. Furthermore, Asphalt cell agents will model asphalt's structural and thermal behavior by dividing the asphalt layer into a grid (Figure 4). Each cell cools at a set rate, while environmental conditions and contractor-provided road geometry (Table 1) define the simulation space and agent interactions.

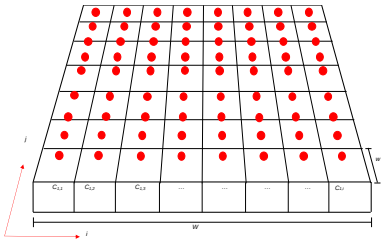


Figure 4. The asphalt pavement cell grid [2]

3.4 Modelling Physical Responses of WMA

Next, the precise physical responses of WMA during the construction process will be obtained, following the workflow shown in Figure 5.

Firstly, a generic physics-based model to accurately simulate the interaction between WMA and the compaction process will be developed. The quality of asphalt pavement is highly dependent on its bulk mechanical properties, which are determined by the contact state of the individual aggregates and the resulting skeleton structure formed between these particles [14]. During the construction process, the compaction force rearranges asphalt mixture aggregates, leading to the development of a denser skeleton structure. Meanwhile, due to the viscoelastic attributes of the asphalt binder, the mechanical properties of asphalt mixtures are also subject to change over different asphalt

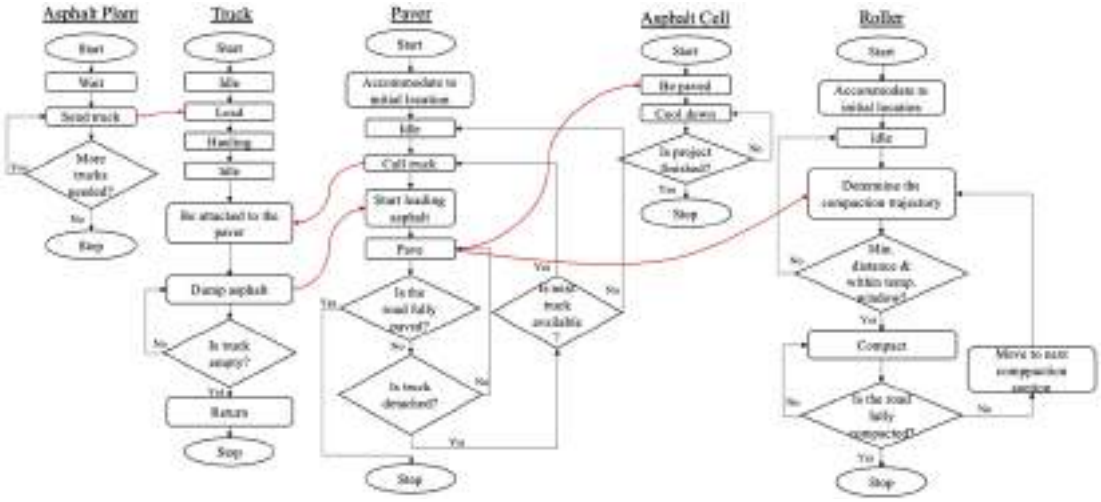


Figure 3. An overview of agents and their states [4]

temperatures. Therefore, this compaction model is expected to be able to capture the WMA's behaviour at the microscopic scale with its granular structure and can be eventually upscaled to derive the macroscopic response of the entire road pavement during compaction.

Overall, the modelling process scales from micro to macro levels. It begins with selecting a contact model to define microscopic interactions, followed by creating mesoscopic models based on WMA mixture designs, calibrated with lab tests at various temperatures. Finally, these models are upscaled to represent WMA pavement, with accuracy tested during field compaction.

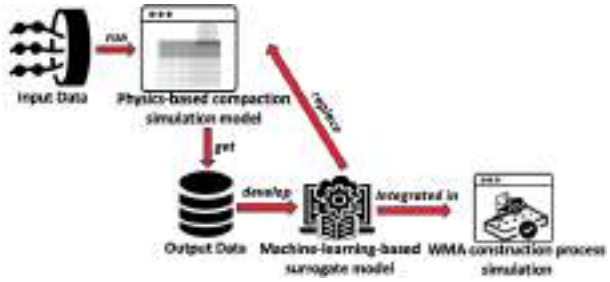


Figure 5. The overview of the WMA construction planning tool development framework

3.5 Surrogate Modelling

The physics-based compaction model will be replaced by a surrogate model to reduce computational costs in analyzing asphalt-compaction interactions. To predict structural responses based on input data, including mixture design, temperature, thickness, density, and applied compaction force.

The surrogate model development will be initiated with the data synthesis to generate a dataset, by performing simulations using the physics-based WMA compaction model under different compaction scenarios (considering different compaction forces from the roller drum and roller speed). In this step, a parametric model will be developed to streamline the data synthesis, and to rapidly construct different compaction scenarios covering the input. Next, based on the established dataset, ML models will be trained and tested, which will be used as the surrogate model to replace the original physics-based compaction model once the former is validated.

3.6 Integrating Surrogate Model into WMA Construction Process Simulation

Lastly, the developed surrogate models for WMA-roller and WMA-environment interactions will be integrated into the WMA construction process simulation model described in Chapter 3.3, to enable a more accurate evaluation of the resultant pavement product quality according to adopted construction strategies. The pavement's mechanical properties can be updated after each roller pass, based on the information including the asphalt temperature and mechanical properties (e.g., density) before compaction, and received compaction energy from the roller.

With the surrogate model, WMA construction simulation can enable automated planning using optimization algorithms like Particle Swarm Optimization (PSO) and Non-Dominated Sorting Genetic Algorithm II (NSGA-II) to generate, assess, and refine strategies for optimal density, productivity, and cost efficiency.

A field test with contractors will validate the planning tool using pre-determined engineering details and weather conditions (Table 1). Optimal construction strategies will be generated where experts from contractors will decide the strategy they will adopt on-site. Two test sections will be prepared: one following the selected optimal strategy and the other using conventional simulation-based planning tools that rely on indirect process quality assessment as a benchmark for quality and efficiency comparison.

4 Preliminary Feasibility Analysis

In this section, a preliminary feasibility study of developing such a planning tool for WMA construction following the framework proposed in Chapter 3 is conducted through a case study. It focuses on demonstrating the feasibility of developing a surrogate model, using the data generated from a physics-based simulation model (which simulates the WMA compaction process on the laboratory scale), to efficiently and accurately capture the temperature-dependent compactability of the target WMA mixture, which is key to precisely map the impact of on-site construction strategies into the resultant pavement quality in the planning tool. Therefore, this preliminary feasibility study concentrates on validating the foundational aspects of the framework and setting stages for future development.

Specifically, five gyratory compaction tests were performed on one type of WMA mixture at five different testing temperatures. Using the gyratory compaction process, the study evaluates the ability of the physics-based model to replicate key structural and thermal responses during compaction. Additionally, it demonstrates the feasibility of calibrating the model using experimental data, ensuring its relevance for future applications. This controlled environment allows for precise analysis and eliminates external complexities, making it an ideal setting for validating the proposed framework's foundational components.

4.1 Gyratory Compaction Test Scheme

Gyratory compaction, a common lab method, simulates field compaction to produce asphalt specimens for mechanical testing. The stabilized mixture in a cylindrical mold undergoes constant vertical pressure

while rotating at a set speed and angle to apply shear forces. Table 3 details the test setup, where compaction stops after 60 gyrations.

Table 3. Gyratory compaction test configurations

Configuration	Value
The internal diameter of the mould	100 mm
Minimum height of the compacted mixture	77 mm
Compaction pressure	600 kPa
Gyration angle	0.82°
Gyration frequency	1/30 Hz
Target number of gyrations	60

The WMA mixture used in the tests was prepared using the design indicated in Table 4. Specifically, this WMA mixture uses a DAT-7 additive, a surface tension reducer, to decrease its production temperature. A total of 5 samples were made. Each sample had a different temperature for performing the compaction test, ranging from 120°C to 40°C with an interval of 20°C (120°C, 100°C, 80°C, 60°C, 40°C).

The testing output is the void ratio of the specimens. The output data was collected after each gyration, to represent the densification process.

Table 4. Gradation of the WMA mixture

Sieving size (mm)	Mass percentage (%)
16 - 11	23.09
11 - 8	3.99
8 - 4	27.33
>2	34.04
Filler	7.14
Asphalt (PEN 40/60)	4.2
Additive (DAT-7)	0.21

4.2 Physics-based Gyratory Compaction Simulation

Following the methodology of Komarairi et al. [10], the gyratory compaction process was simulated using Bullet Physics engine [11]. The engine can effectively detects collisions, calculates contact forces, and updates particle positions [10,11]. Unlike DEM, which simplifies aggregates as spheres, Bullet Physics models real aggregate shapes, improving accuracy while remaining computationally efficient.

Based on the gradation in Table 4, a digital representation of the tested asphalt mixture was created using 10 3D aggregate models obtained via laser scanning (Figure 6). The number of aggregate particles was determined from the asphalt sample mass (1.4 kg) and aggregate density (2686 kg/m³). All particles were scaled to a unit size of 1 mm and adjusted according to gradation to closely match the actual specimen. For computational efficiency, fine aggregates (<2 mm), fillers, and binders were modelled as a mortar layer covering the coarse aggregates, whose thickness is calculated using Equation (1):

$$t = \frac{V_{motar}}{\sum_{i=1}^{n}(SA)_i} \quad (1)$$

where t stands for the thickness of the mortar layer, V_{mortar} refers to the total volume of the mortar, n refers to the total number of coarse aggregates, and $(SA)_i$ refers to the surface area of each coarse aggregate.

These mortar layers also provide viscoelastic contact forces, which are cohesive force in the normal direction between two contacted particles and viscous damping in the tangential direction. Equation (2) specifies the cohesive force:

$$F_c = c\gamma\sqrt{A} \quad (2)$$

where F_c is the cohesive force, c is a proportionality constant, γ is the surface tension of the mortar layer, and A is the contact area between two colliding particles. Together, $c\gamma$ represents the linear stiffness between the two colliding aggregate particles, thus can be replaced by a new parameter k, referring to the contact stiffness.

The viscous force will be simulated using linear and angular velocities [10], as specified in Equation (3) and (4) below:

$$v = v_0\eta_{linear} \quad (3)$$
$$w = w_0\eta_{angular} \quad (4)$$

where v and w stand for the linear and angular velocities of the particle after contact, v_0 and w_0 are the linear and angular velocities of the particle before the contact, and η_{linear} and $\eta_{angular}$ are linear and angular damping factors respectively.



Figure 6. An example of the aggregates

4.3 Calibration and Validation

To test the capability of the physics-based gyratory compaction model in correctly simulating the temperature-dependant compactibility of the target WMA mixture, the compaction model was calibrated using laboratory testing data described in Chapter 4.1. The primary objective is to calibrate the models to determine the corresponding values of k, η_{linear}, and $\eta_{angular}$ at various temperatures. Furthermore, this study assumes that these calibrated parameters will form three temperature-dependent functions for the target WMA mixture through curve fitting. These functions provide a continuous representation of the temperature dependency of each parameter, enabling the derivation of parameter values at any given temperature without the need for additional laboratory tests. Specifically, to validate this assumption, tests conducted at 120°C, 100°C, 60°C, and 40°C were used to calibrate the models to derive corresponding values of k, η_{linear}, and $\eta_{angular}$ and to form the temperature-dependency functions for these parameters. These functions were then used to predict

values at 80°C, where simulation outputs (i.e., the densification curve) were validated against laboratory data.

Conventional calibration of physics-based simulations requires iterative parameter adjustments and repetitive fine-tunings, making it highly time-consuming and computationally expensive. To streamline calibration, the surrogate modelling technique was also applied, where an ML model was developed to replace the physics-based gyratory compaction model. A dataset of 96 instances with varying contact parameters was generated using a full factorial design (Table 5) and simulated to pair input parameters with densification curve outputs. The densification curves were transformed into logarithmic slopes and intercepts, forming inputs for two Random Forest (RF) models: RF-A for slopes and RF-B for intercepts. After training on 80% of the data and testing on the remaining 20%, RF-A achieved an R^2 of 0.9667 and RF-B an R^2 of 0.9871. As for the computational time, using ML-based surrogate models to obtain the densification curve costs 0.003228 seconds, while running each physics-based simulation costs around 7 hours (25200 seconds).

Table 5. Levels of parameters in data generation

Parameter	Levels
Stiffness	1, 100, 1000, 10000, 100000, 1000000
Linear damping factor	0.01, 0.1, 0.5, 0.99
Angular damping factor	0.01, 0.1, 0.5, 0.99

These models facilitated calibration by solving an optimization problem to maximize the R^2 between reconstructed and actual densification curves, as defined by Equation (5):

$$maximize\ R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \quad (5)$$

where y_i is the ith element from the actual curve, $\hat{y}_i$ is the ith element from the reconstructed curve, and $\bar{y}$ is the mean of the actual curve.

RF-A and RF-B were used to predict the values of slope and intercept to reconstruct the (logarithmic) densification curve $\hat{y}$, as shown in Equation (6).

$$\hat{y} = a + b\ln(x) \quad (6)$$

where:

$a = f_{RF-A}(k, \eta_{linear}, \eta_{angular})$

$b = f_{RF-B}(k, \eta_{linear}, \eta_{angular})$

$k \leq 1000000$

$\eta_{linear} \leq 1$

$\eta_{angular} \leq 1$

$k, \eta_{linear}, \eta_{angular} \in Z^+$

Specifically, a and b are the predictions from RF-A and RF-B respectively based on the set of input parameters $k, \eta_{linear}, \eta_{angular}$, and the three input parameters are all positive integers, whose ranges are pre-defined.

This optimization problem was solved by the Genetic Algorithm (GA), to automatically generate, evaluate, evolve, and select optimal sets of input parameters that can fit the actual densifications the best. Each iteration will start with 100 population, where the offsprings will be created with a crossover rate of 0.8 and a mutation rate of 0.2. The GA process will stop after 100 iterations.

Based on the input parameters determined through calibration, three exponential decay functions were determined accordingly, as shown in Figure 7.

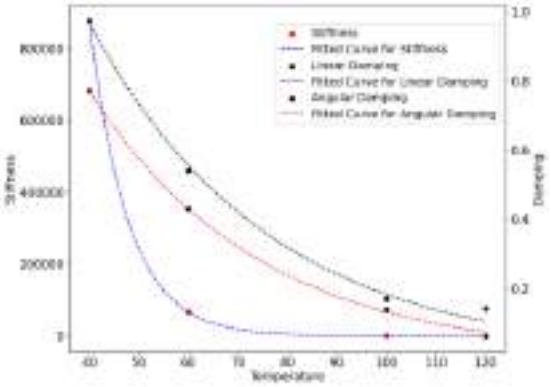


Figure 7. Temperature-dependant decay functions through curve fitting

Values of parameters at 80°C were then determined, resulting in a promising accuracy, with an R^2 score of 0.88 by comparing the obtained densification curve using RF models with the actual laboratory data at this temperature, as shown in Table 6 and Figure 8.

Table 6. Results of the validation

Temperature (°C)	Stiffness (N/mm)	Linear Damping	Angular Damping	R^2
80	5077.99	0.32	0.24	0.88

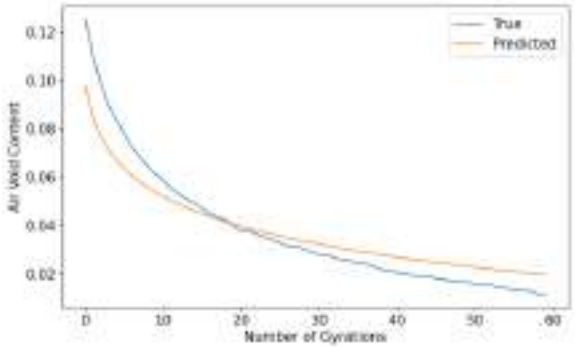


Figure 8. Plots of the validation at 80°C

5 Conclusion

This study proposed a framework for WMA construction planning that incorporates precise physical responses of asphalt, integrating project-specific data

with physics-aware models and simulations. Furthermore, a case study demonstrated the feasibility of developing and calibrating a physics-based simulation model for WMA compaction at the laboratory scale. The model effectively replicated key structural and thermal responses during compaction, capturing granular interactions and material behaviour of the WMA mixture. Calibration using experimental data enabled the derivation of decay functions for contact parameters, which were used to interpolate values across different temperatures. This approach not only ensured a focused validation of the framework but also revealed the model's capability to capture the nuanced relationships between temperature, contact parameters, and compaction dynamics. These findings affirm the foundational premise of the framework, providing a strong basis for extending its application to larger-scale scenarios. However, to validate the surrogate model demonstrated in Chapter 4, only very limited laboratory data was used for the calibration and validation, where it will be more beneficial to include more data and more mixture types to enhance its generalizability.

Additionally, while the integration of a surrogate model is part of the broader research plan, its validation is beyond the scope of this preliminary study and will be pursued in future work. Although real-world adoption of the tool remains challenging due to construction uncertainties and contractor risk aversion, the planning tool can still provide contractors with valuable early-stage planning insights, where the on-site uncertainty can be further overcome by integrating the planning tool with real-time construction data (e.g., pavement temperature, weather conditions, machinery location, etc.), enabling automated adjustments to compaction strategies based on sensor feedback. Ultimately, this would support full automation in WMA construction, enhancing efficiency, consistency, and sustainability in road projects.

6 Reference

[1] G. Cheraghian, A. Cannone Falchetto, Z. You, S. Chen, Y.S. Kim, J. Westerhoff, K.H. Moon, M.P. Wistuba, Warm mix asphalt technology: An up to date review, J Clean Prod 268 (2020) 122128.

[2] D. Makarov, F. Vahdatikhaki, S. Miller, S. Mowlaei, A. Dorée, Real-time trajectory planning for asphalt compaction operator support, Autom Constr 155 (2023) 105034.

[3] C. Hendrickson, T. Au, Project Management for Construction, Second Edition, 2000. https://www.cmu.edu/cee/projects/PMbook/ (accessed October 30, 2023).

[4] A.R.R. Dalence, F. Vahdatikhaki, S. Miller, A.G. Dorée, Integrating VR and Simulation for Enhanced Planning of Asphalt Compaction, Proceedings of the International Symposium on Automation and Robotics in Construction 2022-July (2022) 55–62.

[5] C.G. Arbeider, S.R. Miller, A. Doree, M. Oosterveld, Planning the asphalt construction process: Towards more consistent paving and compaction operations, in: 17th AAPA International Flexible Pavements Conference Exhibition 2017: Roads of the Future, Embracing Disruption, 2017. https://research.utwente.nl/en/publications/planning-the-asphalt-construction-process-towards-more-consistent (accessed October 1, 2023).

[6] J. Kim, D. eun Lee, J. Seo, Task planning strategy and path similarity analysis for an autonomous excavator, Autom Constr 112 (2020) 103108.

[7] S. Park, J. Kim, K. Cho, J. Seo, Integrating Task Component Design in Excavator–Truck Operation Planning: Enhancing Automation Systems through a Web-Based System, Applied Sciences 2024, Vol. 14, Page 6052 14 (2024) 6052.

[8] P. Liu, H. Xu, D. Wang, C. Wang, C. Schulze, M. Oeser, Comparison of mechanical responses of asphalt mixtures manufactured by different compaction methods, Constr Build Mater 162 (2018) 765–780.

[9] L. Al Khateeb, K. Anupam, S. Erkens, T. Scarpas, Micromechanical simulation of porous asphalt mixture compaction using discrete element method (DEM), Constr Build Mater 301 (2021) 124305.

[10] S. Komaragiri, A. Gigliotti, A. Bhasin, Calibration and extended validation of a virtual asphalt mixture compaction model using bullet physics engine, Constr Build Mater 311 (2021) 125257.

[11] E. Coumans, Y. Bai, PyBullet, a Python module for physics simulation for games, robotics and machine learning, (2021).

[12] Q. Shen, Coupling the Road Construction Process Quality Indicators into Product Quality Indicators, EngD Thesis, University of Twente, 2023.

[13] J. Chen, X. Gao, Y. Hu, Z. Zeng, Y. Liu, A meta-model-based optimization approach for fast and reliable calibration of building energy models, Energy 188 (2019) 116046.

[14] F. Chen, N. Wang, T. Ma, Experimental and numerical simulation study on flow behavior of asphalt mixture using smart aggregate, Smart Mater Struct 32 (2023) 075022.

Integrating Spatial Proximity and Visual Feature Similarity for Crew Group Detection in Construction Site

Cheng-Yun Tsai[1], Jacob J. Lin[1], Ci-Jyun Liang[2]

[1]Department of Civil Engineering, National Taiwan University, Taiwan
[2]Department of Civil Engineering, Stony Brook University, USA

r12521709@ntu.edu.tw, jacoblin@ntu.edu.tw, Ci-Jyun.Liang@stonybrook.edu

Abstract -

Crew-level productivity analysis plays a crucial role in construction site management, as it provides a macro-level understanding of workforce performance. While traditional productivity measurement methods, such as work sampling, group timing technique, and five-minute rating, offer valuable insights, they rely heavily on manual observation, making them labor-intensive and prone to inaccuracy. In this study, we propose a deep learning-based framework for automated crew-level identification. The framework employs a graph-based approach that integrates visual feature similarity and spatial proximity of workers, combined with clustering algorithms, to detect and analyze worker groups. The proposed method is validated on a construction site dataset collected from a rebar installation task. Experimental results demonstrate the framework's effectiveness, achieving high accuracy in group detection with robust performance across various evaluation metrics. This work highlights the potential of automated systems to enhance construction site management by reducing reliance on manual observation and providing real-time insights into crew-level productivity.

Keywords -

Group Detection; Clustering; Image Understanding; Human Activity Recognition

1 Introduction

Labor resources play a critical role in the success of construction projects [1]. However, construction sites typically consist of distinct groups of workers collaborating to accomplish tasks. From a macro perspective, analyzing productivity at the crew level—rather than focusing on the productivity of individual workers—provides a more comprehensive view of progress tracking and performance analysis [2]. By incorporating additional contextual information such as objects or spatial locations, this analysis can help managers identify other potential causes of delays, such as inefficient material placement or work environment constraints. Therefore, understanding and monitoring crew-level productivity is essential for improving construction site management.

Traditionally, three primary methods have been used for productivity measurement on construction sites: work sampling, group timing technique, and five-minute rating [3]. Work sampling involves observing workers at random intervals to determine the percentage of time spent on productive and non-productive activities, providing insights into overall labor efficiency. The group timing technique, on the other hand, focuses on measuring crew cycle times by observing work processes at fixed intervals, which allows managers to assess the efficiency of group operations and evaluate the appropriateness of crew size. Finally, the five-minute rating is a simpler and faster approach that estimates delay times by periodically observing crew activities over short durations, such as five minutes. While these methods are effective for gathering productivity data, they rely heavily on manual observation, which is labor-intensive, time-consuming, and dependent on the experience and judgment of observers.

Recent advancements in technology, particularly the widespread deployment of cameras on construction sites and developments in deep learning and computer vision, have made automated productivity monitoring increasingly feasible. These automated systems address the limitations of traditional methods by reducing the need for labor-intensive and time-consuming manual observation. Unlike methods that rely heavily on the observer's experience and continuous monitoring, automated systems offer real-time and accurate productivity tracking. This technological shift enables site managers to focus their efforts on more strategic aspects of project management while ensuring the reliability and efficiency of data collection.

In recent studies, significant progress has been made in motion recognition for individual workers. Luo et al. [4] propose an improved CNN integrating RGB, optical flow, and gray stream data to accurately recognize and assess a single worker's reinforcement installation activity in videos. Luo et al. [5] used spatial and temporal convolutional networks with a novel fusion strategy to recognize multi-workers' actions from surveillance videos. However, these studies primarily focus on the productivity of individual workers and lack an in-depth discussion of

crew-level performance.

A limited number of studies have explored the relationships between workers and workers or workers and objects. The same author from [5] proposes a two-step method, which uses CNN to detect objects and workers and then analyzes semantic and spatial relevance between each entity [6]. Cai et al. [7] introduce a two-step classification method combining working group identification and activity recognition using LSTM networks, leveraging pre-computed positional cues (eq. bounding box overlapping) and attentional cues (eq. head pose, body orientation) to recognize construction activities with multiple entities. Although these studies successfully identify the relationships between objects in construction site images, their methods for worker group identification remain constrained by precomputed handcrafted features or site-specific rules established by domain experts. Therefore, developing an effective method for identifying and analyzing worker groups on construction sites remains a significant research challenge.

Building on these insights, this study contributes to the field by proposing a deep learning-based framework for automated crew-level identification. Unlike previous methods that rely on handcrafted features or predefined spatial heuristics, our approach dynamically integrates visual feature similarity and spatial proximity into a unified relation matrix. This enables the model to adaptively cluster workers based on real-world interactions rather than being constrained by rigid rules or single-modality input. Additionally, this method simplifies the system's deployment requirements, as it only requires image input and a person detection model, eliminating the need for manually predefined spatial constraints or external tracking devices. The framework is validated using a dataset collected from a rebar installation task on a construction site, demonstrating its robustness across diverse scenarios.

2 Methodology

In this study, our objective is to cluster workers in the scene into different groups; we combine both spatial relevance and feature relevance to represent the affinity matrix between the workers in the scene. Figure 1 shows the overview of our framework, which contains four major modules: (a) Feature Extraction, (b) Graph Construction, (c) Spatial Distance Encoding with Masking, and (d) Relational Matrix for Clustering. We will elaborate on the details of these four modules in the following section.

2.1 Feature Extraction

The purpose of this module is to extract the great information of visual feature reliance in the image for each worker. As the main goal of this research is to cluster the workers, we assume that there will be other modules to complete the object detection of workers. Therefore, the inputs of the proposed framework are the RGB image with each worker's bounding box. We first utilize Inception-v3 [8], a CNN-based backbone model combined with ROI-Align [9], to extract spatially precise feature representations for each entity.

In this study, we use Inception-v3 [8] as the feature extraction backbone. To enhance robustness and leverage prior knowledge, we initialize the model with pretrained weights from the Collective Activity Dataset [10], which has been optimized for human-centric activity recognition. This choice allows us to extract meaningful features while mitigating the risk of overfitting. Given the focus of our study on clustering methods rather than CNN architecture comparisons, we selected a well-established backbone known for its balance between efficiency and accuracy in feature extraction.

ROI-Align [9] is a region-based operation designed to extract features corresponding to a specific region of interest (ROI) from the feature maps produced by the CNN backbone. Unlike earlier methods, such as ROI-Pooling, ROI-Align avoids quantization errors by using bilinear interpolation to precisely map the input pixels to the feature space.

2.2 Graph Construction

After extracting the feature from the image, we want to calculate the affinity further according to the similarity of visual features. We can represent the individuals in any scene as a graph, where each node represents the individuals and the edge represents the connection or the relationship between these two nodes. By leveraging this graph structure, we aim to compute the visual feature affinity matrix R_E, which quantifies the affinity between features.

The graph can be denoted as $G = (V, E)$, where V represent the vertices and E represent the edges, and since we are not sure of the exact relationship in the scene in present, there might be any possible pairs of relationships between each worker, so we construct a fully connected graph; the number of vertices N is equal to the number of workers in the scene, and the number of edges is equal to $\binom{N}{2}$. To represent the relationship $e_{u,v}$ between two entities u and v, we adopt the approach introduced in [11]:

$$e_{u,v} = F_1(f_u) \cdot (F_2(f_v))^T, \quad u, v \in \mathcal{V}, \tag{1}$$

where f_u and f_v are the feature vectors of entities u and v, respectively. F_1 and F_2 are learnable linear transformations that project the feature vectors into a relational feature space. The resulting value $e_{u,v}$ represents the connection strength or similarity between u and v. The complete **feature affinity matrix $\mathbf{R_E} \in \mathbb{R}^{N \times N}$**, where N is the number

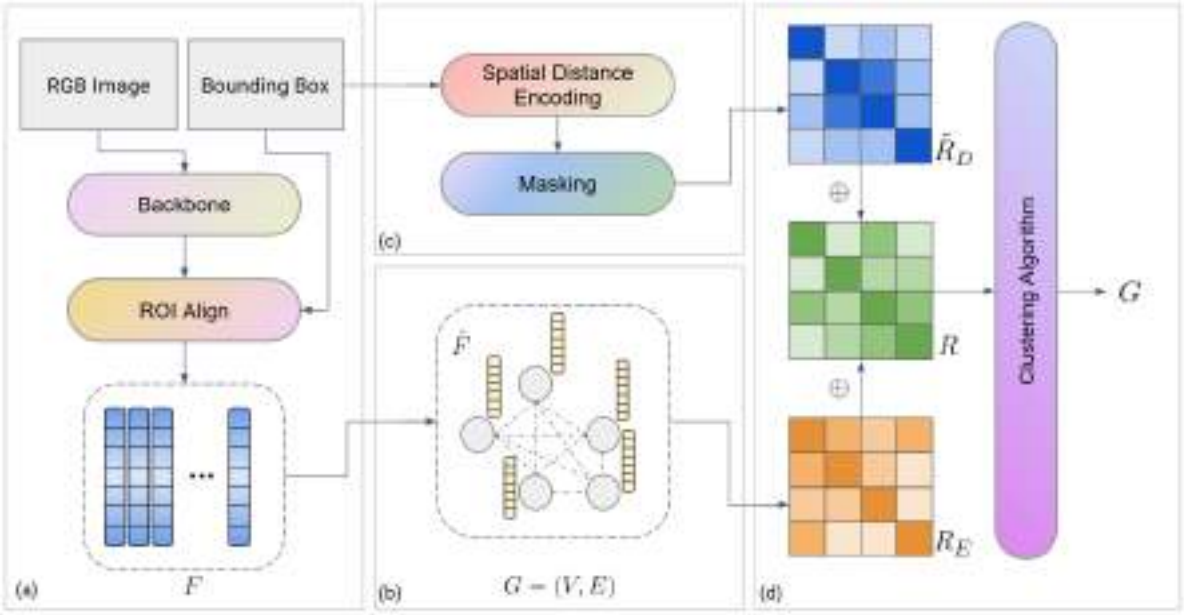


Figure 1. Overview of the Proposed Framework: Key Components – (a) Feature Extraction: extracts high-level features from RGB images and corresponding bounding boxes. (b) Graph Construction: builds a graph structure to represent the extracted features and their relationships, where nodes correspond to features and edges indicate potential connections. (c) Spatial Distance Encoding with Masking: encodes spatial distances and applies masking strategies to enhance the relevance of nearby entities while filtering out irrelevant ones. (d) Relational Matrix for Clustering: computes a relational matrix to facilitate feature grouping and clustering based on the encoded relationships.

of nodes, is constructed as:

$$\mathbf{R_E} = \left[e_{u,v}\right]_{u,v\in\mathcal{V}}. \tag{2}$$

To ensure the weights are normalized, we apply a **softmax operation** to each row of the affinity matrix $\mathbf{R_E}$. This normalization ensures that the sum of the affinity weights for each node (row) equals 1. Mathematically, this process is expressed as:

$$e_{u,v} \leftarrow \frac{\exp(e_{u,v})}{\sum_{v'\in\mathcal{V}} \exp(e_{u,v'})}. \tag{3}$$

This softmax operation converts raw edge weights into probabilities.

2.3 Spatial Distance Encoding & Masking

After we have the feature affinity matrix R_E, We also want to include the relationship between workers in space as a reference. The spatial affinity matrix R_D is computed to encode the physical proximity between entities using their Euclidean distance $D(u, v)$. It is defined as:

$$R_D(u, v) = \exp\left(-\frac{D(u, v)}{\alpha}\right), \tag{4}$$

where:

$$D(u, v) = \sqrt{(x_u - x_v)^2 + (y_u - y_v)^2}. \tag{5}$$

In these equations, (x_u, y_u) and (x_v, y_v) denote the spatial coordinates of entities u and v, typically represented as the midpoint of the bottom edge of their bounding boxes, $\alpha > 0$ is a scaling parameter that determines the sensitivity of the affinity score to the distance.

The inverse exponential function R_D ensures a smooth mapping of distances to affinity scores, where entities that are spatially close (smaller $D(u, v)$) receive higher affinity values, indicating stronger proximity-based interactions, while entities that are farther apart (larger $D(u, v)$) are assigned smaller affinity values, reflecting weaker spatial relationships.

We further apply a **masking technique** to filter out weak spatial relationships. Specifically, for each element in the spatial affinity matrix R_D, if the Euclidean distance value is smaller than a predefined threshold τ, the corresponding entry is set to zero. Mathematically, the masked spatial affinity matrix $\tilde{R}_D$ is given by:

$$\tilde{R}_D(u, v) = \begin{cases} R_D(u, v), & \text{if } D(u, v) \geq \tau, \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

Here, τ is a threshold value that controls the sparsity of the affinity matrix.

2.4 Relation Matrix

To quantify relationships between entities, we define a relation matrix R, which integrates both feature-based and spatial relationships to provide a comprehensive representation of interactions. This matrix is composed of the feature affinity matrix R_E and the masked spatial affinity matrix $\tilde{R}_D$, which represents the spatial relationships between entities based on their positions.

Finally, the combined relation matrix R integrates R_E and $\tilde{R}_D$ to capture both visual feature similarity and spatial proximity. It is calculated as:

$$R = \lambda R_E \oplus (1 - \lambda)\tilde{R}_D, \tag{7}$$

where $\lambda \in [0, 1]$ is a weighting coefficient controlling the relative importance of R_E and R_D, and $\oplus$ denotes element-wise addition of the two matrices. With the relation matrix, we can get the group division results through a post-processing method, e.g., a clustering algorithm.

To achieve robust group clustering, we employ the Spectral Clustering algorithm [12], which effectively operates on the constructed relation matrix R. Unlike traditional K-Means, which assumes spherical clusters and requires predefining the number of clusters, Spectral Clustering leverages the graph Laplacian eigenvectors to segment data based on intrinsic connectivity, making it more suitable for identifying worker groups in complex construction site scenarios. Additionally, given our method's graph-based nature, Spectral Clustering provides a more effective approach to clustering workers based on both feature and spatial affinities.

2.5 Loss Function

The loss function is designed to supervises the accuracy of group detection by evaluating the predicted relationship matrix R_{pred} against the ground-truth relationship matrix R_{gt}. Both matrices are of size $N \times N$, where N is the total number of entities. Each element $R_{i,j}$ in the matrix represents the likelihood of a relationship between entities i and j.

The loss is formulated using cross-entropy, applied element-wise across the matrices:

$$\mathcal{L}_{\text{det}} = -\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(R_{\text{gt}}^{(i,j)}\log R_{\text{pred}}^{(i,j)} + \left(1 - R_{\text{gt}}^{(i,j)}\right)\log\left(1 - R_{\text{pred}}^{(i,j)}\right)\right). \tag{8}$$

Table 1. Average and Total Statistics for Frames and Entities

	Avg	Total
Frame	50	608
Entity	5.28	3211

where $R_{\text{gt}}^{(i,j)}$ is the ground-truth value for the relationship between entities i and j (binary: 1 if related, 0 otherwise).$R_{\text{pred}}^{(i,j)}$ is the predicted probability of the relationship between entities i and j.

3 Experiment

In this section, we elaborate on the details of our proposed dataset, the metrics used to evaluate the performance of our framework, and the results of our experiments. Additionally, we provide an in-depth discussion of the findings.

3.1 Dataset

The dataset consists of 13 videos, capturing real-world construction site activities. The videos were recorded under both sunny and cloudy conditions, ensuring coverage of different natural lighting environments. While the camera angles were not entirely consistent across all videos, the majority were captured from a top-down perspective, typically from elevated positions overlooking the construction site. To evaluate the proposed framework in general construction site, we chose to record the video during the rebar installation work item since there are a lot of workers passing between the scene and different groups of workers working simultaneously. Table 1 shows the detailed statistics of the dataset. On average, each video contains 5.2 workers per frame, reflecting realistic worker density variations. The videos were then set to a resolution of 1920×1080, and downsampled to 1fps.

For group annotation, A group of workers who interact with each other or are working on the same object will be considered a group. Figure 2 shows an annotation sample, bounding boxes of the same color in the picture with red lines between them represent that they belong to the same group.

3.2 Metrics

In this section, we discuss how to evaluate the performence of our framework; the task that our framework does can be classified as **Group Detection Task**. Following [11], we use the **Helf Matrix** [13] to evaluate the accuracy of detected groups. The Helf Matrix quantifies the overlap between a ground truth group and a discovered group using the following formula:



Figure 2. Annotation sample

$$\text{Helf Ratio} = \frac{\text{Number of individuals in the intersection}}{\text{Number of individuals in the union}},$$

the *intersection* refers to the individuals present in both the ground truth and the discovered group, while the *union* accounts for all individuals that appear in either group. If the Helf Ratio exceeds 0.5, the detected group is considered a *true detection*. By this, we can compute the *precision*, *recall*, and *F1 score*.

We also adopted the extended metrics proposed by [11]. Specifically, we evaluate the accuracy at varying thresholds by increasing the criteria from 0.5 to 1 in steps of 0.1, enabling us to plot the accuracy curve and compute the AUC (Area Under the Curve) score, referred to as IoU@AUC. Additionally, we compute the matrix IoU score (Mat.IoU) using the predicted and ground-truth binary human group relation matrices, where element-wise logical AND and OR operations are applied to measure the overlap between groups.

3.3 Network Implementation details

We used Incepction-v3[8] as a frozen backbone pretrained on Collective Activity Dataset [10], and utilize ROI-Align [9] with crop size of 5×5 to extracte features. The parameter λ in Eq. (7) is set to 0.3. The parameter τ in Eq. (6) is set to 0.7 times the width of the input image.

3.4 Training & Inference

In this study, all experiments were conducted on a workstation equipped with an AMD Ryzen 9 5950X CPU and NVIDIA RTX 3090 GPU. The workstation runs on the Ubuntu 18.04 LTS operating system. All networks in this research were implemented under the PyTorch framework. We trained the network using stochastic gradient descent (SGD) with a learning rate initially with 2×10^{-5} and decreasing to 1×10^{-5} at epoch 20. We trained the model for total 30 epochs.

3.5 Result & Discussion

This section presents the results of our proposed framework and discusses the findings. From Table 2, it can be observed that the **Precision**, **Recall**, and **F1 Score** are 64.4%, 82.7%, and 72.2%, respectively. These results indicate that the model performs well in minimizing false detections (high Precision) and capturing most of the true groups (high Recall). However, the significant difference between Recall and Precision suggests that while the model excels in identifying a wide range of groups, it may occasionally misclassify non-groups, leading to false positives. This trade-off highlights the need for further refinement to achieve better balance in precision and recall without compromising detection accuracy.

Table 2. Precision, Recall, and F1 Score (%)

Precision	Recall	F1 Score
64.4	82.7	72.2

Similarly, Table 3 provides insight into the spatial alignment of detected groups, with **IoU@0.5**, **IoU@AUC**, and **Mat.IoU** achieving 72.2%, 33.5%, and 72.3%, respectively. The high **IoU@0.5** and **Mat.IoU** values indicate that the model effectively aligns detected groups with ground truth annotations, demonstrating strong spatial consistency. However, the relatively lower **IoU@AUC** suggests that the model's performance varies across different thresholds, possibly due to challenges in handling edge cases or overlapping groups. This result underscores the importance of refining the model to improve its robustness in more complex scenarios, particularly under varying spatial conditions or occlusions.

Table 3. IoU Metrics (%)

IoU@0.5	IoU@AUC	Mat.IoU
72.2	33.5	72.3

To further analyze the results, we conducted a sensitivity analysis on the visual feature similarity R_e and spatial proximity $\tilde{R}_d$. Table 4 demonstrates that spatial proximity $\tilde{R}_d$ plays a more critical role in group detection performance. As the weight of spatial proximity increases, the performance improves, achieving the highest IoU@0.5 and IoU@AUC values of **72.2%** and **33.5%** when $R_e = 0.3$ and $\tilde{R}_d = 0.7$. However, when the model relies solely on spatial proximity ($R_e = 0$), the performance declines slightly, indicating that visual feature similarity R_e still contributes to learning individual relationships effectively. This suggests that while spatial information dominates, incorporating visual feature similarity helps the model achieve a more balanced understanding of group relationships.

Figure 3 visualizes the correct predictions. In Prediction

Table 4. Sensitivity analysis on visual feature similarity R_e and spatial proximity $\tilde{R}_d$. The highest scores are marked in **bold** and the second highest in underlined.

R_e	$\tilde{R}_d$	Group Detection		
		IoU@0.5	IoU@AUC	Mat.IoU
1		58.5	29.4	**74.1**
0.7	0.3	61.2	29.2	71.7
0.5	0.5	63.5	30.6	71.0
0.3	0.7	**72.2**	**33.5**	72.3
	1	70.6	31.5	69.7

1 (Figure 3b), the framework demonstrates its capability to effectively distinguish multiple groups within a complex scene, even when numerous workers are present and interacting across different locations. This highlights the model's ability to accurately separate groups in dense and dynamic environments. On the other hand, the result for Prediction 2 (Figure 3d) emphasizes the model's ability to correctly identify the true group despite challenges posed by spatial proximity. Specifically, although the two individuals in the upper-right corner appear close to each other, the model successfully leverages visual feature similarity to learn that the actual group consists of the two workers actively engaged in rebar installation. This underscores the model's robustness in learning semantic group relations based on contextual and task-related cues, rather than relying solely on physical distance.

However, Figure 4 highlights the incorrect predictions. The primary source of error in the upper-right corner stems from the intertwining and occlusion of two groups, making it difficult for the model to differentiate between them. Additionally, since the current model takes a single image as input, understanding group dynamics and movement becomes particularly challenging under such conditions. This limitation suggests that without temporal information, the model struggles to resolve ambiguities caused by overlapping worker interactions and visual obstructions. Beyond occlusion, we identified additional error sources: (1) Dynamic role changes, where workers temporarily move between groups, causing assignment inconsistencies; (2) Varying worker proximity, where workers working near each other but not collaboratively are mistakenly clustered together; and (3) Clothing and visual appearance similarities, where workers wearing identical uniforms and safety gear are misclassified into the same group due to similar feature embeddings. These findings suggest that integrating temporal tracking and additional contextual cues could enhance robustness in future iterations of the model.

4 Conclusions

In this study, we introduced a deep learning-based framework for automated crew-level identification on construction sites. The proposed framework leverages a graph-based approach that combines visual feature similarity and spatial proximity, enabling robust detection of worker groups through clustering algorithms. By integrating these relationships into a relational matrix and applying spectral clustering, the method effectively identifies groups of workers collaborating on shared tasks.

The framework was validated on a construction site dataset collected during a rebar installation task. The experimental results showed that our method achieved high performance in group detection, with metrics such as IoU@0.5, IoU@AUC, and Mat.IoU reaching 72.2%, 33.5%, and 72.3%, respectively. Sensitivity analysis further revealed the significance of spatial proximity in improving detection accuracy, while visual feature similarity provided additional contextual support for identifying group relationships.

Despite the promising results, certain limitations were observed, particularly in scenarios involving heavy occlusion or overlapping worker interactions. These challenges highlight the need for incorporating temporal information to better capture group dynamics and movement across sequences of frames. While our dataset captures key worker interaction patterns and spatial relationships, we acknowledge that differences in construction site layouts, worker densities, and environmental conditions may influence the model's performance.

For practical deployment, cameras would be installed on nearby buildings or tower cranes to capture footage of the construction site. The recorded videos would then be transmitted to an office workstation, where the proposed system processes the data to detect and analyze worker groups. This setup enables real-time monitoring without interfering with ongoing construction activities. While the initial setup requires hardware installation and network configuration, the system significantly reduces manual observation efforts and improves productivity tracking, making it a practical and scalable solution for construction management.

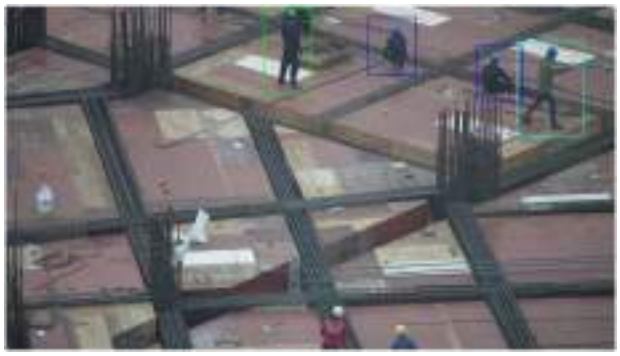
Future improvements could focus on incorporating temporal analysis to enhance worker grouping consistency across frames. One approach is to implement trajectory-based tracking algorithms to maintain worker identities and analyze movement patterns over time. This would help differentiate temporary interactions from persistent group formations. Additionally, sequential modeling techniques such as RNNs or transformer-based video models could be explored to capture temporal dependencies, further refining worker clustering in dynamic construction environments.



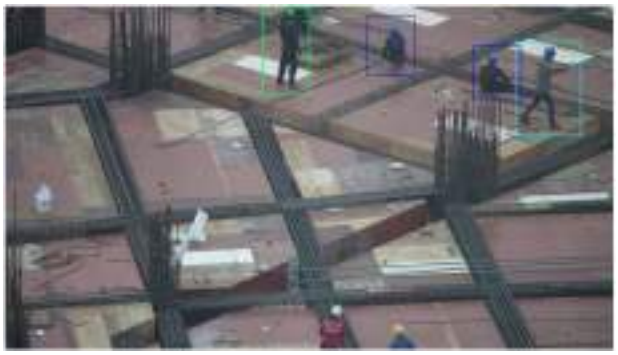
(a) GroundTruth 1



(b) Prediction 1



(c) GroundTruth 2



(d) Prediction 2

Figure 3. Visualization of Correct Predictions: The detected groups (b, d) match the ground truth groups (a, c).



(a) Groundtruth



(b) Prediction

Figure 4. Visualization of Incorrect Predictions: The detected groups (b) do not match the ground truth groups (a).

References

[1] Behnam Sherafat, Changbum R Ahn, Reza Akhavian, Amir H Behzadan, Mani Golparvar-Fard, Hyunsoo Kim, Yong-Cheol Lee, Abbas Rashidi, and Ehsan Rezazadeh Azar. Automated methods for activity recognition of construction workers and equipment: State-of-the-art review. *Journal of Construction Engineering and Management*, 146(6): 03120002, 2020.

[2] Jie Gong and Carlos H Caldas. An object recognition, tracking, and contextual reasoning-based video interpretation method for rapid productivity analysis of construction operations. *Automation in Construction*, 20(8):1211–1226, 2011.

[3] H Randolph Thomas and Jeffrey Daily. Crew performance measurement via activity sampling. *Journal of construction engineering and management*, 109 (3):309–320, 1983.

[4] Hanbin Luo, Chaohua Xiong, Weili Fang, Peter ED Love, Bowen Zhang, and Xi Ouyang. Convolutional neural networks: Computer vision-based workforce activity assessment in construction. *Automation in Construction*, 94:282–289, 2018.

[5] Xiaochun Luo, Heng Li, Dongping Cao, Yantao Yu, Xincong Yang, and Ting Huang. Towards efficient and objective work sampling: Recognizing workers' activities in site surveillance videos with two-stream convolutional networks. *Automation in Construction*, 94:360–370, 2018.

[6] Xiaochun Luo, Heng Li, Dongping Cao, Fei Dai, JoonOh Seo, SangHyun Lee, et al. Recognizing diverse construction activities in site images via relevance networks of construction-related objects detected by convolutional neural networks. *J. Comput. Civ. Eng*, 32(3):04018012, 2018.

[7] Jiannan Cai, Yuxi Zhang, and Hubo Cai. Two-step long short-term memory method for identifying construction activities through positional and attentional cues. *Automation in Construction*, 106: 102886, 2019.

[8] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

[9] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

[10] Wongun Choi, Khuram Shahid, and Silvio Savarese. What are they doing?: Collective activity classification using spatio-temporal relationship among people. In *2009 IEEE 12th international conference on computer vision workshops, ICCV Workshops*, pages 1282–1289. IEEE, 2009.

[11] Ruize Han, Haomin Yan, Jiacheng Li, Songmiao Wang, Wei Feng, and Song Wang. Panoramic human activity recognition. In *European Conference on Computer Vision*, pages 244–261. Springer, 2022.

[12] Lihi Zelnik-Manor and Pietro Perona. Self-tuning spectral clustering. *Advances in neural information processing systems*, 17, 2004.

[13] Wongun Choi, Yu-Wei Chao, Caroline Pantofaru, and Silvio Savarese. Discovering groups of people in images. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13*, pages 417–433. Springer, 2014.

Automated segmentation and analysis of point clouds of pier foundations using Pier Inspection and Evaluation Report (PIER)

Amine Turki[1], Nolan W. Hayes[1], Balaji Selvakumar[1], Bryan P. Maldonado[1], and Diana Hun[1]

[1]Buildings and Transportation Science Division, Oak Ridge National Laboratory, United States of America *

turkia@ornl.gov, hayesnw@ornl.gov, selvakumarb@ornl.gov, maldonadopbp@ornl.gov, hunde@ornl.gov

Abstract -

Pier foundations are commonly used in locations with unstable soil or where other types of foundations are unsuitable or cost prohibitive. A pier foundation consists of vertical columns to support the structure and elevate it above the ground. Common materials for pier foundations include masonry, concrete, timber, and steel. The methods for accurate placement of pier foundations have remained relatively unchanged for decades. For simple installations, construction chalk lines are used to layout the locations of piers to ensure accurate placement and elevation. For more complex installations, surveying instruments operated by trained professionals are employed to accurately locate piers and assess correct elevation before construction. After installation, another survey may need to be performed to assess the quality of the as-built foundation. However, with the advent of terrestrial laser scanners (TLS), the means now exist for contractors to conduct their own assessments of as-built foundations. The major barrier preventing contractors from performing their own assessments of as-built foundation quality is the segmentation and analysis of point cloud data, a skill that often requires a trained user. The objective of this research is to develop a software tool (PIER: Pier Inspection and Evaluation Report) to enable automated segmentation and analysis of point clouds of pier foundations. In this paper, the automated segmentation and analysis algorithms are detailed. A mockup lay out of pier foundations was built using concrete masonry units, and the algorithms were tested to evaluate performance. Limitations of the current algorithms and future research direction are discussed.

Keywords -

foundation, pier, evaluation, automation, real-time

*Notice: This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (https://www.energy.gov/doe-public-access-plan).

1 Introduction

In the United States, pier foundations are often used to support a variety of building types and infrastructure. Common applications for piers include foundations for manufactured housing, site-built housing with crawlspaces, buildings constructed on unstable soil, and bridges, especially those over water. Pier foundations are unique in that they are discontinuously distributed across the footprint of the structure. As a result, the accurate placement of pier locations and elevations is critical.

Pier foundations can take the form of many shapes and sizes. An array of materials are traditionally used to construct piers (Figure 1). For example, timber, steel, cast-in-place concrete, precast concrete, and masonry are all typically used for pier foundations across diverse building types. Due to variability in materials, sizes, and shapes, methods of placement and installation can vary significantly.

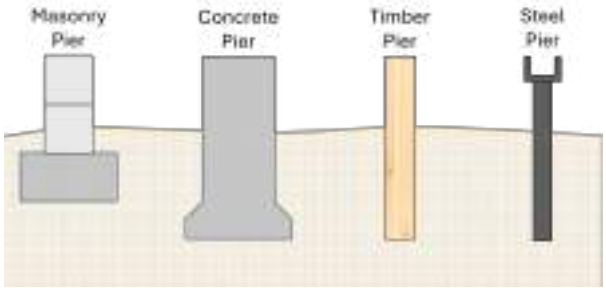


Figure 1. Types of pier foundations.

1.1 Pier foundation positioning and elevation control

The methods for accurate placement of pier foundations have remained relatively the same for decades [1, 2, 3]. For small construction sites, the layout of foundations is made using strings, chalk lines, measuring tapes, or bubble levels. Some combination of these tools is necessary to ensure squarely-set, evenly-spaced pier foundations with accurate top elevations. However, as the size of the construction site increases, the usability of these tools diminishes, and the

likelihood of user error increases. More advanced tools such as laser levels can be employed with a higher degree of accuracy to ensure that the top of each pier foundation is constructed at the correct elevation. For larger construction projects, it is common to hire a professional surveyor to measure and locate foundations prior to construction. After the construction of the foundations, it is sometimes necessary to perform another survey to determine if the as-built elevations and locations are within tolerance. Recent research has developed methods to analyze piers in challenging locations, such as underwater [4]; however these methods are not typically applicable under normal conditions in open air.

Because these tools rely on skilled labor to assess distances and read measurements, there is a chance for errors. Accurate placement of piers is critical because the structure above must be adequately supported at the correct locations. As a result, it is common to incorporate some tolerance for pier foundations so that if the placement of a pier is inaccurate to some degree, the subsequent construction is minimally affected. However, excessive errors can cause constructability issues, which can delay the entire construction schedule. For example, in a series of piers supporting an array of beams, if a pier is misplaced by more than the width of a beam, either the pier must be demolished and correctly reconstructed or the layout of the beam structure must be altered to accommodate the misplaced pier. Errors can often go unnoticed during the foundation construction process, causing significant delays in the construction timeline.

94% of construction firms report having difficulty finding workers to hire [5]. Current quality assessments of pier foundations for complex construction sites are conducted by professional surveyors, who are becoming harder to find and more expensive. The transfer of knowledge and continued development of skilled labor is necessary to ensure that construction demand can meet the growing needs of society. Digital tools can increase productivity and alleviate some of the strain on the diminishing workforce. However, for foundation layout and construction, the availability of digital tools to improve productivity is low. Some options exist on the market to automate some aspects of foundation evaluation; however, no option currently exists specifically to evaluate the quality of construction for pier foundations.

1.2 Laser scanning and point cloud segmentation

Modern laser scanning devices can collect more than 2 million points per second and are accurate at long ranges with peak range accuracy of 2 mm + 2 ppm or better [6]. These simple, easy-to-use scanners have the potential to expedite the measurement of as-built quality of construction and generate real-time guidance to correct issues. With further development, the affordability and accessibility of these devices will also add to their potential. The accompanying software with modern laser scanners can automate and simplify the process of registering multiple scans into a single frame, which is usually a time-consuming and difficult process. Laser scanners have the potential to replace manual measurement methods and tools used during the traditional process of construction if point cloud data and analysis can be streamlined and automated.

The major aspect of laser scanning that currently limits its adoption is the segmentation of point cloud data. The fastest method of scanning for most TLS is typically a full 360° scan during which every object in line of sight is measured. Once the scan is complete, the data is imported into software and the points belonging to the objects of interest (the piers in this case) are retrieved from the rest of the points. This process is called segmentation. Automated segmentation of points clouds for varying applications is a major point of interest in current research [7, 8, 9, 10]. Automated point clouds segmentation techniques exist to perform tasks ranging from prefabricated tolerance compliance assessment [11] to automated digital twin generation [12]. The segmentation process can be turned from a several-hour manual process into a few second-long automated process by incorporating automation through artificial intelligence. Previous research by the authors developed automated segmentation and analysis algorithms for the flatness and levelness of concrete slabs [6]. The current research expands upon the existing algorithms to automate segmentation and analysis of pier foundations. This paper details automated segmentation and analysis algorithms to identify and analyze the quality of construction for pier foundations.

2 Methodology

The focus of this research is the development of automated segmentation and analysis algorithms for point clouds of pier foundations. The algorithms are incorporated into a software tool called the Pier Inspection and Evaluation Report (PIER). As part of the analysis, the user is informed in real-time about differences in the top elevations of piers and misalignment of pier locations based on an assumed orthogonal planar layout. The method takes as input a dense point cloud and does not require user interaction to automatically segment points belonging only to piers and analyze the quality of construction. For the initial design of the algorithms, masonry piers were used.

2.1 Segmentation

The segmentation process is performed to identify and isolate points belonging only to piers from a dense point

cloud. This segmentation is necessary to identify and isolate pier locations before the automated analysis can be conducted. Figure 2 illustrates the steps in the segmentation procedure.

The first step is to remove points above the scanner origin. The top surfaces of piers will not be visible from the scanner unless the scanner is located at a higher elevation than all the top piers surfaces. As a result, only points below the scanner are relevant. This step helps to reduce the total number of points initially and ensure that any elements above the scanner origin are not incorrectly identified as piers. To perform this removal, the point cloud is cropped such that all points with an elevation above zero are removed from the point cloud.

It is essential to remove the points belonging to the ground around the piers. The next step consists of removing points belonging to the ground by using random sample consensus (RANSAC). RANSAC consists of randomly sampling points to generate planes and select inlier point sets that fall within a specified normal distance from that plane [13]. All points that fall within a specified distance threshold are considered inliers within the selected plane. The procedure in PIER randomly selects five points in the cloud to fit a plane. The distance threshold represents the normal distance from the plane generated and assigns points within this distance as inliers. After a specified number of iterations, the plane with the largest number of inlier points is returned as the plane representing the ground [6]. For instances where the ground around the piers is largely non-planar, the procedure may struggle to effectively remove the ground. However, it is only necessary to remove the ground around the piers to ensure there is clear separation between the points at the bases of piers and points belonging to the ground.

The next step is to prepare the point cloud data for computationally expensive algorithms. To ensure the code is not too computationally expensive and that the run time is minimized, data reduction is achieved by voxel-down sampling [14]. A grid of cubes with a specified dimension is computed. All points within each cube are averaged and reduced to a single point. In this work, a voxel size of 12.5 cm appears to be effective in reducing the point cloud while still accurately providing enough fidelity in the data for the remaining steps in the procedure [6].

Points at the edge of the point cloud are also removed to make the code run more quickly and improve robustness. To do so, convex hulls and K-D tree are used. The convex hull is an algorithm that computes the smallest convex shape of spatial data points: given $P = \{p_1, p_2, \ldots, p_n\}$, Convex Hull$(P) = \bigcap\{H \mid P \subseteq H, H \text{ is convex}\}$. The implementation of the convex hull algorithm in the Open3D library (Qhull algorithm) is used in this work [14]. In this implementation, a tetrahedron is initialized with vertices at extreme points (maxima and minima of x, y, and z). Then the remaining points outside the shape are checked and included as long as the convex property is maintained [15]. Computing the convex hull of the point cloud allows us to identify points composing the vertices or the borders of the hull. The next step is to find and remove the neighbors of these border points by using KD-Tree (k-dimensional tree) for computational optimization purposes. The K-D tree is a space-partitioning binary search tree for organizing points and enables the selection of points based on nearest neighbors [16]. Space is partitioned with an axis-aligned hyperplane. Embedded functions in Open3D easily allow the organization of the point cloud in a KD-Tree and selection of neighbors within a radius of the border points identified with the convex hull procedure to be removed.

After removing the ground and the points at the extremities of the point cloud, a density-based scan (DBSCAN) clustering method ([14] [17]) is used to identify the different clusters in order to isolate the point sets belonging to the piers. A set of filters are used to correctly identify which sets of clusters are pier foundations. The algorithm iterates through the projection of each cluster on xy plane to compare the area difference between the minimum oriented bounding box and the contour of the cluster. The minimum oriented bounding box algorithm determines the smallest rectangle at any orientation that encapsulates all points within the set [18]. The contour is computed using alpha shape: a generalization of convex hull that is capturing all vertices and allows concavity [19]. In addition, the contour area is compared with the user-specified dimensions of piers to ensure a robust selection of clusters belonging to piers. In most cases, the dimensions of the piers are normalized or easily quantified in the construction industry. The user input pier dimensions are not required but do improve the robustness of pier identification. All individual point clusters identified as piers are joined to create a single point cloud. At this stage, the point cloud includes only points that have been identified as piers and is ready for analysis.

2.2 Analysis

After segmenting the piers, labels are set based on the distance between the centroid of each pier and the origin of the point set. By using this naming convention, the identification of each pier remains consistent even if the user takes multiple scans and runs the segmentation and analysis multiple times. Labels are set as P_i. The current objectives of the analysis are to produce two plots: (1) a quantified illustration of the differences in elevations between the top surfaces of piers that allows the user to shim piers such that the top surface of each pier is at the same elevation, and (2) a quantified illustration of the

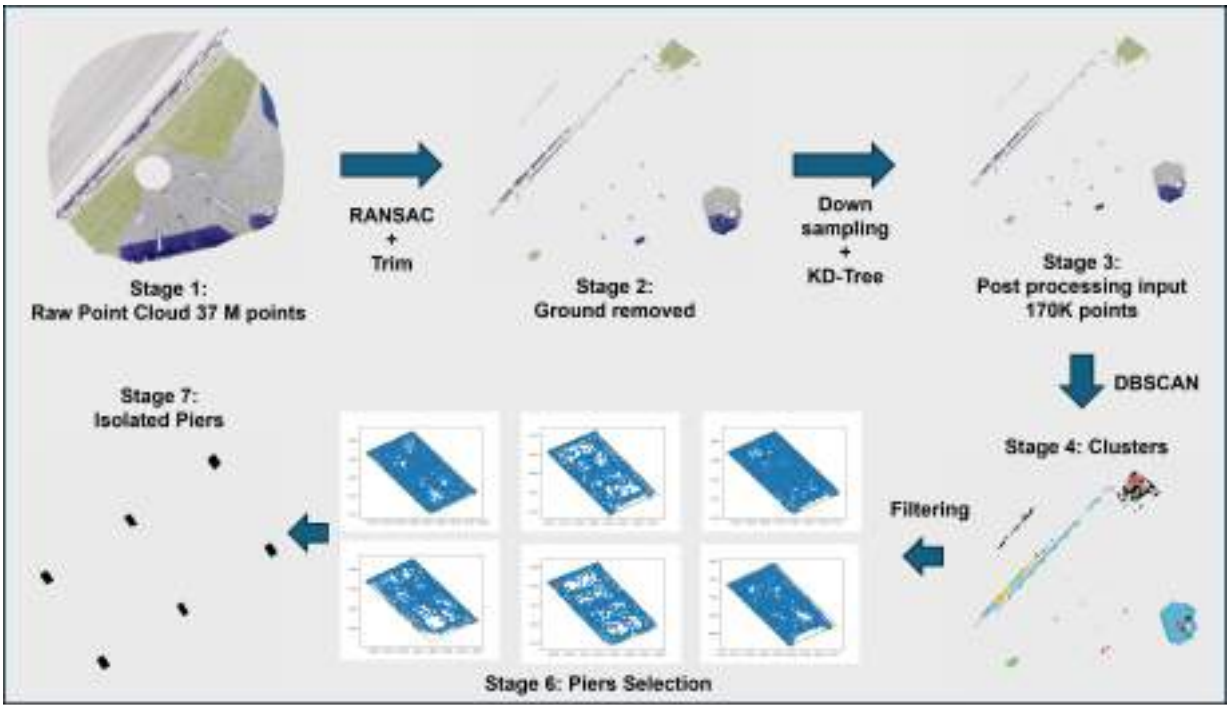


Figure 2. Steps in the automated segmentation process for pier foundations.

misalignment that informs the user about incorrect planar locations of piers.

To determine the difference in elevation between the top surfaces of piers, a point of reference must be established. The highest point of the point cloud, including only points identified as piers, is computed and referred to as h to be used as the reference. For each pier, the local maximum elevation h_i is identified, and a plane parallel to the xy-plane is created using the local maxima h_i at centroid c_i. For each of these planes, a perpendicular vector is computed, originating from the corresponding plane's centroid c_i and extending to the corresponding point where the z-coordinate matches the global maximum elevation h. This analysis results in the reported minimum distance from the top surface of the pier at the highest elevation to the top surface of each other pier as shown in Figure 3. With this information, the user can adjust piers top elevation by shimming appropriately. For example, in Figure 3, P2 is shown to be 18 cm below P6 (the datum). To correctly bring the top elevation of P2 to a uniform elevation, a shim or block with total thickness of 18 cm would need to be added to the top of the second pier. After shimming, the site can be scanned again, and the automated segmentation and analysis can be repeated to determine if the piers were correctly shimmed.

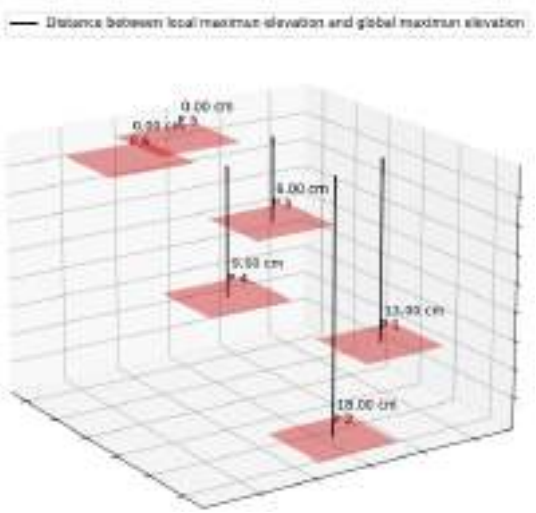


Figure 3. Elevation differences between top surfaces of piers.

During the foundation construction, the planar locations of piers need to be accurate to avoid issues with the construction of the structure above. To give corrective guidance for planar location, the automated analysis generates an illustration of planar misalignment. To generate this guidance, the total point cloud including only piers is projected onto xy plane (Figure 4). In the current development, the algorithms assume a rectangular layout of piers. The measured diagonals formed between piers at corner locations are presented to show the quality of the

rectangle produced by pier layout. If the measured distance of diagonals are substantially different, the user can know that the interior angles of the rectangle forming the layout of the piers are not orthogonal indicating an error in the construction.

Additionally, local guidance is presented to the user on how to correct the location of each pier to achieve an orthogonal rectangular layout. The contour and minimum oriented bounding box of the total point cloud including only piers are computed. For each pier, a minimum oriented bounding box is computed for all points on the xy plane. The closest point on the segment of the total bounding box from each pier extremity is given by their respective orthogonal projection (Figure 5). For example, let H be the minimum oriented bounding box of the total point cloud, and let a_i, b_i, c_i, d_i be the vertices generating the minimum oriented bounding box of the each pier P_i. For each a_i, b_i, c_i, d_i, the minimum distance between these points and H is computed and presented to the user. For example, as shown in Figure 5, the pier in the top left corner needs to be adjusted nearly 14 cm to be correctly set to create a planar rectangular layout with the remaining piers.

If the displacement between one of the points a_i, b_i, c_i, d_i for each P_i and H is less than 2 mm, it is not going to be plotted. Also if this displacement is higher than the dimension of the piers specified by the user, it is not going to be computed since it does not bring valuable information.

2.3 Limitations

In this version, the automated segmentation and analysis algorithms include some limitations. Each limitation will be addressed in future research.

First, this initial study focuses on the general performance and feasibility of the proposed algorithm for a single case study. Future research will focus on multiple demonstrations to evaluate robustness and accuracy of outputs.

The final alignment plot can be misleading if the errors are too large. Since the code is looking for the displacement needed so that all the piers are within an optimized rectangle, if large errors were made during placement, the assumption for a rectangular layout is not valid, and the algorithms generate incorrect guidance.

The number of points attributed to the piers is low compared to the remaining elements such as the ground. During the step where the ground is removed from the point cloud, a few failures occur where a pier is inadvertently removed in some instances.

In the current form, the automated analysis algorithms do not provide guidance on the levelness of the top surface of each pier. Information of the levelness of the top surface is critical to ensuring that constructability issues are not created. A feature to extract and report on the levelness of the top surface of each pier will be added to the algorithms.

In the current development, the algorithms assume a rectangular layout of piers. The measured diagonals formed between piers at corner locations are presented to show the quality of the rectangle produced by piers layout. In future developments, a digital twin is planned to be incorporated to generate guidance for complex layouts. However, in this case, alignment of the digital twin with the physical space will need to be performed to generate applicable correct guidance for users.

Due to the automated nature of the analysis procedure for correct planar rectangular layout, the algorithm assumes that the most exterior pier within the minimum oriented bounding box is the reference point. As a result, all guidance is to move the remaining piers outwards from the center of the layout. This assumption may not always be correct as a certain pier may need to be used as the reference. In future development, the user will be allowed to choose which piers should be treated as the reference point and remain unmoved.

3 Demonstration

The automated segmentation and analysis algorithms were tested on a mockup pier foundation construction located on the Oak Ridge National Laboratory site. Six masonry piers were placed in a planar rectangular layout with design dimensions of 7.5 m by 3.6 m. The mockup pier foundation was scanned using a terrestrial laser scanner.

The scan was made with a Leica MS60 which was placed in front of the piers and was set to do a 360° scan of the area to have a full view of the mockup construction site. The same scan parameters as in the previous paper [6] were used. Figure 6 shows the raw .pcd file of the scan that was loaded into the PIER digital tool. The masonry piers are clearly visible by their "shadows" cast from the laser scanning procedure.

The pier segmentation and analysis algorithms were written using Python. Table 1 shows the list of parameters and values used to analyze the mockup foundation. Voxel Size is the dimension of the voxel cube used for downsampling. The Plane Distance Threshold is the minimum normal distance between the RANSAC plane and points to be considered as inliers. Cluster Neighbor Radius is the maximum distance between points to be considered members of the same cluster. Cluster Minimum Points is the minimum number of points that a cluster must contain. KD-Tree Minimum Points is the minimum number of neighbors a point must have in order to not be considered a border point in the point cloud. KD-Tree search radius is the maximum distance where neighbors are taken into

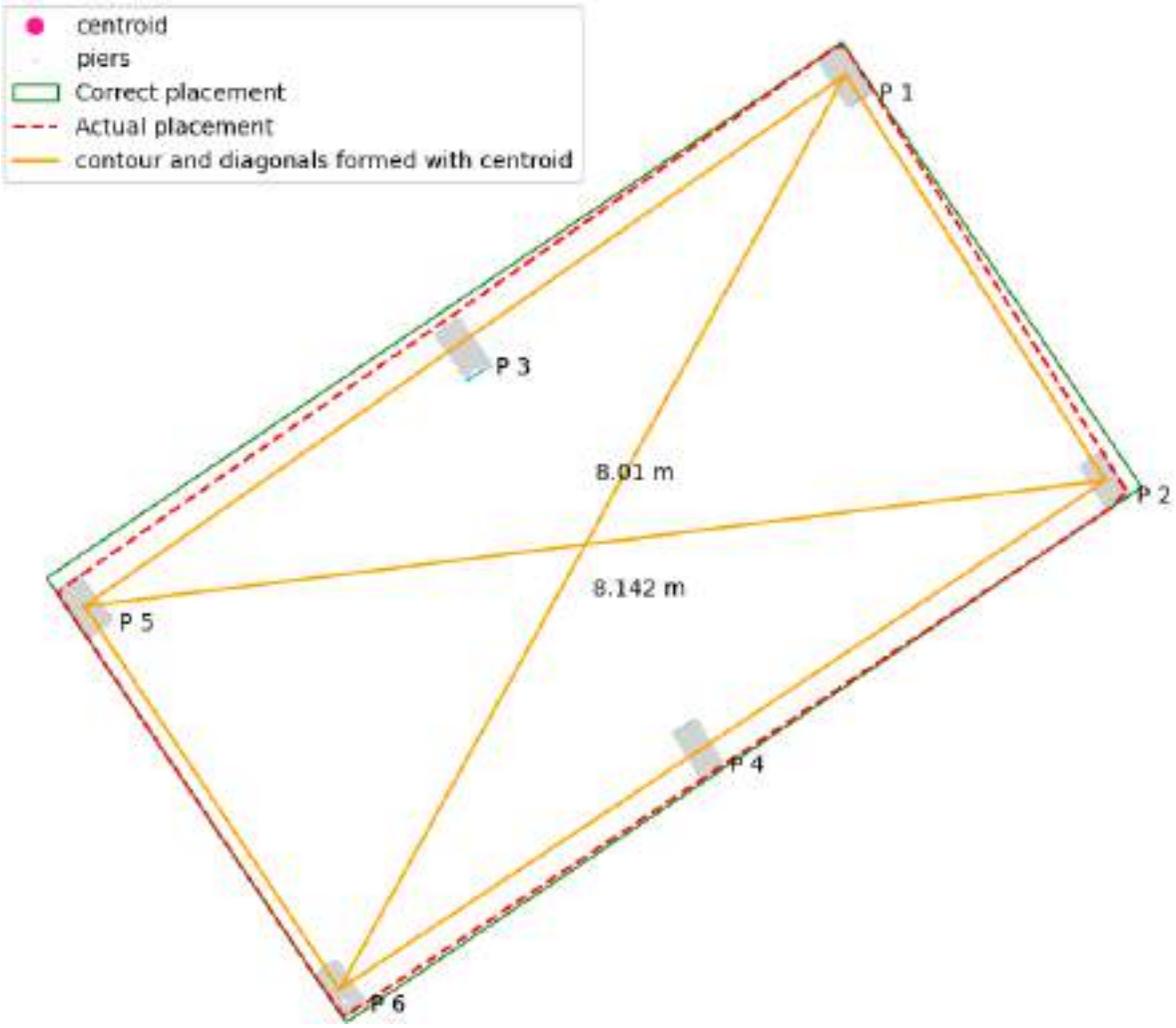


Figure 4. Planar alignment guidance for pier locations

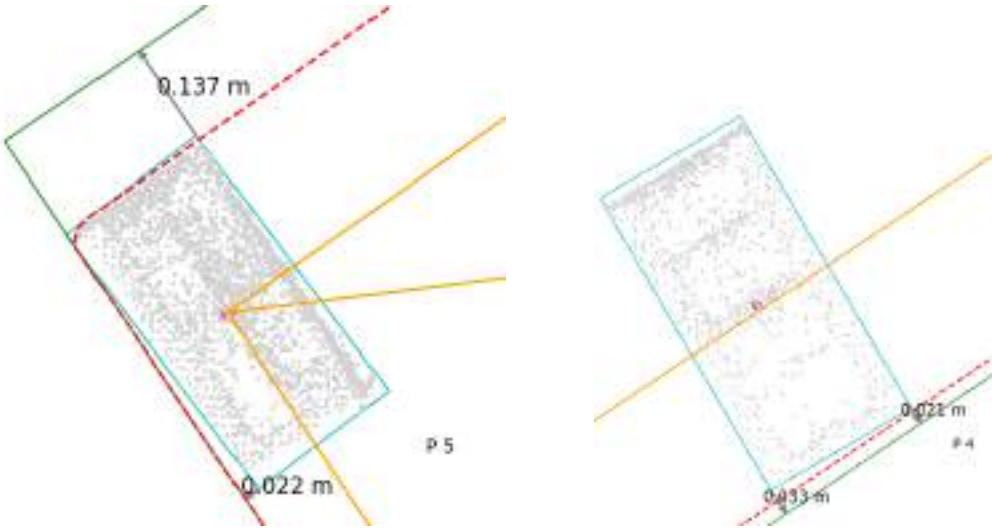


Figure 5. Local planar correction to locate each pier



Figure 6. A portion of the full-dome scan of the site with mockup pier foundations.

account.

Table 1. Parameters for the demonstration.

Parameter	Value
Voxel Size	12.5 cm
Plane Distance Threshold	20 cm
Kd-Tree Search Radius	20 cm
Kd-Tree Minimum Points	1000
Cluster Minimum Points	25
Cluster Neighbor Radius	50 cm

For the MS60 laser scanner used in this study, the single-view dense point cloud with a specified resolution was collected in approximately 45 minutes. However, many TLS scanners boast faster scanning rates and can capture point cloud data with higher resolution in a fraction of the time (3 minutes or less).

Table 2 gathers the running time for each main task required to segment and analyze the piers foundation on a standard laptop computer. Tasks before segmentation included the PCD file reading, trimming, and pre-processing with RANSAC, voxel-down sampling, and KD-Tree. The clustering by density-based scan (DBSCAN) and the cluster filtering are reported collectively within the segmentation task. Finally elevation and alignment guidance tasks represent the computation and analysis of piers elevation and placement.

The mean total run time was approximately 27.20 seconds. Most tasks within the algorithms have consistent run time except for pre-processing and segmentation. This is due to the random nature of RANSAC. Since the tests were not conducted on a flat surface, some parts of the ground may not be removed and taken into account during the segmentation which increases the computation time. Another method to identify the ground may be introduced in the future in order to address this issue.

Among 40 iterations, two runs were considered failures. One failure was due to RANSAC removing the entire collection of points belonging to a single pier. The other failure consisted of an additional element in the construction site being incorrectly identified as a pier. This variability is mainly due to the ground removal with RANSAC and the clustering method that use unsupervised machine learning, resulting in different cluster identifications between attempts. Another cause of failure is if there is an object on site that has approximately the same dimension as the piers since the algorithm is filtering the different identified clusters based on the dimension specified by the user.

Table 2. Run times. 40 iterations.

Task	Run Times (sec)			
	Min	Max	Mean	Std
Read PCD	3.77	4.11	3.89	0.09
Trim	0.01	0.03	0.01	0.01
Pre processing	4.27	5.80	5.10	0.36
Segmentation	13.06	25.25	16.33	3.08
Elevation	0.21	0.38	0.28	0.04
Alignment	1.78	1.04	1.4	0.15
Entire Procedure	23.47	36.52	27.20	3.14

4 Conclusions and next steps

Pier Inspection and Evaluation Report (PIER) is a digital tool that uses algorithms to automate the analysis and inspection of point clouds of pier foundations. Tests were conducted on masonry piers at Oak Ridge National Lab site. The results show that PIER can provide quick and accurate feedback on piers installation.

In future work, an alternative way to remove the ground to reduce the variability of the segmentation will be implemented. The selection algorithm for identifying point clusters of piers will be improved to increase robustness. Futhermore, a feature to extract and report on the levelness of the top surface of each pier will be added to the algorithms.

Additionally, future work will include research on the robustness and accuracy of the proposed methods through multiple demonstrations at construction sites where piers are being constructed. PIER will be demonstrated and verified to produce accurate results and guidance.

5 Acknowledgements

This research was supported by the DOE Office of Energy Efficiency and Renewable Energy, Building Technologies Office, under the guidance of Sven Mumme, and used resources at the Building Technologies Research and Integration Center, a DOE-EERE User Facility at Oak Ridge National Laboratory.

References

[1] IRC 2024. International Residential Code (IRC). Building code, International Code Council, 2024.

[2] ACI SPEC-336.1-01. Specification for the Construction of Drilled Piers. Standard, American Concrete Institute, 2001.

[3] ACI SPEC-336.1-24. Construction of Drilled Piers - Specification. Standard, American Concrete Institute, 2024.

[4] Tao Wu, Shitong Hou, Weihao Sun, Jing Shi, Fujun Yang, Jian Zhang, Gang Wu, and Xiaoyuan He. Visual measurement method for three-dimensional shape of underwater bridge piers considering multirefraction correction. *Automation in Construction*, 146:104706, 2023. ISSN 0926-5805. doi:https://doi.org/10.1016/j.autcon.2022.104706.

[5] Associated General Contractors. New Survey Shows How Failure To Invest In Construction Education Makes It Hard For Firms To Build - AGC News — news.agc.org. https://news.agc.org/labor-hr/new-survey-shows-how-nations-failure-to-invest-in-construction-education-training-programs-makes-it-hard-for-firms-to-build/, 2024. [Accessed 16-12-2024].

[6] Nolan W Hayes, Bryan P Maldonado, Mengjia Tang, and Diana Hun. Flat and Level Analysis Tool (FLAT) for real-time automated segmentation and analysis of concrete slab point clouds. In *Proceedings of the 41st International Symposium on Automation and Robotics in Construction*, pages 838–846, Lille, France, June 2024. International Association for Automation and Robotics in Construction (IAARC). ISBN 978-0-6458322-1-1. doi:10.22260/ISARC2024/0109.

[7] Shuangping Li, Bin Zhang, Junxing Zheng, Dong Wang, and Zuqiang Liu. Development of automated 3d lidar system for dimensional quality inspection of prefabricated concrete elements. *Sensors*, 24(23), 2024. ISSN 1424-8220. doi:10.3390/s24237486.

[8] Yan Xu, Yi Luo, and Jian Zhang. Laser-scan based pose monitoring for guiding erection of precast concrete bridge piers. *Automation in Construction*, 140:104347, 2022. ISSN 0926-5805. doi:https://doi.org/10.1016/j.autcon.2022.104347.

[9] Min-Koo Kim et al. Automated dimensional quality assessment for formwork and rebar of reinforced concrete components using 3d point cloud data. *Automation in Construction*, 112:103077, 2020.

[10] Mads A Kuhlmann-Jørgensen et al. Plastr: Planning for autonomous sampling-based trowelling. *IEEE Robotics and Automation Letters*, 8(8):5069–5076, 2023.

[11] Pingbo Tang et al. Automatic reconstruction of as-built building information models from laser-scanned point clouds: A review of related techniques. *Automation in Construction*, 19(7):829–843, 2010.

[12] Bryan P. Maldonado et al. Automatic point Cloud Building Envelope Segmentation (Auto-CuBES) using Machine Learning. In *Proceedings of the 40th International Symposium on Automation and Robotics in Construction*, pages 48–55, Chennai, India, 2023.

[13] Martin A. Fischler and Robert C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981. doi:10.1145/358669.358692.

[14] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. Open3D: A modern library for 3D data processing. *arXiv:1801.09847*, 2018.

[15] C. Bradford Barber, David P. Dobkin, and Hannu T. Huhdanpaa. The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software*, 22(4):469–483, December 1996. doi:10.1145/235815.235821. URL `http://www.qhull.org`.

[16] Jon Louis Bentley. Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18(9):509–517, 1975. doi:10.1145/361002.361007.

[17] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. pages 226–231, 1996.

[18] Joseph O'Rourke. Finding minimal enclosing boxes. *International Journal of Computer & Information Sciences*, 14(3):183–199, Jun 1985. ISSN 1573-7640. doi:10.1007/BF00991005.

[19] Kevin Bellock. alphashape: A python implementation of alpha shapes. `https://github.com/bellockk/alphashape`, n.d.

A Synthetic Data Generation Pipeline for Point-Cloud-Based Rebar Segmentation

Tao Sun[1], Yingtong Luo[2] and Yi Shao[1]

[1]Department of Civil Engineering, McGill University, Canada
[2]Department of Mechanical Engineering, McGill University, Canada
tao.sun@mail.mcgill.ca, yingtong.luo@mail.mcgill.ca, yi.shao2@mcgill.ca

Abstract

Automated rebar cage assembly and quality inspection require reliable rebar recognition. Although rebar segmentation from point clouds has been extensively studied, its generalizability remains limited. One key challenge is the scarcity of real data for training the segmentation models. To address this issue, we propose, for the first time, a pipeline for generating synthetic data for the rebar point cloud instance segmentation task. Using this pipeline, we applied the state-of-the-art Oneformer3d on rebar mesh instance segmentation. The model trained on our synthetic dataset achieved 92.1 mAP in real-world experiments, showing strong synthetic-to-real transfer capability. By eliminating the need for manual data collection and annotation, the proposed method facilitates advancements in automated rebar cage assembly and dimensional quality inspection technologies.

Keywords –

Point cloud instance segmentation; Rebar Recognition; Synthetic-to-Real; Computer Vision

1 Introduction

The automation of rebar cage assembly and dimensional quality inspection has garnered significant attention in the construction industry due to its potential to enhance productivity and reduce labor costs [1]. Accurate rebar recognition is a key technology that enables the automation of the above tasks. It serves as a critical foundation for many downstream tasks, such as rebar manipulation based on pose estimation and the dimensional measurement of rebar cages.

Existing studies have extensively explored rebar segmentation from images or point clouds to accurately identify the shapes and dimensions of rebars. Depending on the type of sensor data, these methods can be categorized into image-based rebar segmentation methods and point cloud-based segmentation methods. The advantage of the former lies in the ease of data collection and annotation and lower hardware costs. For instance, Kardovskyi et al. [2] applied the Mask R-CNN to segment rebars from images, which was then integrated with stereo vision techniques to measure rebar length and spacing. Chang et al. [3] present a hybrid rebar segmentation method to enable accurate rebar dimension measurement. This method first employs Mask R-CNN to segment rebars from images, followed by mapping the segmented results into 3D space, where point cloud processing techniques are used to cluster and identify 3D rebar instances.

Compared to object segmentation from images, direct segmentation from 3D point clouds offers two significant advantages: (1) it enables end-to-end processing of point clouds from laser scanners, and (2) it leverages prior knowledge of objects' 3D shapes more effectively. Consequently, many studies have focused on rebar segmentation from 3D point clouds [4], [5], [6]. They all relied on point cloud processing techniques, leveraging prior knowledge of rebar shapes to segment point clouds from complex scenarios. For example, Kim et al. [7] proposed a method based on the line RANSAC method and circular RANSAC to efficiently identify straight rebars within rebar cages. However, relying solely on point cloud processing to identify rebars faces limitations in generalizability [3].

Deep learning-based point cloud instance segmentation has demonstrated remarkable performance in various computer vision tasks, such as indoor scene understanding [8] and industrial part dimension inspection [9]. They typically rely heavily on large quantities of annotated data for training [10]. However, such large-scale data in the context of rebar segmentation is still lacking [3], which directly hinders the advancement of rebar segmentation. Moreover, the labor-intensive and time-consuming nature of manual annotation processes further exacerbates this challenge. Wang et al. [11] proposed a synthetic dataset using BIM and Lumion 11 rendering software, which contains automatically labeled 2500 synthetic images of rebar meshes. Mask R-CNN was trained and tested on their proposed dataset, showing high accuracy of bounding

box and mask prediction. However, to the best of the authors' knowledge, no point cloud dataset specifically designed for rebar instance segmentation has been proposed. Therefore, the applicability of advanced point cloud instance segmentation networks to rebar segmentation tasks also remains unexplored.

To address these critical gaps, we first develop a synthetic data generation pipeline for generating synthetic point clouds. The pipeline simulates realistic rebar point clouds from image-based 3D reconstruction and their corresponding 3D labels, enabling the training of advanced segmentation networks without the need for manual data collection and annotation. Using the proposed synthetic dataset, we apply a state-of-the-art point cloud instance segmentation network to the real-world task of rebar mesh recognition. The feasibility and accuracy of the developed pipeline are validated through real-world experiments. Additionally, the experiments show both the applicability and limitations of the state-of-the-art method in rebar segmentation tasks.

2 Method

Point cloud instance segmentation networks typically require a substantial amount of training data to enable the model to effectively learn the visual and geometric features of target objects. Therefore, we developed an end-to-end data generation pipeline for rebar instance segmentation based on the modeling and rendering software Blender, as shown in Figure 1. This pipeline efficiently simulates 3D point clouds reconstructed using Structure-from-Motion (SfM) and Multi-View Stereo (MVS) algorithms and automatically generates labels. The proposed pipeline consists of three stages: (1) scene generation, (2) point cloud reconstruction, and (3) model training.

2.1.1 Scene generation

We first developed a parameterized modeling method using Blender and Infinigen [12]. The method can generate large-scale rebar models with diverse shapes and appearances. Specifically, Blender is used to model rebars, while Infinigen converts the parametric geometry nodes into executable code for batch processing. Figure 2 shows different rebar models generated by our method. We use Cylinder primitives in Blender to represent straight rebar segments. To simulate different real rebars, our straight rebar model supports parameters required for typical design specifications, including rebar diameter, length, bending degree, and rib type. The rib structures are created using meshes, enabling high-fidelity and geometric details.

To simulate different rebar appearances, we randomize visual parameters such as base color, metallic properties, roughness, and texture to randomize the visual appearance of the rebar models. Additionally, we use a combination of Noise Texture and Musgrave Texture nodes to simulate surface defects on rebars, such as stains and corrosion.

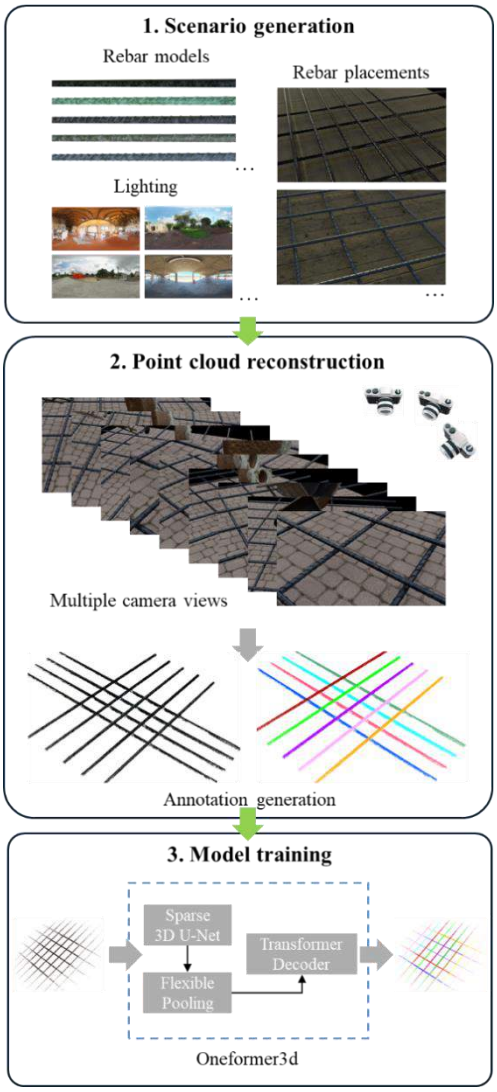


Figure 1. Proposed pipeline for rebar instance segmentation

By defining heuristic rules and graph structures, we can use automatically generated rebar models to establish various rebar cage configurations. In this study, we specify the positions and orientations of straight rebars to construct reinforcement mesh scenarios for slab production. For lighting, the approach of high dynamic range images (HDRI) is applied to illuminate the scenario. To capture a wide range of lighting conditions, 733 HDRI images are collected from the open-source asset library PolyHaven.

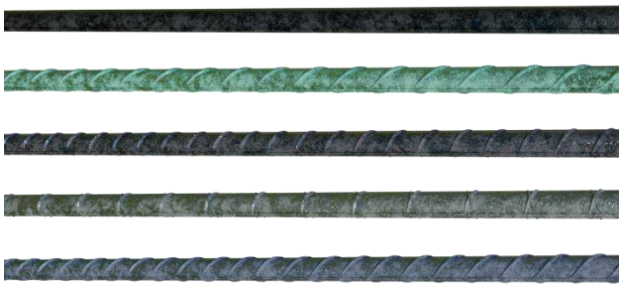


Figure 2. Examples of generated rebar models

2.1.2 Point cloud reconstruction

A straightforward method for generating point clouds from 3D scene models is to downsample the mesh directly to obtain the point cloud [13]. However, this approach struggles to effectively replicate the distribution of point clouds reconstructed from real images, leading to reduced segmentation accuracy [14]. To better simulate the point clouds derived from real image reconstruction, we sample cameras within the 3D scene and employ traditional SfM [15] and MVS [16] algorithms to generate synthetic point clouds. Blender's Cycles renderer is applied to generate photorealistic images of the scenarios. The Cycles renderer utilizes physically based ray tracing to simulate complex interactions between light, shadows, and material reflections.

To accurately label the reconstructed point cloud P_{re}, we first perform uniform downsampling on the rebar mesh model generated in the synthetic scenario, denoting it as P_{GT}. Subsequently, we align P_{re} and P_{GT} by utilizing the camera poses estimated via SfM and the ground-truth poses of the virtual cameras. Based on this alignment, the nearest neighbor for each point in P_{re} is identified in P_{GT}, allowing instance labels to be assigned to P_{re}.

2.1.3 Instance segmentation network

A SOTA transformer-based instance segmentation network, Oneformer3d [17], was used for the instance segmentation model. The method achieved the best performance on three different benchmarks (ScanNet, ScanNet200, and S3DIS), and excels in scenarios involving slender objects like chair legs, which share similar geometric characteristics as with rebars.

3 Experiment

To evaluate the adaptability and performance of the developed pipeline, a large-scale synthetic dataset was first generated. This dataset was designed to be diverse, incorporating randomized rebar models, spacing configurations, lighting conditions, camera numbers and camera poses. Based on this dataset, we trained the state-of-the-art point cloud instance segmentation model, Oneformer3d. To assess the model's syn-to-real generalization capability, real-world experiments were conducted on rebar mesh arrangements, simulating slab reinforcements that are commonly used in construction sites.

3.1 Dataset generation

Based on the proposed pipeline, a synthetic dataset containing 238 samples was reconstructed automatically, as shown in Figure 3(a-b). Each sample was annotated with detailed instance labels. The dataset was divided with an 8:2 ratio into a training set (190 samples) and a validation set (48 samples).

To evaluate syn-to-real transfer performance, we arranged slab rebar meshes to collect real-world point cloud data. A smartphone was used to capture 48 to 79 multi-angle images of each rebar placement, with a resolution of 4032 × 3024 pixels. The camera was positioned approximately 70 cm from the target. To introduce variability, we adjusted the number of rebars (8 to 14) and their placement intervals (5 cm to 25 cm), aiming to challenge the algorithm and assess its robustness under diverse spatial configurations. The final test set comprises 12 manually annotated samples. Figure 3(c-d) provides examples from our real-world test set.

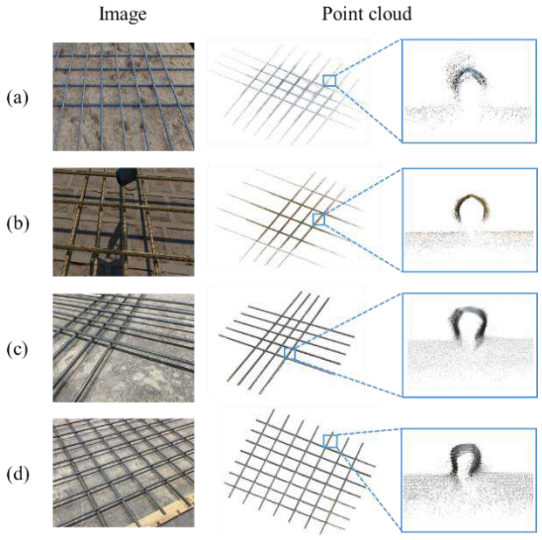


Figure 3. Examples of our synthetic training set and real-world test set: (a-b) synthetic training set; (c-d) real-world test set

3.2 Evaluation Metrics

We evaluated the performance of the 3D rebar instance segmentation using mean average precision (mAP), a standard metric in point cloud segmentation. The mAP was computed as the average precision over Intersection over Union (IoU) thresholds ranging from 50% to 95% in increments of 5%.

3.3 Training setting

All training settings of Oneformer3d are basically following the original implementation [17]. For the learning rate, we used an AdamW optimizer with an initial learning rate of 0.0001 and weight decay of 0.05. Taking computational efficiency into account, we employed a batch size of 2 and constrained the training to 160 epochs. Furthermore, we have experimentally observed that the scale diversity of the point cloud can significantly improve the sensitivity of the segmentation network to the scale of the point cloud. Therefore, we applied a larger range of random scaling, from 0.5 to 1.5, compared to the original implementation.

3.4 Results

Figure 4 illustrates the progression of the mAP on the validation set throughout training. The trend indicates that training on our synthetic dataset is stable and effective, with no signs of significant overfitting. Oneformer3d demonstrates strong segmentation performance on the synthetic dataset, achieving the highest mAP of 99.4. When evaluated on the real-world test set, the model achieved a high mAP of 92.1, indicating its ability to generalize from synthetic training data to real-world scenarios.

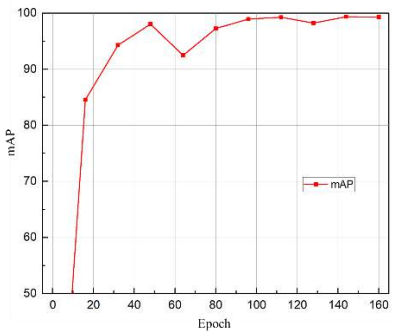


Figure 4. mAP variation with training epochs on the synthetic validation set

Figure 5 presents several results obtained from the real-world dataset. The majority of rebar outputs exhibit high segmentation accuracy with clearly defined instance boundaries However, a few cases exhibit poor segmentation performance. For instance, as illustrated by the red arrows in Figure 5(c), two rebars were incorrectly merged into a single instance. This may be attributed to the global context awareness of transformer-based models, which, in rare cases, can lead to objects with similar features being recognized as the same instance. This merged detection is also noted in indoor object segmentation [18] using a similar transformer-based model with Oneformer3d. Employing a distance-based clustering filter, such as DBSCAN[18], enhanced by shape priors, offers a potential solution to mitigate this challenge.

Additionally, missing points were observed in the segmentation results of certain rebars, as indicated by the blue circles in Figure 5(d). This can be attributed to the model's insufficient accuracy, particularly in handling incomplete edges in point clouds. This minor issue can be addressed using various post-processing techniques, depending on the downstream tasks. For instance, in the case of rebar 6D pose estimation, known geometric priors of the rebar can be utilized to accurately compute the pose even in the presence of missing points.

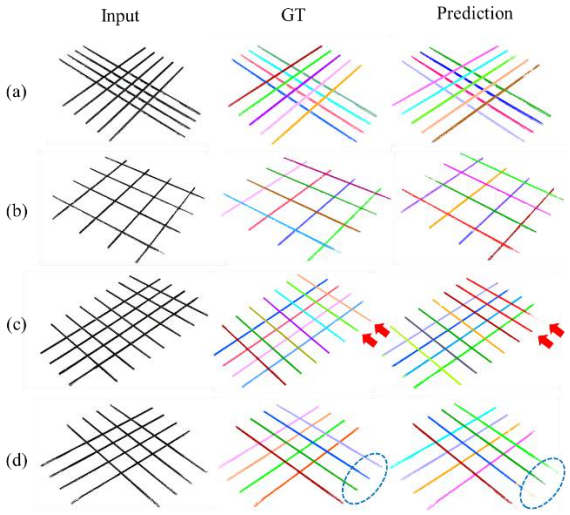


Figure 5. Examples of segmentation results on our real-world test set

4 Discussion

This study demonstrates the adaptability of synthetic data for training the rebar point cloud instance segmentation model. This validates that the approach can generate reliable rebar visual and geometric features, enabling deep neural networks to effectively learn and recognize real-world rebars. Therefore, it is expected that the proposed pipeline can automatically generate 3D data for a wide range of rebar cages. For example, as shown in Figure 6(a), our pipeline can generate complex rebar cage point clouds and annotations. Besides, the pipeline

shows promise for training networks to estimate 6D poses and 2D masks of rebars from 2D images. Furthermore, synthetic models can be integrated into physical simulations to train reinforcement learning policies for dexterous rebar manipulation in future work, as illustrated in Figure 6(d).

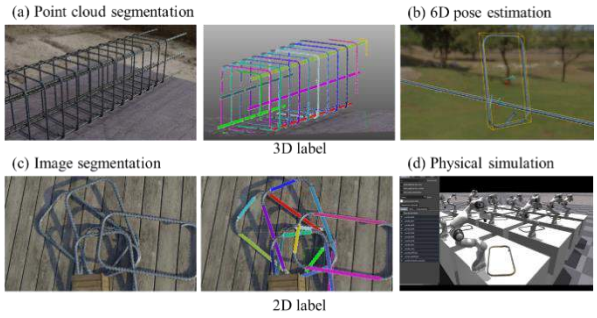


Figure 6. Our pipeline is promising to be extended in the future to: (a) 3D instance segmentation for rebar cage; (b) 6D pose estimation of rebars; (c) 2D image segmentation; (d) Physical simulation

This preliminary exploratory study has certain limitations. First, our point cloud segmentation experiments did not explicitly address the challenges posed by real-world variations, such as more complex rebar structures, occlusions, lighting conditions, and different scanner resolutions. While our synthetic dataset performs well in controlled settings, its generalization to diverse real-world environments remains an open question. Future work should focus on incorporating a broader range of 3D data that captures these variations, ensuring robustness across diverse rebar structures and scanning conditions. Additionally, while preprocessing techniques such as RANSAC [19], can effectively remove the ground, achieving end-to-end rebar segmentation in the presence of complex background interference using neural networks remains a valuable research direction.

Second, the proposed method faces challenges in achieving real-time running in practice due to the high computational cost of the point cloud segmentation network and the offline nature of the reconstruction method. In the future, adopting 3D cameras directly, combined with real-time reconstruction and lightweight segmentation or keypoint detection from 3D point clouds, will be a promising research direction. Third, future work could focus on comprehensive benchmarking to identify the optimal network architecture for rebar point cloud segmentation and facilitate model selection in real-world applications.

Finally, while this study focuses on demonstrating the effectiveness of synthetic data for rebar segmentation, a detailed ablation study on the impact of individual augmentation parameters, such as texture variations, occlusions, and lighting conditions, remains an important area for future work.

5 Conclusion

To address the challenge of lacking training data for rebar point cloud instance segmentation, we first developed a synthetic data generation pipeline. This pipeline enables high-fidelity annotated point clouds, facilitating the training of instance segmentation networks without the need for manual data collection and annotation. The adaptability of the pipeline was validated through a real-world experiment. The state-of-the-art point cloud instance segmentation network Oneformer3d trained on our synthetic dataset achieved a 92.1 mAP on the real-world test set, showing good syn-to-real capacity.

This work contributes to the advancement of automated rebar cage assembly and dimensional quality inspection technologies, paving the way for the broader adoption of intelligent construction solutions.

References

[1] A. Klarin and Q. Xiao, “Automation in architecture, engineering and construction: a scientometric analysis and implications for management,” Engineering, Construction and Architectural Management, vol. 31, no. 8, pp. 3308–3334, 2024.

[2] Y. Kardovskyi and S. Moon, “Artificial intelligence quality inspection of steel bars installation by integrating mask R-CNN and stereo vision,” Automation in Construction, vol. 130, p. 103850, Oct. 2021, doi: https://doi.org/10.1016/j.autcon.2021.103850.

[3] C.-C. Chang, T.-W. Huang, Y.-H. Chen, J. J. Lin, and C.-S. Chen, “Autonomous dimensional inspection and issue tracking of rebar using semantically enriched 3D models,” Automation in Construction, vol. 160, p. 105303, Apr. 2024, doi: 10.1016/j.autcon.2024.105303.

[4] H.-L. Chi, M.-K. Kim, K.-Z. Liu, J. P. P. Thedja, J. Seo, and D.-E. Lee, “Rebar inspection integrating augmented reality and laser scanning,” Automation in Construction, vol. 136, p. 104183, Apr. 2022, doi: 10.1016/j.autcon.2022.104183.

[5] X. Yuan et al., “Automatic evaluation of rebar spacing and quality using LiDAR data: Field application for bridge structural assessment,” Automation in Construction, vol. 146, p. 104708, Feb. 2023, doi: 10.1016/j.autcon.2022.104708.

[6] G. Hodge and J. M. Gattas, “Geometric and semantic point cloud data for quality control of bridge girder reinforcement cages,” Automation in Construction, vol. 140, p. 104334, Aug. 2022, doi:

10.1016/j.autcon.2022.104334.

[7] M.-K. Kim, J. P. P. Thedja, and Q. Wang, "Automated dimensional quality assessment for formwork and rebar of reinforced concrete components using 3D point cloud data," Automation in Construction, vol. 112, p. 103077, Apr. 2020, doi: 10.1016/j.autcon.2020.103077.

[8] Y. Sun, X. Zhang, and Y. Miao, "A review of point cloud segmentation for understanding 3D indoor scenes," Visual Intelligence, vol. 2, no. 1, p. 14, 2024.

[9] Y. Li, Y. Wang, and Y. Liu, "Three-dimensional point cloud segmentation based on context feature for sheet metal part boundary recognition," IEEE Transactions on Instrumentation and Measurement, vol. 72, pp. 1–10, 2023.

[10] H. Jiang, F. Yan, J. Cai, J. Zheng, and J. Xiao, "End-to-end 3D point cloud instance segmentation without detection," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 12796–12805.

[11] H. Wang, Z. Ye, D. Wang, H. Jiang, and P. Liu, "Synthetic Datasets for Rebar Instance Segmentation Using Mask R-CNN," Buildings, vol. 13, no. 3, p. 585, 2023, doi: https://doi.org/10.3390/buildings13030585.

[12] A. Raistrick et al., "Infinite Photorealistic Worlds using Procedural Generation," Jun. 26, 2023, arXiv: arXiv:2306.09310. Accessed: Jan. 10, 2024. [Online]. Available: http://arxiv.org/abs/2306.09310

[13] J. Shu, W. Li, C. Zhang, Y. Gao, Y. Xiang, and L. Ma, "Point cloud-based dimensional quality assessment of precast concrete components using deep learning," Journal of Building Engineering, vol. 70, p. 106391, Jul. 2023, doi: 10.1016/j.jobe.2023.106391.

[14] M. Chen et al., "Stpls3d: A large-scale synthetic and real aerial photogrammetry 3d point cloud dataset," arXiv preprint arXiv:2203.09065, 2022.

[15] P. Moulon, P. Monasse, R. Perrot, and R. Marlet, "OpenMVG: Open multiple view geometry," in International Workshop on Reproducible Research in Pattern Recognition, Springer, 2016, pp. 60–74.

[16] D. Cernea, "OpenMVS: Multi-View Stereo Reconstruction Library," 2020. [Online]. Available: https://cdcseacave.github.io/openMVS

[17] M. Kolodiazhnyi, A. Vorontsova, A. Konushin, and D. Rukhovich, "Oneformer3d: One transformer for unified point cloud segmentation," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 20943–20953.

[18] J. Schult, F. Engelmann, A. Hermans, O. Litany, S. Tang, and B. Leibe, "Mask3D: Mask Transformer for 3D Semantic Instance Segmentation," in 2023 IEEE International Conference on Robotics and Automation (ICRA), London, United Kingdom: IEEE, May 2023, pp. 8216–8223. doi: 10.1109/ICRA48891.2023.10160590.

[19] L. Li, F. Yang, H. Zhu, D. Li, Y. Li, and L. Tang, "An improved RANSAC for 3D point cloud plane segmentation based on normal distribution transformation cells," Remote Sensing, vol. 9, no. 5, p. 433, 2017.

Towards the Automation of Last Planner System ®

Michael Awe[1], Avleen Malhi[1], Nicholas Mavengere[1], Marcin Budka[1] and Bhargav Dave[2]

[1] Department of Computing and Informatics, Bournemouth University, UK
[2]VisiLean, UK

mawe@bournemouth.ac.uk, amalhi@bournemouth.ac.uk, nmavengere@bournemouth.ac.uk
mbudka@bournemouth.ac.uk, bhargav@visilean.com

Abstract -

The increasing complexity and uncertainty in construction projects necessitate more efficient and adaptive construction management systems. This research studies the automation of the Last Planner System (LPS), a collaborative lean scheduling tool, by integrating predictive strategies to enhance scheduling accuracy, constraint resolution, and insights for decision-making. A conceptual framework is proposed for LPS automation while also representing the automation of LPS as an optimisation problem and determining the core aspects of LPS, such as communication management, constraint management, and planning and scheduling. This framework provides a foundation to which integration of predictive strategies in digitised Last Planner Systems is possible, focusing on process improvement, delay mitigation, and performance optimisation. A model that determines the constraint resolution duration was presented. This model used feature extraction techniques to determine constraint complexity score and preprocessed other features such as priority, location, and category to predict constraint resolution duration. The results indicated a Mean absolute error of 21.38 days (MAE) from the actual constraint resolution duration. Although the outcome and model can help in constraint management towards an automated LPS, the research identifies key limitations, including challenges in dataset quality, unfavourable sample sizes, and challenges with parsing complex textual and tabular data, highlighting areas for improvement. Future research should prioritise exploring emerging techniques, such as Table parsing (TAPAS) and Retrieval-augmented generation (RAG), to enhance the integration of data-driven strategies into Last Planner System workflows.

Keywords -

Automation, Predictive Strategies, Last Planner System (LPS)

1 Introduction

In the face of increased complexities and uncertainties of projects in the construction sector, more efficient and intuitive project controls, planning, and scheduling systems are highly sought. Pressing factors contributing to inefficiencies in existing systems include inadequate management of constraints, communication, and lack of adaptability in scheduling and project control systems [1]. However, automatically classifying constraints, intuitively classifying project communication, instantly generating insights, and detecting deviations in these systems are crucial for navigating complex and high-risk projects such that predictability is enhanced and project outcomes are favourable [1, 2, 3]. To achieve this, a more adaptive project management system (Last planner system) is considered as prototyped by Ballard [4]. While existing project control and scheduling systems have indicated significant improvement in addressing inefficiency in project implementation, there remains a substantial gap for improvement in ascertaining project status, instant generation of insights from project trends, and ensuring predictable project outcomes [5, 6]. From the first principle, a last planner system (LPS) can, therefore, be represented as a combination of constraint management system (CN), collaboration/communication system (CB), and Planning and scheduling systems (PS), where x..n is the increasing project instances of elements contained that make each system functional;

$$LPS(x..n) \rightarrow CN(x..n)+CB(x..n)+PS(x..n) \quad (1)$$

Improved efficiency in existing project control and scheduling systems depends on the integration of digitised project control and scheduling systems with other technologies and methods such as lean systems, visual simulation, BIM, artificial intelligence, AR VR, etc. [1, 5]. These integrations have been proposed as solutions to identified fundamental limitations, including unreliable schedules, unreliable data, manual processes leading to errors, and overarching challenges of delay, cost overrun, and reworks [7]. These challenges are expected for most projects, especially in complex and high-risk projects where project managers faced with trade-offs in cost, quality, time, resources, changes, etc., have to rely on tacit knowledge, experience, and intuition to make accurate decisions quickly rather than relying on insights from analysed operation data to draw conclusions and make data-driven decisions. Studies have shown that although experience and tacit knowledge are essential in managing projects, it is more reliable for businesses to find data-driven paths to making operational decisions to control current and future project outcomes [8, 9].

While predictive strategies can improve most of the existing digitised project control and scheduling systems, this study

aims explicitly to automate digital last planner systems. LPS is a lean and collaborative tool that provides short-term schedules and lookahead information that enhances planning efficiency, improves the reliability and accuracy of plans, enables the selection of alternatives, and reduces time-space conflicts on construction sites. The system pulls information from all forms of communication regarding constraints, deviations, changes, etc., to determine tasks to be scheduled within specific lookahead periods. The benefits of LPS are evident in LPS-related research sources, and an automated Last Planner System (LPS) has been proposed to exert an even greater impact on construction planning and implementation processes. [10, 11, 7, 12, 13].

As part of the contribution of this research, this study presents a streamlined flow for automating the digitised Last planner system, considering all factors which enable it to be fully functional. In achieving this aim, the study follows a procedure that allows the identification of all relevant elements and features of a last planner system, represents the system as an optimisation problem and integrate predictive strategies in one of the identified elements (constraint management). Section 1 presents a summary of the conceptualisation of the topic, and Section 2 analyses existing literature and discusses the gaps and limitations. Section 3 presents the flow for the proposed automated last planner system. Section 4 initialises a basis for integration of artificial intelligence and presents a representation of the problem mathematically. Section 5 showcases the predictive strategies and offers implementation and discussion of results. finally, Section 6 and 7 discusses the research outcomes, limitations and future works.

2 Literature Review

The last planner system (LPS) is a project control and scheduling system created in the early 90s by Glen Ballard [4]. Research highlights that LPS emerged by applying lean construction (LC) principles in construction projects. It is a system based on the need to enhance workflow reliability by field workers implementing the task specified in planning processes. Key elements of LPS identified from existing research include 1) collaborative planning, 2) constraint management, 3) hierarchy planning systems and levels, 4) pull planning, and 5) continuous improvement [1, 3, 2].

Challenges of LPS were first discussed expansively by Aslam et al., [7]. The shortcomings identified include the inability to manage project variability correctly, lack of visualisation of LPS capabilities, incorrect lookahead planning and many more [14]. However, The conceptual framework for integrated LPS included necessary features such as collaboration, constraint identification and communication methods, which led to increased collaboration, improved visualisation, risk analysis, continuous improvement, etc. The implemented digitised LPS also has shortcomings, even though it is digitised. These shortcomings include limited support and integration, lack of consistency in planning and scheduling, inadequate real-time information, data management challenges, and limited constraint management [2]. A comprehensively automated digital LPS is expected to improve efficiency in all processes and activities, enhance constraint management, allow real-time data analysis and visualisation, allow knowledge reuse and to improve collaboration and communication through streamlining processes and enhancement of information sharing.



Figure 1. Last planner transformation stages

Developing a workable automated planning system is predicated on identifying all required base elements that make the system automated, collaborative, and manage constraints. Automation, as identified in existing research, is indicated in 3 levels; the first level is research based on the digitisation of LPS systems as against the manual forms [4]; the second level is the research instances which developed linkages and union nodes for BIM and digitised LPS as an upgrade to overcome data homogeneity and enhance information sharing [15, 16, 12]; the third level is the point where predictive strategies and AI techniques are being integrated into digitised LPS platforms to aid automation and streamline processes. [2]

A review study by Agrawal et al. [1] identified several existing research where automated planning processes were categorised into master and phase scheduling, lookahead scheduling, weekly work planning and learning stage. Although the study excluded actual automation of planning and control in LPS, it systematically reviewed existing works. The research identified research works which presented automation in all three identified levels in Figure 1. Further reviews and classification of related works are summarised in Table 1.

Overall, the early research by Ballard [4] has since been transformed to become more encompassing from the manual linear scheduling methods to the digitised methods. Then, the digitised systems that link LPS with BIM are often integrated for automation, allowing for improved scheduling of activities and streamlining processes. Existing related research has focused on other specific applications and stages of projects but neglected the three core aspects of a functional LPS. This is the gap which this study seeks to cover. More specifically, automation in the Last Planner system has been applied mainly in performance evaluation. A framework or prototype of an integrated LPS system that considers all the factors and

Table 1. An overview of existing literature

Sections	Overview	Stage	Citations
Last planner sytems	LPS concept application manually using pilot projects	Ist	[4]
Schdeuling Automation Review	A systematic review of existing works.		[1]
Master and phase scheduling	Optimisation of buffer projects used varying optimisation techniques to achieve maximised schedule allocations	2nd	[17, 18, 19]
Safety, risk and logistics constraints	Constraints that may impact projects using BIM as a source of information	2nd	[20, 21]
Lookahead planning	Automation-driven research which used transformer models to predict lookahead and generate plans quickly	2rd	[22, 3]
Constraints management	Identification and removal of constraints by first achieving integration between different data sources	3rd	[23, 24, 25, 26]
Last planner system (performance)	Used machine learning techniques to forecast and predict performance using identified LPS-based performance indicators such as the Percent plan complete (PPC) and other indicators.	3rd	[2]
Integrated last planner system ILPS	helps smoothen the implementation process and overcome the shortcomings identified in the LPS.	2nd	[12, 7, 13]

features of an LPS system appears non-existent, motivating the need for this study.

3 Last planner system Automation

Project control and scheduling systems ensure the flow of work and control the progress and outcomes of the projects. As shown in Figure 2, a typical project will be guided by a developed master schedule, which serves as input detail to other scheduling stages, including phase scheduling, lookahead, weekly work scheduling and daily scheduling. Once the master schedule exists, LPS principles can be integrated into processes, including short-term planning and lookahead. Figure 2 contains the flow chart for logical automation of Last Planner systems.

3.1 Lookaheads

In lookahead plans or short-term plans, determining what tasks will be ready to start in a lookahead period is very important. It can be based on the probability that all tasks before this period are completed [27]. The probability of task completion can be calculated using historical productivity rates. Measuring the percentage of work completed (PPC) relative to the time spent on a similar previous task provides a more accurate estimate of productivity rates than relying on generalised project or organisational productivity rates, as suggested by Salama et al., [27]. Their method involves determining performance based on forecasted productivity rates (using linear regression) and the remaining quantity of work. However, this approach assumes that the same organisation's employees consistently perform the tasks across all projects, which is often untrue. In reality, various contractors frequently complete tasks, each contributing to different aspects of the work. This variability complicates accurately determining productivity rates, especially on-site, where numerous contractors are involved in diverse activities. Consequently, there is a pressing need to maintain detailed performance data so that individual contractors can evaluate their work completion speeds better.

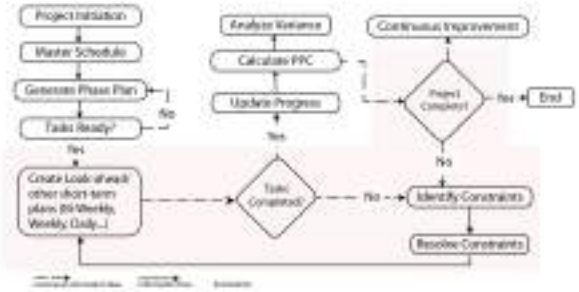


Figure 2. Proposed automated Last planner systems

4 LPS Aspects For AI Integration

As described already in equation 1, an LPS, on a fundamental basis, is a combination of three elements, which include collaboration, constraints and planning systems. This entities are further elaborated below;

AS in figure 3, the Project schedule contains a list of tasks (LT) detailing the following features: Task Name TN_1 .. TN_n, a unique identifier for each task. Location L_i .. L_n, the location identifier for each task. Trade: CT_i .. CT_n, the construction trade responsible for each task. Scheduled start and end dates: $t_{s,i}$ and $t_{e,i}$ for tasks TN_i, Duration: $D_{t,i} = t_{e,i} - t_{s,i}$, the duration of each task.

Constraint Table (CN) contains project constraints and their features: Constraint Name C_i .. C_n, a unique identifier for each constraint. Location L_i .. L_n, location identifier where the constraint exists. Trade CT_i .. CT_n, the trade associated with the constraint. Commitment start and end dates cts_i and cte_i. Constraint resolution time $D_{c,i} = cte_i - cts_i$, the time to resolve the constraint. finally, the tables LT and CN are linked through the linked task names, which associates constraints with corresponding tasks (TN_i).

Problem Definition

Given the tables LT and CN, we aims to optimize the inclusion of tasks ($TL \subset LT$) into a specified Lookahead Period (LAP). The goal is to dynamically determine an optimal subset of tasks while resolving associated constraints and efficiently allocating resources, subject to the following conditions:

- Binary Decision Variables (x_i); $x_i = 1$ if task i is included in the Lookahead Task List (TL), and $x_i = 0$ otherwise.
- Tasks left with status (ready) are well-defined for scheduling.
- All constraints are valid and removed for tasks ready for scheduling.
- The Lookahead Period is defined by the Lookahead Start Time (LAT_{start}) and Lookahead End Time (LAT_{end}).
- Tasks and constraints are resolved and aligned with resource availability and stakeholder collaboration.

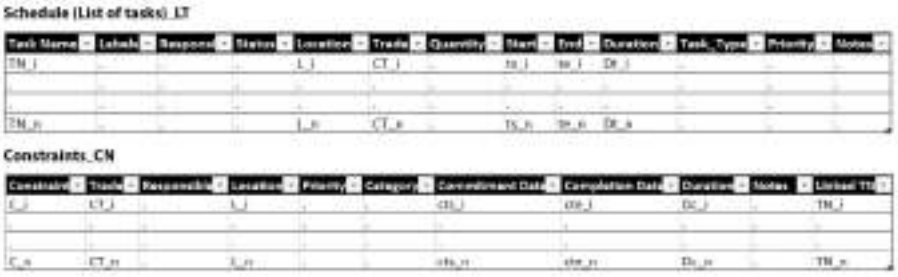


Figure 3. Standard Data-frame of the schedules and constraints entities.

Objective Function

As given in the problem definition, the objective function Z is to find the tasks TL in LT, which maximises Z.

amax Z =

$$\sum_{i \in \mathrm{TL}} w_i \cdot x_i \cdot D_{t,i} - \lambda_1 \cdot \mathrm{Pen}_{\mathrm{cons}} - \lambda_2 \cdot \mathrm{Pen}_{\mathrm{colab}} \tag{2}$$

Where w_i is priority weight for all tasks i (i.e., high = 3, medium = 2, low = 1), $D_{t,i}$ is duration of all tasks x_i, $\mathrm{Pen}_{\mathrm{cons}}$ is penalty for unresolved constraints and $\mathrm{Pen}_{\mathrm{colab}}$ is penalty for unresolved collaboration conflicts. Z is subject to the PS constraints, CN constraints, CB constraints and dynamic updates.

Schedule PS: This entity entails planning for different projects at all levels. Here, the critical dependencies, sequences, lookahead and buffers are represented accordingly.

- Task Dependency Constraint: tasks must satisfy their dependencies before being included in TL.

$$ts_i \geq te_j, \quad \forall (j \rightarrow i) \text{in dependencies.} \tag{3}$$

- Lookahead Time Period Constraint; Only tasks whose durations overlap with the Lookahead Period ($\mathrm{LAT}_{\mathrm{start}}$ to $\mathrm{LAT}_{\mathrm{end}}$) are considered

$$t_{s,i} \geq \mathrm{LAT}_{\mathrm{start}}, \quad t_{e,i} \leq \mathrm{LAT}_{\mathrm{end}} \tag{4}$$

- Lookahead Duration Constraint; The total duration of selected tasks must not exceed the Lookahead Duration (LAD)

$$\sum_{i \in \mathrm{TL}} x_i \cdot D_{t,i} \leq \mathrm{LAD} \tag{5}$$

- Buffers: This justifies the necessary lack of periods needed for selected tasks.

$$t_{e,i} + \mathrm{buffer}_i \leq t_{s,j}, \quad \forall (i \rightarrow j) \tag{6}$$

Constraints CN: This entity manages issues and emerging risks to enhance quicker constraint resolutions by teams and domains for different projects. Sometimes, tasks are usually hard-constrained. However, this component is primarily controlled by responsible persons. Sometimes, the impediment may be forced removed or managed to get the task completed. In the event of force removal or manoeuvres, the system should be updated to ensure the continuous correctness of scheduled tasks.

- Status constraint; The status constrain represents the readiness indicator for each task x_i, we introduce a binary variable R_i:

$$R_i = \begin{cases} 1, & \text{if all constraints for task } x_i \text{ are resolved} \\ 0, & \text{otherwise} \end{cases}$$

- Resource Constraint; To ensure that the selected tasks do not exceed resource limits for materials M and workers W.

$$\sum_{i \in \mathrm{TL}} x_i \cdot m_i \leq M_{\mathrm{available}} \tag{7}$$

$$\sum_{i \in \mathrm{TL}} x_i \cdot w_i \leq W_{\mathrm{available}} \tag{8}$$

- Space Constraint (S): Only tasks with available space for the job are marked ready to be scheduled. i.e available workspace $S_(avail)$ must be sufficient for the selected tasks.

$$\sum_{i \in \mathrm{TL}} x_i \cdot S_i \leq S_{\mathrm{avail}}, \quad \forall t \in [\mathrm{LAT}_{\mathrm{start}}, \mathrm{LAT}_{\mathrm{end}}] \tag{9}$$

Collaboration CB: This entity represents communication and cooperation among teams (planners, quality, design, engineering, supply chain, etc.) and domains to pre-empt and resolve potential conflicts and plan for projects to withstand emerging risks.

For a given project schedule PS, $N_i k$: Notes in LT and CN tables are associated with task x_i and constraint vector k. $CN_i k$: The k-th constraint vector in the constraint table.

The collaborative conflict for the Lookahead task list (TL) is defined as:

$$\mathrm{Pen}_{\mathrm{colab}} = \sum_{i \in \mathrm{TL}} \sum_{s} x_i \cdot (CN) \cdot \alpha(N_{i,k}) \tag{10}$$

Other than constraint CN, Each constraint type (e.g., safety, design, quality) is explicitly modelled as a separate binary

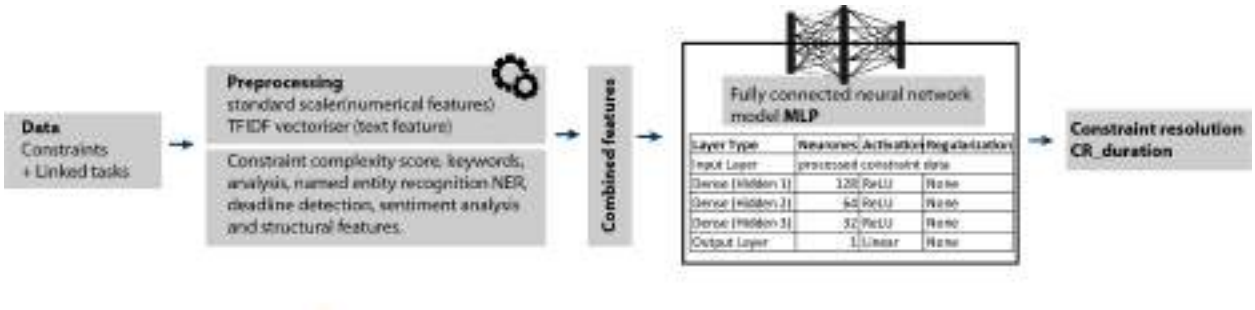


Figure 4. Model Architecture - Fully connected neural network

variable $Sf_{i,k}$ safety for task x_i:

$$Sf_{i,k} = \begin{cases} 1, & \text{if constraint of type } k \text{ for task } i \text{ is resolved} \\ 0, & \text{otherwise} \end{cases}$$

Dynamic Updates

At each time change t of a project, there are consequent changes in the amount of work completed, PPC, constraints, and communication, as these entities are not static but usually evolving. Based on this, the updated Lookahead plan must evolve. Hence, we perform rolling updates for every new iteration as the final constraint to the objective function:

$$\mathrm{LAT}_{\mathrm{start}}^{(t+1)} = \mathrm{LAT}_{\mathrm{end}}^{(t)} + \Delta T \tag{11}$$

Recompute lookahead plan task list TL dynamically. TL accuracy lies in the accuracy of the data that is fed into the process.

$$TL_{accuracy} \simeq \begin{array}{ll} Pr(y_1)=1 & \text{if upcoming tasks;} \\ CN_{x_i}=0 & \text{no constraint exist;} \\ Pr(x_n)=1 & \text{if previous tasks;} \end{array}$$

where x_n are work tasks x_i before specific lookahead interval and y_1 are WTs within the lookahead or short-term interval. The performance indication will then be the range of deviation of actuals from planned, estimated or predicted values.

5 Implementation

As illustrated in Figure 5, advanced deep learning techniques, including Table Parsing (TaPas), deep implicit layers, self-querying mechanisms, and large language model (LLM) frameworks such as Retrieval-Augmented Generation (RAG), represent pivotal architectures for enabling the digital integration and automation of the Last Planner System (LPS) [28, 29, 30, 31].

5.1 Deep learning for ALPS

Deep learning DL, a subset of machine learning, utilises advanced architectures such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to process and generalise complex data with high accuracy [32]. These models, known for scalability and pattern recognition, are well-suited for enhancing project control and scheduling within frameworks like the Last Planner System as indicated in Figure 5. Also, deep implicit layers, which are an ordinary differential equation technique that is able to represent constant changes and understand multi-level reasoning as needed in an automated last planner system (ALPS), would be crucial for automation. It is noted that complex problems such as one represented in Section 4, can be crunched into a single fixed iteration formula, which can then guide LLMs to make feasibility and multilevel reasoning solutions [33].

5.2 RAG for ALPS

Retrieval-augmented generation (RAG) enhances LLMs by integrating data from external databases, enabling more accurate and context-aware outputs. RAG operates through three interconnected modules: the retriever, retriever fusion, and generator[34]. RAG techniques are particularly effective for automating text summarisation, question answering, information extraction, text classification, etc. In the context of ALPS, RAG can significantly enhance processes by enabling constraint classification, automating notifications and communication categorisation, and summarising extensive text data. This integration offers a powerful tool for streamlining complex workflows and improving project management efficiency.



Figure 5. LPS - AI integration



Figure 6. Sample of Constraints; Highlighting keywords, date and NER

5.3 Towards ALPS: Constraint Resolution prediction

The potential for improved augmentation can be significantly enhanced through the integration of predictive strategies in digitised LPS, and the aspects that this integration can impact are enormous. In terms of application, we only looked at ascertaining the resolution duration for constraints, which is an integral aspect of ALPS.

5.3.1 Data

VisiLean, a major cloud-based construction management software, supplied constraint and schedule data for major ongoing projects. Facts and details of the project are withdrawn for privacy reasons. A total of 173 constraints of a large multi-stage developmental project are contained in the dataset, and the data is framed as in Figure 3 and analysed. A sample of the constraint and notes can be seen in Figure 6.

5.3.2 Process

To build a model that predicts constraint resolution duration, we used the constraint dataset. This data was analysed using a cloud-based Python IDE called Collab. First, preliminary data preprocessing was carried out to understand the essential features needed for the model. The unnecessary columns in our dataset were dropped, and constraint resolution duration for all constraints was calculated by determining the duration between the date of constraint resolution and the date of constraint creation. As indicated in Table 2, complexity score, which depicts the difficulty level of a constraint, was determined using keywords, named entity recognition NER, deadline detection, sentiment analysis and structural features, including the length of constraints. After this, both textual and numerical columns were preprocessed. This step regularised the dataset for accurate model training purposes. The final set of features used for training these models includes TDIF vectorisation of combined textual data (Constraint title, description, notes and linked tasks), priority, and Trades. The model architecture is indicated in Figure 4. Afterwards, the created model was evaluated and accessed in the following section. The model is simplified with reduced layers, removed dropouts, and batch normalisation to allow the model to map the complexity of the small datasets fully. Epoch was set to 50, and batch size was reduced to 4 to optimise weight updates and better utilize limited data.

Table 2. Complexity score calculation

Feature	Summary
NER	This recognises named entities via SpaCy
Keywords	keywords are classified based on complexity -critical and mild
Sentiment Analysis	Scoring based on perceived negativity or positivity.
Structure	Words depicting dependencies signifies complexity eg and, but etc.
Length combined	Length of character of the combined textual data
Presence of dates	dependency, date or deadline count
Counts of keywords	Number of keywords, dependencies per constraints

6 Results

The results of the fully connected neural network model are shown in Table 3. The model predicts constraint resolution duration, and the output is explained here. The average duration error or variance for each of the test instances was determined as 23.18 days (MAE), and the standard deviations of each of the error instances in our datasets, indicating effectively the deviation of the actuals from the target variable is 30.12 days (RMSE).

Table 3. Fully connected neural network results

Model	MAE	MSE	RMSE
Neural Network (MLP)	23.18	907.22	30.12

Figure 7 indicates the results plot. These plots, the RMSE and MAE values, indicate that the model overall might not be fully capturing the patterns in the data. Although the target variable is a variable with increased variability and the dataset size is small, the model will keep improving as it is trained on more data. The model can help automate the last planner system as it will assist in pre-empting when constrained jobs can be rescheduled or included in certain lookahead

Table 4. Error comparison using other models

Model	MAE	MSE	RMSE
SVR	21.78	935.93	30.59
LightGBM	23.43	1048.46	32.38
Random Forest	24.06	1108.12	33.29

7 Discussions and Limitations

In further discussion and comparison of the results in Tables 3 and 4, it was discovered that the support vector regression SVR achieves a slightly lower MAE of 21.78 days but exhibits a comparable RMSE of 30.59 days, indicating that SVR has a lower deviation from the ideal when compared with neural network model. On the other hand, the lightGBM and random forest models display higher absolute error values (MAE: 23.43 and 24.06 days), suggesting that they are unable to map the patterns in the data as much as SVR or the Neural network model. This higher error variability is also replicated in the RMSE values, indicating that neural network models perform relatively better. Although SVR is not completely outperformed in terms of MAE by the neural network model, the neural network model maintains a competitive RMSE, signigfiying its ability to generalise despite the small dataset and variability in target

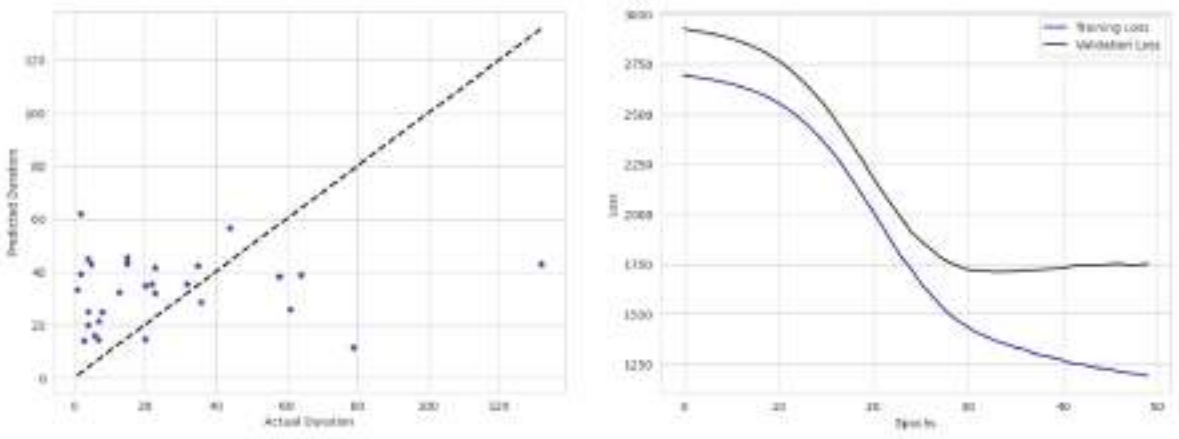


Figure 7. Plot of constraint-resolution-duration actuals and prediction compared with ideal prediction line. Validation loss and loss plot indicates the efficiency at which the model learns.

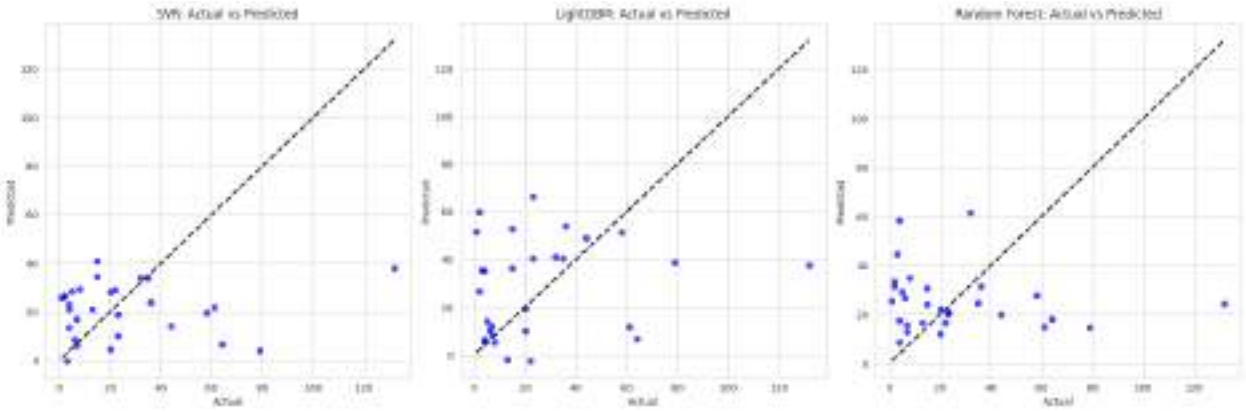


Figure 8. support vector regression SVR, LightGBM, and Randomforest Plot of constraint-resolution-duration actuals and prediction compared with ideal prediction line.

variable. Furthermore, as the dataset expands, the neural network has greater potential for improvement through training and hyper-parameter tuning. The plots in Figures 7a and 8a showcases the prediction of each constraint instance to the ideal with specific improvement in predictions especially when plots of both neural network and SVR are weighed side by side.

While this work highlighted the core aspects of LPS where automation can be applied, the potential for automated constraint management systems, automated collaborative systems, and automated scheduling systems all integrated into one system will increase the efficiency and accuracy of the Last Planner system. To initiate this ambitious research concept, the study conceptualized the system as an optimization problem, which can be reformulated into a fixed iterative equation, such as a gradient descent operation. This equation can subsequently be integrated as a constraint within a large language model (LLM) to facilitate the generation of accurate semantic queries and the determination of feasible solutions for lookahead scheduled tasks. Secondly, an analysis of constraint data of major ongoing projects to determine constraint or conflict resolution time will be handy in quick insight generation and accuracy of Lookahead tasks. However, certain limitations were evident throughout the study.

- The dataset revealed several instances where the creation_date and committed_date occurred after the completion date, presenting a distorted representation of the actual outcomes. This discrepancy primarily arises from the fact that actual data inputs are often recorded not for process improvement or performance evaluation but rather merely to fulfil reporting requirements. Such limitations hinder the model's ability to generalise effectively over the training dataset, as many data inputs exhibited either zero

or negative values for constraint resolution time.
- The available dataset was relatively small in size. In cases involving larger datasets, the model would likely demonstrate improved predictive capability regarding constraint resolution durations, thereby minimising associated errors.
- The textual data utilised in the analysis, which includes fields such as Constraint_title, Description, Notes, and Linked_tasks, exhibited a variety of named entities and structural inconsistencies. These variations complicated the analysis of constraint complexity and hindered the appropriate feature engineering necessary for enhancing model accuracy.

7.1 Recommendations and Implications for Practitioners

While the ideas presented in this research are still evolving, they can be viewed as an augmentation of the Last Planner System (LPS) rather than a complete automation of last planners. This is reflected in the title: Towards Automation of the Last Planner System. The first concept focuses on optimizing the scheduling management system to categorize the most suitable list of tasks for specific look-ahead periods while considering existing constraints and communication dynamics. Although the optimization was evaluated mathematically, the proposed LPS-AI framework suggests employing methods such as Retrieval-Augmented Generation (RAGs), TAPAS, Deep Implicit Layers, Deep Learning, and self-querying to automate the entire Last Planner System.

The second concept involves predicting constraint resolution durations. This predictive model assists in determining the period within which specific constraints will be resolved. As a result, it facilitates the scheduling of optimized look-ahead tasks by providing foresight into constraint resolution timelines. Traditionally, in the absence of such models, planners and foremen have had to replan and reschedule tasks at the beginning of each work period due to their inability to anticipate constraint resolution times, thereby losing valuable productive work hours.

Although the target variable exhibits high variability and the dataset is currently limited, the model is expected to improve as it is trained on more data. By preemptively identifying when constrained tasks can be rescheduled or incorporated into specific look-ahead plans, this model contributes to automating the Last Planner System and enhancing workflow efficiency

7.2 Future Work

Overall, the limitations are pointers to what must be done to achieve either an augmented system or a self-sufficient system that can rely on information at hand to generate accurate lookahead schedules for desired periods. The quality of the generated actual data is very important, and a structured list of possible constraints and tasks should be adhered to. This will help analysis and integration easily. While the problem is dynamic, as highlighted in the definition of the optimisation problem, constrained iterations with implicit and explicit functions would be coded into an LLM as constraints for outputs in a RAG structure to improve and augment the existing LPS. Also, improved research on table parsing (TAPAS) to map and understand complex linked tables with several rows would advance the possibility of augmentation of last planner systems.

8 Conclusion

This research highlights the potential for automating the Last Planner System (LPS) by integrating advanced AI techniques to enhance constraint management, scheduling, and collaboration. Rather than aiming for full automation, the proposed framework augments the Last Planner System by optimizing lookahead scheduling and predicting constraint resolution durations. By framing LPS as an optimization problem, this study explored the use of deep learning, Retrieval-Augmented Generation (RAG), Large Language Models (LLMs), and predictive modeling to improve decision-making and efficiency.

One key contribution of this research is the introduction of a predictive model that estimates constraint resolution durations, thereby enabling more proactive scheduling of lookahead tasks. Traditionally, foremen and planners face challenges in anticipating when constraints will be resolved, leading to frequent replanning and inefficiencies. By leveraging AI-driven predictions, this study lays the foundation for reducing uncertainty and minimizing lost productive hours.

Despite its promise, limitations such as dataset inconsistencies, small sample sizes, and challenges in complex table parsing highlight areas for further development. To overcome these challenges, future research should focus on refining the optimization representation into a constrained iterative framework, embedding implicit and explicit functions into LLM structures for better integration within RAG architectures. Additionally, improving structured data collection processes and enhancing table parsing techniques (e.g., TAPAS) will be critical in augmenting LPS capabilities.

By addressing these gaps, this research paves the way for an AI-augmented Last Planner System that enhances construction workflow efficiency, improves predictability, and ultimately transforms project planning in complex, dynamic environments.

9 Acknowledgments

This study was supported by Bournemouth University and VisiLean under a match-funded scholarship program in the UK. I acknowledge the contributions of my supervisors for their constructive comments and feedback throughout the preparation of this research work.

References

[1] Ajay Kumar Agrawal, Yang Zou, Long Chen, Mohammed Adel Abdelmegid, and Vicente González. Moving toward lean construction through automation of planning and control in last planner system: A systematic literature review. *Developments in the Built Environment*, page 100419, 2024. ISSN 2666-1659.

[2] Camilo I Lagos, Rodrigo F Herrera, Alejandro F Mac Cawley, and Luis F Alarcón. Predicting construction schedule performance with last planner system and machine learning. *Automation in Construction*, 167:105716, 2024. ISSN 0926-5805.

[3] Daniel Heigermoser, Borja Garcia de Soto, Ernest Leslie Sidney Abbott, and David Kim Huat Chua. BIM-based Last Planner System tool for improving construction project management. *Automation in Construction*, 104:246–254, 2019. ISSN 0926-5805.

[4] Glenn Ballard. The Last Planner System of Production Control. *University of Birmingham Research Archive*, 1 2000.

[5] Bhargav Dave, S Boddy, and Lauri Koskela. Visilean: Designing a production management system with lean and BIM. *19th Annual Conference of the International Group for Lean Construction 2011, IGLC 2011*, pages 477–487, 1 2011.

[6] Bhargav Dave. *Developing a construction management system based on lean construction and building information modelling*. PhD thesis, University of Salford, 2013. URL https://www.proquest.com/openview/5a51dcdb6d531ebc4a7a35c9b42869d6/1?cbl=2026366&diss=y&pq-origsite=gscholar.

[7] Mughees Aslam, Zhili Gao, and Gary Smith. Development of innovative integrated last planner system (ILPS). *International Journal of Civil Engineering*, 18:701–715, 2020. ISSN 1735-0522.

[8] Thomas H Davenport, Jeanne G Harris, David W De Long, and Alvin L Jacobson. Data to knowledge to results: building an analytic capability. *California management review*, 43(2): 117–138, 2001. ISSN 0008-1256.

[9] Nowshade Kabir and Elias Carayannis. Big data, tacit knowledge and organizational competitiveness. In *proceedings of the 10th international conference on intellectual capital, Knowledge Management and Organisational Learning: ICICKM*, page 220, 2013.

[10] Olli Seppänen, Glenn Ballard, and Sakari Pesonen. The combination of last planner system and location-based management system. *Lean construction journal*, pages 43–54, 2010.

[11] Roshan Poudel, Borja Garcia de Soto, and Eder Martinez. Last Planner System and Scrum: Comparative analysis and suggestions for adjustments. *Frontiers of engineering management*, 7:359–372, 2020. ISSN 2095-7513.

[12] Ankit Bhatla and Fernanda Leite. Integration framework of bim with the last planner systemtm. In *IGLC 2012-20th Conference of the International Group for Lean Construction*, 2012.

[13] M C Garrido, R Mendes Jr, S Scheer, and T F Campestrini. Using BIM for last planner system: case studies in Brazil. In *Computing in Civil Engineering 2015*, pages 604–611. 2015.

[14] Bhargav Dave, Juho-Pekka Hämäläinen, and Lauri Koskela. Exploring the Recurrent Problems in the Last Planner Implementation on Construction Projects. *Proceedings of the Indian Lean Construction Conference (ILCC 2015)*, 1 2015.

[15] Conrad Boton, Yaya Pitti, Daniel Forgues, and Ivanka Iordanova. Investigating the challenges related to combining BIM and Last Planner System on construction sites. *Frontiers of Engineering Management*, 8:172–182, 2021. ISSN 2095-7513.

[16] Rein Terje Thorstensen, Bo Terje Kalsaas, John Skaar, and Sigmund Jensen. Last planner system innovation efforts on requirements for digital management system. In *Annual conference of the international group for lean construction*, volume 21, 2013.

[17] Mani Poshdar, Vicente A González, Gary M Raftery, Francisco Orozco, and Guillermo G Cabrera-Guerrero. A multi-objective probabilistic-based method to determine optimum allocation of time buffer in construction schedules. *Automation in Construction*, 92:46–58, 2018. ISSN 0926-5805. doi:https://doi.org/10.1016/j.autcon.2018.03.025. URL https://www.sciencedirect.com/science/article/pii/S0926580517300432.

[18] Vicente González and Luis Fernando Alarcón. A methodology for integrated buffer design and management in repetitive construction projects. In *Proceedings of the 17th Annual Conference of the International Group for Lean Construction*, pages 393–406, 2009.

[19] David Jason Gerber, Burçin Becerik-Gerber, and Alex Kunz. BUILDING INFORMATION MODELING AND LEAN CONSTRUCTION: TECHNOLOGY, METHODOLOGY AND ADVANCES FROM PRACTICE. In *IGLC -Enabling Lean with IT*, Haifa, Israel, 7 2010. URL https://api.semanticscholar.org/CorpusID:13917181.

[20] Ophir Rozenfeld, Rafael Sacks, and Yehiel Rosenfeld. 'CHASTE': Construction hazard assessment with spatial and temporal exposure. *Construction Management & Economics*, 27:625–638, 7 2009. doi:10.1080/01446190903002771.

[21] Rafael Sacks, M Treckmann, and O Rozenfeld. Visualization of Work Flow to Support Lean Construction. *Journal of Construction Engineering and Management-asce - J CONSTR ENG MANAGE-ASCE*, 135, 12 2009. doi:10.1061/(ASCE)CO.1943-7862.0000102.

[22] Fouad Amer, Yoonhwa Jung, and Mani Golparvar-Fard. Transformer machine learning language model for auto-alignment of long-term and short-term plans in construction. *Automation in Construction*, 132, 9 2021. doi:10.1016/j.autcon.2021.103929.

[23] Luis F Alarcon and Rodrigo Calderon. A Production Planning Support System for Construction Projects. In *11th Annual Conference of the International Group for Lean Construction*, Virginia, USA, 2003. URL http://iglc.net/Papers/Details/222/pdfhttp://iglc.net/Papers/Details/222.

[24] Hyun Jeong Choo, Iris Tommelein, Glenn Ballard, and Todd Zabelle. WorkPlan: Constraint-Based Database for Work Package Scheduling. *Journal of Construction Engineering and Management-asce - J CONSTR ENG MANAGE-ASCE*, 125, 6 1999. doi:10.1061/(ASCE)0733-9364(1999)125:3(151).

[25] Hyun Jeong Choo and Iris D Tommelein. WorkMovePlan: Database for Distributed Planningand Coordination. In *Iglc*, 2000. URL https://api.semanticscholar.org/CorpusID:55674696.

[26] E Sriprasert and Nashwan Dawood. Multi-constraint information management and visualisation for collaborative planning and control in construction. *Electronic Journal of Information Technology in Construction*, 8:341–366, 10 2003.

[27] Tarek Salama, Ahmad Salah, and Osama Moselhi. Integrating critical chain project management with last planner system for linear scheduling of modular construction. *Construction innovation*, 21(4):525–554, 2021. ISSN 1471-4175.

[28] Weizheng Lu, Jing Zhang, Ju Fan, Zihao Fu, Yueguo Chen, and Xiaoyong Du. Large language model for table processing: A survey. *arXiv preprint arXiv:2402.05121*, 2024.

[29] Jonah Philion and Sanja Fidler. Lift, splat, shoot: Encoding images from arbitrary camera rigs by implicitly unprojecting to 3d. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIV 16*, pages 194–210. Springer, 2020. ISBN 3030585670.

[30] Malte Helmert. The fast downward planning system. *Journal of Artificial Intelligence Research*, 26:191–246, 2006. ISSN 1076-9757.

[31] Weiye Xu, Min Wang, Wengang Zhou, and Houqiang Li. P-RAG: Progressive Retrieval Augmented Generation For Planning on Embodied Everyday Task. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 6969–6978, 2024.

[32] Laith Alzubaidi, Jinglan Zhang, Amjad J Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, J Santamaría, Mohammed A Fadhel, Muthana Al-Amidie, and Laith Farhan. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, 8(1):53, 2021. ISSN 2196-1115. doi:10.1186/s40537-021-00444-8. URL https://doi.org/10.1186/s40537-021-00444-8.

[33] Kolter Zico, Duvenaud David, and Johnson Matt. Deep Implicit Layers - Neural ODEs, Deep Equilibirum Models, and Beyond, 2023. URL https://implicit-layers-tutorial.org/.

[34] Shangyu Wu, Ying Xiong, Yufei Cui, Haolun Wu, Can Chen, Ye Yuan, Lianming Huang, Xue Liu, Tei-Wei Kuo, Nan Guan, and Chun Jason Xue. Retrieval-Augmented Generation for Natural Language Processing: A Survey. 7 2024. URL http://arxiv.org/abs/2407.13193.

EA-YOLOv8: An Enhanced Crack Segmentation Algorithm for Infrastructure Maintenance

Feiyu Chen[1], Jinghui Qin[2]*, Jianyu Yin[3] and Xianfei Yin[4]

[1]Department of Civil and Transportation Engineering, Guangdong University of Technology, Guangzhou, China
[2]Department of Information Engineering, Guangdong University of Technology, Guangzhou, China
[3]Department of Civil and Environmental Engineering, University of Macau, Macau, China
[4]Department of Architecture and Civil Engineering, City University of Hong Kong, Hong Kong, China

flyversechen@gmail.com qinjinghui@gdut.edu.cn dc12714@um.edu.mo xianfyin@cityu.edu.hk

Abstract -

Automated crack detection plays a critical role in infrastructure maintenance and safety assessment. The detection and segmentation of concrete and pavement cracks present substantial technical challenges. These challenges stem from the inherent complexity of crack morphology, environmental lighting variations, and the textural similarities between cracks and their surrounding surfaces. In this paper, we introduce a novel segmentation algorithm named EA-YOLOv8 to address two critical gaps in current crack detection techniques. The first gap involves limited spatial information utilization. To solve this, we integrate the Efficient Local Attention (ELA) mechanism into the YOLOv8 architecture's Bottleneck structure. ELA enhances spatial feature extraction without compromising computational efficiency or channel information. This approach directly tackles the challenge of precise crack localization. The second gap relates to imprecise crack boundary detection. We develop an adaptive region-specific loss (ARS loss) that specifically targets the complexities of crack edge delineation. Traditional loss functions often fail to capture the nuanced boundaries of microscopic cracks. Experimental validation of EA-YOLOv8 was conducted on the Roboflow crack dataset. Through ablation studies and baseline comparisons with YOLOv5, YOLOv8, and the latest YOLOv11, the experimental results show the superior performance of EA-YOLOv8. In terms of results, EA-YOLOv8 improved precision by 10% compared to YOLOv5 and the latest YOLOv11, and increased precision by 5% over the original YOLOv8. These experimental findings validate the theoretical framework and practical applicability of our proposed EA-YOLOv8 in crack detection scenarios.

Keywords -

Crack segmentation; YOLOv8; Efficient local attention; Adaptive region-specific loss

1 Introduction

Infrastructure has experienced extensive environmental erosion and service load increases during recent decades. These factors significantly impact structural safety throughout the infrastructure lifecycle. Concrete and pavement cracks represent fundamental indicators of structural health conditions [1]. Precise and timely crack detection remains essential for infrastructure maintenance and public safety assurance.

*Corresponding author

The advancement of inspection technologies has transformed crack detection methodologies. Manual inspection methods maintain widespread application but exhibit limitations in labor requirements, time consumption, and subjective variations [2]. The integration of optical technologies with deep learning algorithms provides enhanced automated solutions for crack detection [1, 3]. While YOLO networks have shown potential in crack detection, existing methods still struggle with fundamental limitations [4, 5].

The primary challenges in current crack detection techniques include: Crack morphology presents extreme complexity. Cracks exhibit irregular patterns with dramatic variations in width and length, making feature extraction highly difficult [2, 6]. Most existing deep learning models fail to capture the nuanced characteristics of these structural irregularities.Background and crack region imbalance creates significant segmentation challenges. Current algorithms struggle to distinguish subtle crack boundaries from surrounding surface textures. This limitation leads to inaccurate detection, especially in regions with complex surface conditions. Boundary detection precision remains a critical weakness. Existing models typically cannot accurately delineate crack edges, particularly in areas with minimal contrast or complex surface textures.

To address these critical gaps, we developed EA-YOLOv8, an advanced segmentation algorithm with two key innovations. First, we integrated the Efficient Local Attention (ELA) mechanism to improve spatial feature recognition. ELA enables more precise crack localization by dynamically distributing attention weights across different image regions. This approach overcomes the traditional limitations in feature extraction. Second, we designed an Adaptive Region-Specific Loss (ARS loss) function specifically targeting crack boundary detection challenges. This innovative loss function enhances the model's ability to differentiate crack edges from background regions during training. By focusing on the most challenging detection areas, ARS loss significantly improves edge detection precision.

Experimental validation on the Roboflow crack dataset demonstrates the substantial improvements of EA-YOLOv8. Compared to existing methods, our approach increased precision by 10% over YOLOv5 and the latest YOLOv11, and improved performance by 5% compared to the original YOLOv8. These results validate the theoretical contributions and practical utility of our method in infrastructure crack detection.

2 Related Work

Advancements in deep learning has created substantial opportunities in complex detection tasks. Its application in crack detection demonstrates extensive implementation and profound impact. The field has witnessed numerous research achievements that establish the foundation for technological progression.

N.A.M. Yusof et al. [1] implemented convolutional neural network (CNN) methodologies for concrete crack classification. Their research validated the discriminative capabilities of deep learning in crack detection applications. Liu et al. [2] developed an enhanced deep convolutional neural network incorporating deep supervision mechanisms. This approach addressed segmentation accuracy limitations in complex environments and improved fine crack detail capture. Hassan Hosseinzadeh et al. [3, 7] utilized the YOLOv5 framework for pavement crack detection and segmentation. Their methodology achieved precise crack characterization while maintaining computational efficiency, addressing practical requirements in road maintenance applications. Similarly, Yin et al. [7, 8] implemented the YOLOv3 network for automated defect detection in municipal drainage systems. Their system detected six types of defects including cracks, achieving a mean average precision of 85.37%. This work demonstrated the practical application of deep learning in infrastructure maintenance, particularly for sewer pipe inspection where thousands of kilometers of CCTV footage require analysis. Some recent research has focused on minimizing the annotation burden for sewer defect detection [9, 10]. For example, a framework utilizing weakly supervised object localization for automated sewer defect detection was developed by Yin et al. [10], achieving high localization accuracy while requiring minimal annotation effort compared to traditional supervised approaches.

Current deep learning models exhibit specific limitations in crack detection applications. These limitations manifest in edge detection precision and fine-grained segmentation capabilities, particularly when processing irregular crack morphologies and complex background conditions. Crack characteristics, including width variations, morphological complexity, and environmental interactions, present significant challenges for accurate boundary delineation and feature extraction. Addressing annotation challenges in infrastructure inspection, Yin et al. [9] applied Weakly Supervised Object Localization (WSOL) for sewer defect detection using only image-level labels. Their approach achieved mean MaxBoxAccV2 scores of 64.33% with ResNet-50 backbone and established a benchmark for weakly supervised learning in infrastructure applications, eliminating the need for labor-intensive bounding box annotations while maintaining high classification accuracy.

Research efforts focus on architectural optimization strategies to address these technical limitations. Xiang et al. [7, 11] integrated Transformer modules into YOLOv5 architecture. The self-attention mechanism enhanced long-range dependency capture and contextual understanding within crack regions, expanding model detection capabilities. Dong et al. [4, 7] incorporated the Convolutional Block Attention Module (CBAM) into YOLOv5. This modification enabled adaptive attention distribution across feature dimensions, improving crack feature focus while reducing background interference. Liu [1] enhanced the DeepLabV3+ network through Dual Attention Module integration and HDD loss implementation. These modifications strengthened feature emphasis and optimized edge detection performance, advancing crack detection methodologies.

Contemporary research demonstrates that architectural enhancement through specialized modules and loss functions represents an effective approach for performance improvement. This study implements the Efficient Local Attention (ELA) mechanism [12] and Adaptive Region-Specific Loss (ARS loss) [11] within the YOLOv8 framework. Systematic ablation experiments validate the model's effectiveness. The research provides enhanced solutions for crack detection applications, contributing to technological advancement in this domain.

3 EA-YOLOv8

3.1 Overview

To improve the performance of crack segmentation, we introduce two architectural improvements to YOLOv8. The first modification integrates the Efficient Local Attention (ELA) mechanism into the C2F module. This integration creates the C2ELA module for crack feature extraction enhancement. The second modification implements Adaptive Region-Specific Loss (ARS loss) within the segmentation component. ARS loss operates in conjunction with BCE Loss for crack boundary localization. These improvements enhance the segmentation performance of the base architecture. Figure 1 depicts the architecture of the proposed EA-YOLOv8 framework.

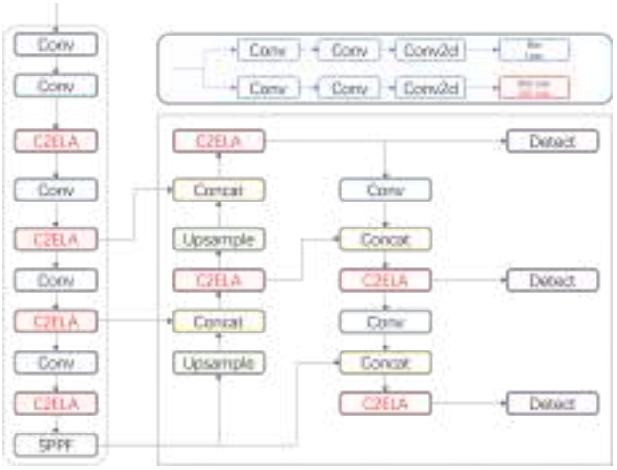


Figure 1. The network architecture of EA-YOLOv8



Figure 2. The structure of Efficient Local Attention.

3.2 Efficient Local Attention (ELA)

We introduce the operational mechanisms of ELA (Efficient Local Attention) for crack recognition in this section. Figure 2 illustrates the structural architecture of ELA. ELA utilizes strip pooling operations for coordinate information extraction. The pooling process extracts horizontal and vertical features from input feature maps. This directional processing method enables orientation-specific pattern recognition. Crack features demonstrate distinct directional properties. The coordinate extraction mechanism captures feature variations along crack extension directions. This process establishes detailed morphological analysis capabilities.

The architecture deploys 1D convolution for bidirectional positional enhancement. Unlike standard 2D convolutions that process pixels in all directions simultaneously, 1D convolution works like a signal filter that processes information along a single dimension (either horizontal or vertical). This approach is similar to how radar systems scan in specific directions to detect objects. 1D convolution has demonstrated computational efficiency advantages over 2D convolution operations, requiring fewer calculations while still capturing essential directional information. This convolution method integrates directional features and maintains spatial continuity information, helping the model "connect the dots" when detecting elongated crack structures.

ELA incorporates Group Normalization (GN) [13] for feature distribution stabilization. In simple terms, GN works like adjusting the contrast of different parts of an image separately to ensure all important details remain visible regardless of lighting conditions. Environmental factors in crack images introduce feature value variations. These factors include illumination conditions and background complexities. GN [13] executes group-based feature normalization procedures. This normalization method reduces environmental interference effects. The stabilized feature distributions support consistent pattern recognition during model training.

The integration of ELA into YOLOv8's architecture is implemented by replacing the C2f modules with our proposed C2ELA modules. The bottleneck structure in YOLOv8's C2f module performs feature extraction operations through convolutional processing and residual connections. Our implementation embeds the ELA mechanism within this Bottleneck structure according to the following steps:

- We initialize the model with pretrained YOLOv8 weights from Ultralytics.
- The bottleneck components within C2f are identified and modified to include ELA.
- Each bottleneck module processes the input through a 1×1 convolution to reduce channel dimensions.
- The ELA placement occurs after these convolutional layers but before the residual connection.
- The output from ELA is then passed through another 1×1 convolution to restore channel dimensions.
- Finally, a residual connection adds the original input to the processed features.

This architectural arrangement optimizes feature processing capabilities for crack detection. Figure3 presents the C2ELA module configuration. The structural integration combines ELA and Bottleneck functionalities, enhancing the model's ability to capture the distinctive direc-

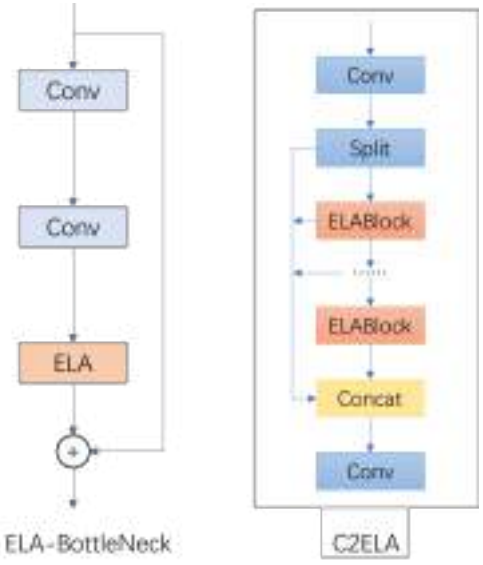


Figure 3. The structure of C2ELA.

tional patterns of cracks while maintaining computational efficiency.

3.3 Adaptive Region-Specific (ARS) Loss

We propose the Adaptive Region-Sensitive Loss (ARS-Loss) mechanism, which employs Tversky Loss [14] for sub-region computation. This approach surpasses traditional Dice Loss limitations through parameters α and β, which regulate false positive (FP) and false negative (FN) penalties. The formulation is expressed as:

$$L_{Tversky} = 1 - \frac{\sum_{i \in V} p_i \cdot g_i + \varepsilon}{y} \quad (1)$$

$$y = \sum_{i \in V} p_i \cdot g_i + \alpha \sum_{i \in V} p_i (1 - g_i) + \beta \sum_{i \in V} (1 - p_i) \cdot g_i + \varepsilon \quad (2)$$

The equation incorporates specific variables: p_i represents pixel-wise crack probability prediction, g_i denotes corresponding ground truth labels, V specifies the sub-region pixel set, and ε maintains numerical stability.

The core innovation of ARS-Loss lies in its adaptive parameters that evolve according to regional error distributions through:

$$\alpha_{\text{Adaptive}} = A + B \cdot \frac{FP}{FP + FN} \quad (3)$$

$$\beta_{\text{Adaptive}} = A + B \cdot \frac{FN}{FP + FN} \quad (4)$$

The ARS-Loss implementation follows a systematic procedure integrated with the YOLOv8 framework:

1) Forward Pass and Feature Extraction: Input images are processed through the YOLOv8 backbone and neck, generating feature maps at different scales.

2) Segmentation Head Processing: The segmentation head transforms the feature maps to produce prediction probability maps $P \in \mathbb{R}^{H \times W}$, where each pixel value $p_i \in [0, 1]$ represents the probability of belonging to the crack class.

3) Region Subdivision: Both prediction maps and ground truth labels are subdivided into a regular grid structure of 16×16×16 (width × height × channels) sub-regions. This subdivision is implemented as follows:

a. For an input of size $H \times W$, we compute region dimensions $r_h = H/16$ and $r_w = W/16$

b. Each sub-region $R_{m,n}$ covers pixels from indices $(m \cdot r_w, n \cdot r_h)$ to $((m+1) \cdot r_w - 1, (n+1) \cdot r_h - 1)$

c. The grid structure creates 256 sub-regions, enabling localized error analysis

4) Sub-region Error Analysis: For each sub-region $R_{m,n}$, we compute:

a. False Positives (FP): $\sum_{i \in R_{m,n}} p_i \cdot (1 - g_i)$

b. False Negatives (FN): $\sum_{i \in R_{m,n}} (1 - p_i) \cdot g_i$

c. True Positives (TP): $\sum_{i \in R_{m,n}} p_i \cdot g_i$

5) Adaptive Parameter Calculation: For each sub-region $R_{m,n}$, the adaptive parameters are calculated as:

a. $\alpha_{m,n} = 0.3 + 0.7 \cdot \frac{FP_{m,n}}{FP_{m,n} + FN_{m,n}}$

b. $\beta_{m,n} = 0.3 + 0.7 \cdot \frac{FN_{m,n}}{FP_{m,n} + FN_{m,n}}$

6) Sub-region Loss Computation: The Tversky loss for each sub-region is calculated using the region-specific adaptive parameters:

$L_{Tversky} = 1 - \frac{\sum_{i \in V} p_i \cdot g_i + \varepsilon}{y}$

$y = \sum_{i \in V} p_i \cdot g_i + \alpha \sum_{i \in V} p_i (1 - g_i) + \beta \sum_{i \in V} (1 - p_i) \cdot g_i + \varepsilon$

7) Total Loss Aggregation: The overall ARS-Loss is computed as the mean of all sub-region losses:

$L_{ARS} = \frac{1}{256} \sum_{m=0}^{15} \sum_{n=0}^{15} L_{m,n}$

8) Loss Integration with YOLOv8: The ARS-Loss is integrated with the standard YOLOv8 loss components:

$L_{total} = \lambda_{box} \cdot L_{box} + \lambda_{cls} \cdot L_{cls} + \lambda_{ARS} \cdot L_{ARS}$

where $\lambda_{box} = 0.05$, $\lambda_{cls} = 0.5$, and $\lambda_{ARS} = 1.0$ are weighting coefficients determined through ablation studies to balance the different loss components.

9) Backpropagation and Weight Update: During the backward pass, gradients are computed with respect to the total loss. The adaptive nature of ARS-Loss ensures that regions with higher FP receive stronger penalties for false positives (higher α), while regions with higher FN receive stronger penalties for false negatives (higher β). This region-sensitive adaptation guides the model to focus on challenging areas with imbalanced error distributions.

The implementation of ARS-Loss starts with the pre-trained YOLOv8 model (weights from Ultralytics) and modifies the loss computation module while maintaining

the original network architecture. This approach requires minimal changes to the YOLOv8 framework, making it easily applicable as a drop-in enhancement for crack segmentation tasks.

The region-sensitive methodology enhances feature discrimination capacity and crack boundary detection precision by dynamically adjusting penalties based on local error characteristics. This is particularly effective for thin crack structures where traditional global loss functions often struggle to provide sufficient gradient signals.

4 Experiments

4.1 Dataset and Implementation details

The experimental dataset consists of the Roboflow Crack Segmentation Dataset. Roboflow developed this public dataset for traffic and infrastructure safety research applications. The dataset incorporates 4,029 static images from road and wall environments. Each image contains pixel-level annotations. The image collection represents multiple infrastructure conditions and crack characteristics. The dataset supports comprehensive safety analysis tasks.

The research implementation adopts Roboflow's original data organization. The dataset division follows the provided training, testing, and validation partitions. Table 1 specifies the image distribution across these subsets.

Table 1. The dataset splitting of Roboflow Crack Segmentation dataset.

Dataset	Roboflow Crack Seg
Train	3717
Validation	200
Test	112

All experiments are conducted with an NVIDIA GeForce RTX 4070 GPU (12GB VRAM). The model is implemented with PyTorch[1] 2.4.1 framework running on CUDA 12.6. Additional libraries include NumPy 1.26.3 for numerical operations and Ultralytics for YOLOv8 implementation. All experiments are conducted with Python 3.10.14. The model initialization incorporates YOLOv8s-seg pre-trained weights from Ultralytics. The training configuration is concluded in Table 2.

Table 2. Hyperparameters for model training.

Parameter		Parameter	
Epoch	150	Momentum	0.937
Optimizer	SGD	Warmup Epoch	3
lr0	0.01	Warmup Momentum	0.8
lrf	0.01	Warmup Bias lr	0.1

[1]https://pytorch.org/

Table 3. Performances of various models on the Roboflow dataset.

	Precision	Recall	mAP50	F1
YOLOv5	0.75592	0.68675	0.68545	0.71968
YOLOv8	0.80325	0.6506	0.67648	0.718911
YOLOv11	0.75672	**0.69478**	0.65088	0.72442
ARS-YOLOv8	0.80307	0.66265	0.69921	0.726134
ELA-YOLOv8	0.83489	0.6506	0.69087	0.731313
EA-YOLOv8	**0.85669**	0.65462	**0.70297**	**0.742146**

5 Results and Discussion

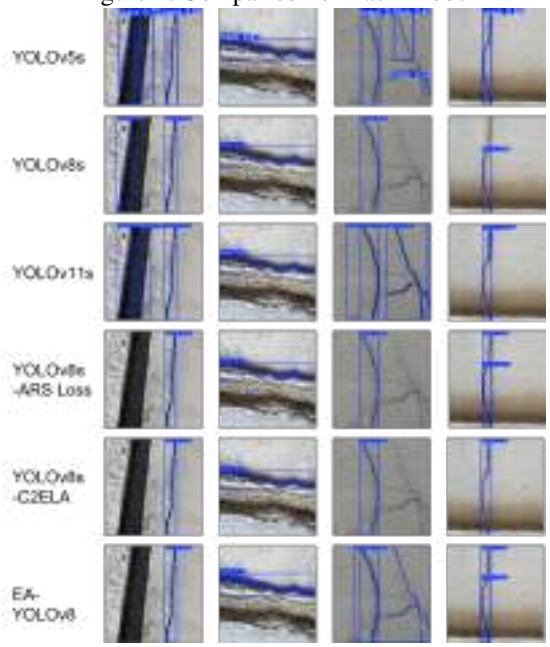
This investigation was executed utilizing the Roboflow dataset corpus. Our methodology encompasses a systematic evaluation between EA-YOLOv8 and preceding YOLO variants: YOLOv5, YOLOv8, YOLOv11, ELA-YOLOv8, and ARS-YOLOv8. The architectural design of ELA-YOLOv8 incorporates the ELA attention mechanism explicitly for feature extraction augmentation. The ARS-YOLOv8 framework implements the ARS loss function specifically for optimization procedures. Statistical metrics were acquired through methodological validation protocols. The empirical results, as documented in Table 3, elucidate the comparative analysis. The novel EA-YOLOv8 architecture, synthesizing the ELA attention mechanism and ARS loss function, manifests enhanced detection efficacy compared to singular-component implementations.

The experimental evaluation compares EA-YOLOv8 with the YOLOv5 baseline model. EA-YOLOv8 exhibits performance enhancements across evaluation metrics. YOLOv5 achieves baseline performance with 0.75592 precision and 0.68675 recall. YOLOv11, despite being the most recent architecture in the YOLO family, achieves 0.75672 precision and 0.69478 recall, showing only marginal improvement over YOLOv5 in crack detection tasks. YOLOv8 and its architectural variants demonstrate precision improvements. These results indicate YOLO architecture progression.

Precision defines the ratio of correctly predicted positive samples (True Positives, TP) among total positive predictions. EA-YOLOv8 achieves 0.85669 precision in crack identification. This result surpasses both YOLOv5 (0.75592), YOLOv11 (0.75672) and YOLOv8 (0.80325) baselines. The visualized results in Figure 4 column one show detection errors in baseline models. EA-YOLOv8 reduces false detections through two mechanisms. The ELA mechanism enhances directional feature extraction capabilities. The ARS mechanism enables regional attention optimization. These components reduce False Positive (FP) predictions.

Recall represents the proportion of actual positive samples correctly predicted as positive (TP). EA-YOLOv8 achieves 0.65462 recall. While this is lower than

Figure 4. Comparison of Each Model



YOLOv11's recall (0.69478), which demonstrates the highest recall among the baseline models, the third column in Figure 4 shows detection limitations in YOLOv8, ELA-YOLOv8, and ARS-YOLOv8. EA-YOLOv8 combines ARS loss regional weight adjustment with ELA feature extraction. This integration improves positive sample detection.

Mean Average Precision (mAP50) evaluates model performance in object detection tasks. EA-YOLOv8 achieves 0.70297 mAP50 ,which significantly outperforms YOLOv11's 0.65088 mAP50. The ELA mechanism extracts discriminative features. The ARS loss focuses on challenging regions. These components enhance detection accuracy across scales.

F1 score provides the harmonic mean of precision and recall. EA-YOLOv8 achieves 0.742146 F1 score , surpassing YOLOv11's 0.724428 F1 score. This metric demonstrates the effectiveness of ELA and ARS loss integration.

This research presents EA-YOLOv8 for crack detection applications. The model integrates ELA attention mechanism and ARS loss function. Ablation experiments on the Roboflow dataset validate model effectiveness. EA-YOLOv8 demonstrates performance improvements over YOLOv5, YOLOv11 and YOLOv8 baselines. The model surpasses ELA-YOLOv8 and ARS-YOLOv8 variants in precision, recall, mAP50, and F1 metrics.

6 Conclusion and Limitation

In this paper, we introduce a novel segmentation algorithm named EA-YOLOv8 to incorporate two technical innovations into the YOLOv8 architecture. First, we deploy the Efficient Local Attention (ELA) mechanism following the final convolutional layer within the Bottleneck structure of the C2f module. ELA can optimize spatial information utilization while preserving channel dimensions and computational efficiency, resulting in improved crack localization capabilities. Second, we develop an adaptive region-specific loss (ARS loss) as an auxiliary loss to specifically target the difficulties in crack boundary delineation, leading to refined edge detection results. Experimental validation of EA-YOLOv8 was conducted on the Roboflow crack dataset. Through comprehensive ablation studies and baseline comparisons with YOLOv5, YOLOv8, and the latest YOLOv11, the experimental results demonstrate the EA-YOLOv8's superior performance. Quantitatively, EA-YOLOv8 exhibits a 10% precision improvement over YOLOv5, a 10% improvement over YOLOv11 and a 5% performance gain compared to the original YOLOv8. While YOLOv11 demonstrates slightly higher recall than EA-YOLOv8, our model achieves superior precision, mAP50, and F1 scores, indicating better overall performance for crack detection tasks. These experimental results validate both the theoretical framework and practical applicability of our proposed EA-YOLOv8 in crack detection scenarios.

Research limitations require further investigation. EA-YOLOv8 recall performance remains below ARS-YOLOv8 levels. This indicates potential feature extraction misalignment between ELA and ARS loss components. ARS loss weight adjustment mechanisms may experience constraints from ELA-generated features. These interactions affect regional learning processes. Future research directions include attention mechanism compatibility analysis with ARS loss. Additional investigation of YOLOv8 detection head, feature pyramid, and SPPF module modifications may enhance crack segmentation performance. Furthermore, this study utilized only a single dataset for training and testing. Future work should incorporate multiple diverse datasets for cross-validation to better verify the model's reliability and generalizability across different crack detection scenarios.

References

[1] NAM Yusof, MK Osman, MHM Noor, A Ibrahim, NM Tahir, and NM Yusof. "Crack detection and classification in asphalt pavement images using deep convolution neural network". In: *2018 8th IEEE international conference on control system, computing and engineering (ICCSCE)*. IEEE. 2018, pp. 227–232.

[2] Yahui Liu, Jian Yao, Xiaohu Lu, Renping Xie, and Li Li. "DeepCrack: A deep hierarchical feature

learning architecture for crack segmentation". In: *Neurocomputing* 338 (2019), pp. 139–153.

[3] Hassan Hosseinzadeh, Ali Ghiami Bajgirani, and Mohadeseh Delavarian. "Detection and classification of asphalt pavement cracks using YOLOv5". In: *Journal of Modeling in Engineering* 21.72 (2023), pp. 175–193.

[4] Zhaohui Dong, Guijie Zhu, Zhun Fan, Jiacheng Liu, Huanlin Li, Yuwei Cai, Huaxing Huang, Ze Shi, Weibo Ning, and Liu Wang. "Automatic Pavement Crack Detection Based on YOLOv5-AH". In: *2022 12th International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER)*. IEEE. 2022, pp. 426–431.

[5] Xuezhi Xiang, Zhiyuan Wang, and Yulong Qiao. "An improved YOLOv5 crack detection method combined with transformer". In: *IEEE Sensors Journal* 22.14 (2022), pp. 14328–14335.

[6] Jie Liu. "Road Crack Detection Using HDD LOSS and Dual Attention Module with DeepLabv3+". In: *2023 3rd International Conference on Digital Society and Intelligent Systems (DSInS)*. IEEE. 2023, pp. 148–152.

[7] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. "YOLO by Ultralytics, January 2023". In: *URL https://github. com/ultralytics/ultralytics* 2 ().

[8] Xianfei Yin, Yuan Chen, Ahmed Bouferguene, Hamid Zaman, Mohamed Al-Hussein, and Luke Kurach. "A deep learning-based framework for an automated defect detection system for sewer pipes". In: *Automation in construction* 109 (2020), p. 102967.

[9] Jianyu Yin, Xianfei Yin, Yifeng Sun, and Mi Pan. "Bridging the Annotation Gap: Innovating Sewer Defects Detection with Weakly Supervised Object Localization". In: *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*. Vol. 41. IAARC Publications. 2024, pp. 669–674.

[10] Jianyu Yin, Xianfei Yin, Mi Pan, and Long Li. "Scalable and transparent automated sewer defect detection using weakly supervised object localization". In: *Automation in Construction* 174 (2025), p. 106152.

[11] Yizheng Chen, Lequan Yu, Jen-Yeu Wang, Neil Panjwani, Jean-Pierre Obeid, Wu Liu, Lianli Liu, Nataliya Kovalchuk, Michael Francis Gensheimer, Lucas Kas Vitzthum, et al. "Adaptive region-specific loss for improved medical image segmentation". In: *IEEE transactions on pattern analysis and machine intelligence* 45.11 (2023), pp. 13408–13421.

[12] Wei Xu and Yi Wan. "ELA: Efficient local attention for deep convolutional neural networks". In: *arXiv preprint arXiv:2403.01123* (2024).

[13] Yuxin Wu and Kaiming He. "Group normalization (2018)". In: *URL https://arxiv. org/abs* (1803).

[14] Seyed Sadegh Mohseni Salehi, Deniz Erdogmus, and Ali Gholipour. "Tversky loss function for image segmentation using 3D fully convolutional deep networks". In: *International workshop on machine learning in medical imaging*. Springer. 2017, pp. 379–387.

Process Coordination and Automation in Modular Construction Supply Chain Using Smart Contract and RPA

Ningshuang Zeng[1], Xuling Ye[2], Shiqi Chen[1], Yan Liu[3], Markus König[2], and Qiming Li[1]

[1]School of Civil Engineering, Southeast University, Nanjing, China
[2]Department of Civil and Environmental Engineering, Ruhr-University Bochum, Bochum, Germany
[3]School of Management and Engineering, Nanjing University, Nanjing, China
ningshuang_zeng@seu.edu.cn, xuling.ye@ruhr-uni-bochum.de, shiqi_chen_@seu.edu.cn, ly@nju.edu.cn, koenig@inf.bi.rub.de, njlqming@163.com

Abstract –

The integration of off-site manufacturing with on-site construction in the Modular Construction Supply Chain (MCSC) presents significant challenges due to the differences in production methods and geographical distances. While various technologies have transformed construction practices, substantial improvements are still needed in on- and off-site coordination to mitigate delays, ensure quality, and lower costs. Blockchain-enabled smart contracts offer a promising workflow engine for enhancing process-oriented MCSC coordination. Additionally, Robotic Process Automation (RPA) holds the potential to further refine the modules within MCSC processes and automate tasks. This paper, therefore, aims to develop a new workflow engine that combines blockchain-enabled smart contracts, RPA, and related visualization technologies to address the coordination challenges inherent in MCSC. Demand-driven process modeling and smart contracts conversion with RPA actions are conducted with the MCSC coordination unit configuration, and the initial system is developed and validated in a real-world case study. Industrial requirements and feedback are collected and analyzed via questionnaires and interviews. Finally, practical implications are discussed, and the integration of RPA and AI is identified as a promising exploration for future work in MCSC coordination.

Keywords –

Smart Contract; RPA; Modular Construction; Supply Chain Coordination

1 Introduction

In dense urban environments such as Asian regions, the industrialization of construction has been significantly influenced by factors e.g., rapid urbanization and increasing sustainability demands. This context has spurred the fast development of modular construction (MC). The merging of off-site manufacturing with on-site construction in the modular construction supply chain (MCSC) presents challenges due to different production approaches and geographical separation [1,2]. These challenges include amplified delays, disjointed quality control systems, and discrepancies in accountability [3,4]. Numerous studies aim to enhance coordination within the construction supply chain at a macro level through strategies such as supplier alliance, organizational partnership, and value co-creation [5,6].

However, the effective implementation of these macro-level strategies requires a solid material and process foundation. Establishing a domain-spanning production flow and developing effective on-site and off-site coordination mechanisms are essential prerequisites for holistically improving the performance of the modular construction supply chain. A series of technological advancements provide a foundation for better process coordination in MCSC.

Business Process Management (BPM) and its process modeling standards, e.g., Business Process Model and Notation (BPMN) and Business Process Execution Language (BPEL), have been widely applied to construction information management with BIM (Building Information Modeling) models [7,8]. Recently, blockchain-enabled smart contracts have played a role in executing the process models in a traceable and semi-automated manner [9]. Meanwhile, the development of Robotic Process Automation (RPA) provides a possibility to further detail the modules within the MCSC processes and automate tasks. It comprises software agents that mimic the manual path taken by a human through integrated computer applications when performing certain tasks in a business process [10]. Currently, AI-enhanced RPA has been explored, and it shows the potential to elevate BPM from semi-automation to full automation [11,12].

Therefore, this paper aims to develop a new workflow

engine integrating blockchain-enabled smart contracts, RPA and related visualization technologies to address the coordination issues existing in MCSC. The demand analysis is conducted at the beginning with a series of industrial engagement activities. The process models and the triggering and driving mechanisms are further developed and integrated into MCSC coordination multi-terminal devices. Finally, a real-world case study is carried out to validate the proposed methods.

2 Literature Review

2.1 BPM and BPA in Construction

Business Process Management (BPM) addresses the topic of process support from a broader perspective by integrating methods of analysis (such as simulation, verification, and process mining) and connecting processes to business and societal aspects [13]. Different process languages have been gradually developed, e.g., Petri nets, Workflow nets, YAWL, and BPMN.

In construction, the supply chain process has been early modeled and decomposed from conceptual representation to process element specification using BPMN and BPEL [14]. Recently, the integration of blockchain-enabled smart contracts and workflow engines has been newly explored and developed for Business Process Automation (BPA) [9,15]. Construction supply chain coordination and payment automation were frequently taken as use cases and application examples [16].

Robotic Process Automation (RPA) is an emerging automation technology in BPA that uses robots to partially or fully automate simple or complex repetitive tasks [17]. These robots, for example, process transactions, manipulate data, and communicate with other digital systems or agents [11]. Therefore, RPA and smart contract-enabled workflow engines can work together in a complementary way when deploying and monitoring automated processes. RPA automates a specific set of tasks, while smart contract-based workflow engines automate the flow among tasks and rule-based decision-making in each logical branch [9].

2.2 MCSC Coordination

The development of information technologies to improve visibility, enable coordination, and promote performance through information sharing in the construction supply chain is a continuously explored topic. The applications of blockchain-enabled smart contracts in conventional construction supply chain management mainly concern aspects including procurement, inventory management, traceability, sustainability, cross-border logistics, etc. [18].

In the general construction supply chain research field, regarding boundary-spanning interdependence, the off-site and on-site domains and domain-spanning processes and activities underneath have been analyzed [3]. In MCSC, the domain-spanning effect shows great importance. The typical MCSC process-oriented scenarios, including production, logistics, and business processes, have been identified as critical application areas [19]. The use of blockchain-enabled smart contracts and automation technologies as an MCSC coordination enabler is still in the early phase. According to a systematic review, currently, only about 18% of studies in modular construction practically implemented digital tools and applications [20]. It also indicated that the on-site and off-site fragmentation is a critical issue for MCSC and requires more digital solutions [20].

3 Methodology

3.1 Top-level Design for MCSC Process Collaboration and Automation

Smart contract generation necessitates accurate process models (e.g., in the BPMN format). Establishing such process models requires clear specifications of each participant's responsibilities and requirements, combined with BIM models, plans, and real-time site information. In different scenarios, the process participants and task complexity may vary. The MCSC process encompasses complex tasks and many participants, including the contractor, client, designer, and suppliers. Process splitting and subprocess modeling can be performed to simplify highly complex processes.

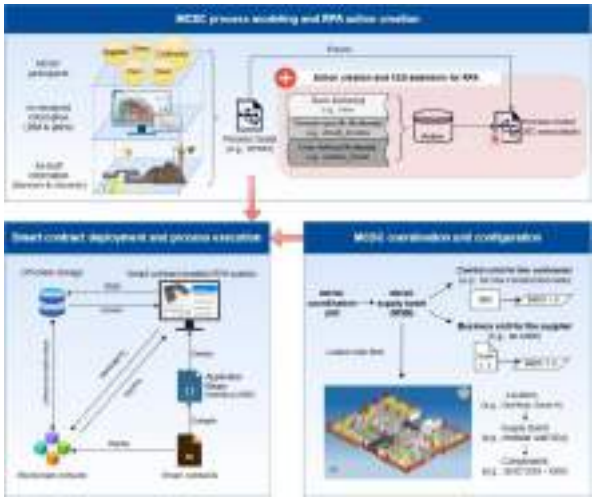


Figure 1. The framework of MCSC process coordination and automation

Because the initiation of tasks in the MCSC process is affected not only by the participants but also by as-

designed and as-built information, process modeling and automation can be conducted by combining BIM and IoT applications. Figure 1 shows the details of the top-level design of the MCSC coordination framework.

There are three key aspects that form the foundation of MCSC coordination: 1) demand-driven process modeling and smart contracts conversion with RPA actions; 2) MCSC coordination unit configuration from a managerial perspective; and 3) blockchain-enabled smart contracts deployment and MCSC process automation.

3.2 MCSC Process Modeling and Conversion

The industrial opinions were collected from a survey with 146 questionnaires and a series of interviews with practitioners involved in MCSC. Regarding the functionality of smart contract and RPA-enabled coordination systems or tools, the majority of respondents recognize that these tools ensure the authenticity of construction supply chain data (with 94% of respondents assigning high scores), traceability (with 81% of respondents assigning high scores), and automation capability (with 96% of respondents assigning high scores).

In the demand analysis of MCSC coordination, the majority of respondents identified the following key needs: preventing supply delays, interacting with suppliers to control progress and costs, obtaining quick responses from suppliers, planning and controlling inventory based on production activities, implementing more automated supply management systems, and ensuring traceability.

Based on industrial requirements, this study modeled the MCSC process using precast components as an example, as shown in Figure 2.

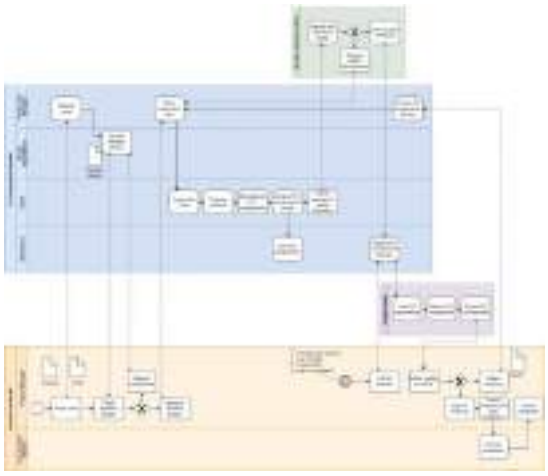


Figure 2. MCSC process model example

In this process model, all essential schedules and quality control gateways are set, and the information flow among the contractor, precast component supplier, logistics service provider, and quality inspection unit are modeled.

Based on the process model, blockchain-enabled smart contracts can be generated using our previously developed generation method [9], as shown in Figure 3. Smart contracts work as the global workflow engine for process execution and flow automation, which can provide traceable reports of the execution records.

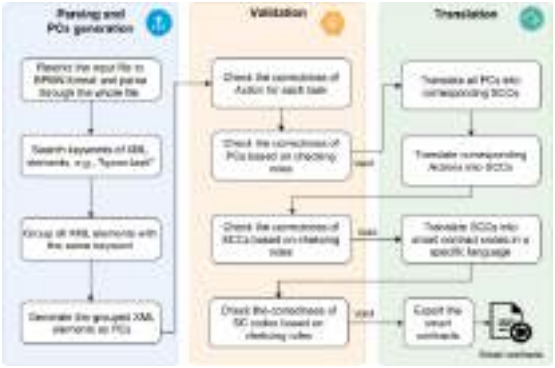


Figure 3. Smart contract generation method (permission obtained)

Meanwhile, a series of RPA actions were defined for internal task automation. To conform to the context of MCSC, RPA actions are defined in three levels: general (basic level for general purpose), domain-specific (e.g., logistics, payment, warehouse and construction), and user-defined (customized level under the requirements of a specific user).

Some manual actions can be replaced with fully or semi-automated RPA actions using smart contracts and blockchain (e.g., a “pay” RPA action), where the smart contracts are used as robots to realize RPA. Other manual actions, which cannot be simply executed via smart contracts (e.g., “check progress” action or “tracking” action), can be transformed into RPA actions by further integrating other devices or agents (e.g., sensors).

3.3 MCSC Coordination Configuration

MCSC Supply Batch (MSB) is defined as the smallest element within the coordination framework, which helps facilitate interactions and processes across different stages and parties in the supply chain, as shown in Figure 4. The attributes of MSB can be generated from predefined templates and integrated IFC properties from the BIM model.

The configuration of MCSC coordination with the basic unit, i.e., MSB, is shown in Figure 4. The defining and planning of MSB are based on project decomposition and a location-based perspective. A construction project can be decomposed from division work, sub-division work, and sub-item work, into construction tasks, and the MSB is linked to a construction zone, which contains a

series of construction tasks following the sequence. The location-based perspective [21] is for better collaboration to provide the MCSC participants with a BIM model view. This project decomposition and a location-based perspective is mainly for the construction and project management side, e.g., the contractor, client, and consultant.

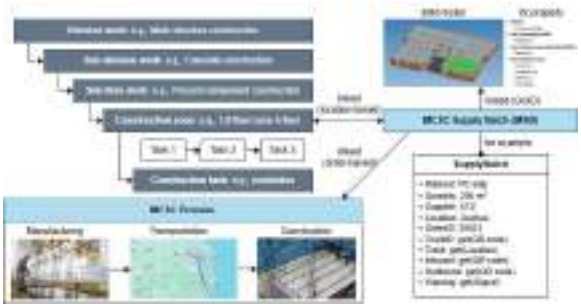


Figure 4. The basic unit of MCSC coordination

Another essential perspective is the order-based perspective, which is mainly for the supply side, e.g., the suppliers and logistics service providers. The order can be generated automatically after the contractor sets up the MSB in the system. The supplier can prepare the delivery and related documents according to an MSB order or link the MSB order to an existing purchase contract. All automatically collected data (e.g., sensors), semi-automatically collected data (e.g., QR code), and manually input data (e.g., mobile device or computer) can be linked to the MSB unit. Therefore, the life-cycle process of an MSB from off-site to on-site can be easily tracked and monitored. With inbound and outbound functions, the as-built information of MSB can also be tracked and shown in the BIM model view.

3.4 MCSC Process Automation

The MCSC process can be executed through semi-automatic operations by authorized personnel and predefined automatic triggers. Figure 5 shows the MCSC coordination process execution with RPA actions.

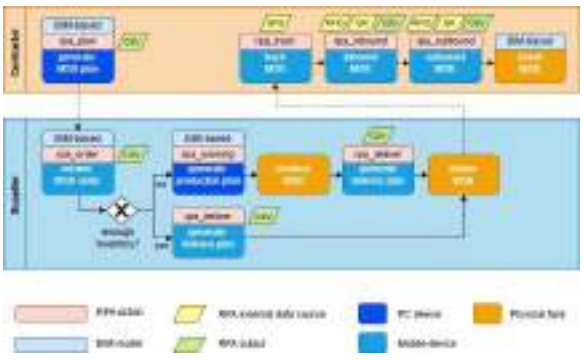


Figure 5. MCSC coordination process execution with RPA actions

In the initiation stage, an MSB plan can be generated with BIM models (components selection) and schedules, including the MSB name, construction zone, and required date. The quantity and associated components can be extracted from the IFC properties. The generation process can be semi-automatically executed using the rpa_plan action. For example, a CSV document of the MSB plan can be automatically created with manual adjustments on quantity unit and required date.

When the supplier receives the MSB plan, the rpa_order action will automatically convert it into an order so that it can be better understood from a business perspective, which also can be a part of a purchase contract.

When there is enough inventory, the supplier can confirm the MSB order and provide a delivery plan with a planned arrival date and other essential information including related staff account information. The rpa_deliver can automatically create a CSV document and send it back to the contractor. This situation is common in practice when a series of MSBs belong to one purchase contract. However, when the MSB is an ETO or MTO-type material or equipment, the production process is required, and the rpa_warning action will calculate if there will be any potential production delay according to real-time data.

After the supplier delivers the MSB, the rpa_track action can update the transportation location automatically, integrating GPS data, while the rpa_warning action will also calculate if there will be any potential transportation delay.

When the MSB arrives, the rpa_inbound action will automatically generate an inbound receipt according to previous WSB data and related staff account information to better assign responsibility. When the MSB is taken for use, the rpa_outbound action will automatically generate an outbound record and assign as-built information, such as construction zone, material requisitioner and outbound clerk.

All manual operations can be performed using mobile devices, except MSB planning and production planning, which are usually integrated into existing PC systems.

The overall workflow is deployed on the blockchain network, and all tasks and gateway decisions will be recorded into transactions to ensure traceability. The automation mechanism of blockchain-enabled smart contracts has been introduced in our previous study [9]. Meanwhile, the status of internal task RPA actions can also be recorded and ensured via the blockchain network.

4 Implementation

This study applied the initially developed system to a project in Nanjing South New Town in China, one of the smart construction pilot projects in Nanjing city,

covering commercial, office and residential areas, to test the proposed blockchain-enabled smart contracts and RPA integrated method.

The demo system of the blockchain-enabled smart contracts for precast components supply coordination is shown in Figure 6.



Figure 6. The demo system of the precast components supply coordination

Based on the further requirements from the project managers and communication with suppliers, the first version of the MCSC system has integrated the MSB concept and RPA actions. In this project, the MCSC coordination is not only about the precast components but also includes the prefabricated decoration components, for example, wood flooring materials and decorative wall aluminum plates. An example of the user interface of this system translated into English is shown in Figure 7.



Figure 7. Example interface of MCSC coordination system in case project

In this case, an example of MSB is "Building B Zone A 3rd floor slab", of which quantity amount is 200 m^2, and the required arrival date is 2024-09-01. This MSB was assigned to the process model and triggered by predefined conditions and manual decision input in the process model, which was displayed using the BPMN standard in the interface, while on the backend, it was deployed on the blockchain in the form of smart contracts.

In this project, the B/S system was mainly applied for the main contractor to plan, generate, and track the MSBs. The planning of MSBs is the most difficult step since it guides all the following data collection, tracking and warning. Therefore, in this project, the MSBs were defined and planned by a small group of one project manager, one construction manager, and one researcher. The supplier only used mobile devices to receive MSB orders and deliver MSB products. All the logistics processes utilized GPS devices for tracking, and the inbound and outbound processes used QR codes, which are attached to each MSB and scanned by the on-site staff and construction workers' phones.

The tools and system provided solutions to issues such as on- and off-site process fragmentation, information silos, data redundancy, untimely logistics tracking and untraceable.

During this project execution, several issues arise, such as a lack of understanding of new technologies, especially among material suppliers with low levels of information technology adoption. Distinguished by the abstract, on-chain coding mechanisms, smart contracts differ fundamentally from more visually intuitive technologies like BIM or Virtual Reality (VR). This often requires significant effort and time for communication and system adjustments. This is one of the key reasons for choosing mobile devices on the supplier side. Most experienced project managers have no problems with the B/S system and can easily deal with BIM elements. Meanwhile, most MCSC participants are not familiar with BPMN tools, and sometimes, additional UI design is required to simplify the process models for better coordination.

5 Conclusion

Given the practical issues of the modular construction supply chain and industrial requirements, this study proposed a blockchain-enabled smart contract and RPA-integrated method to improve process coordination. Three key steps are elaborated: 1) modeling processes driven by demand and converting them into smart contracts with RPA actions; 2) configuring the MCSC coordination unit from a managerial viewpoint; and 3) deploying blockchain-enabled smart contracts and automating MCSC processes. A real-world project validation provided positive results from both technical and managerial perspectives.

Further work can be improved in several aspects. First, the process models and RPA actions can be decomposed into process pattern base and RPA action base for more flexible assembly for various scenarios. Second, the deployment and execution of the smart contract can be optimized by adopting advanced blockchain or other encrypted database technology. Third, practical training and tool usability are essential for better MCSC

coordination practices, and the design of systems based on unrealistic assumptions should be avoided. Additionally, the integration of RPA and AI will be a promising research topic in future MCSC coordination.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant Nos. 72301068 and 72201125), the China Postdoctoral Science Foundation funded project (No. 2023M740592).

References

[1] H. Arshad, T. Zayed, Critical influencing factors of supply chain management for modular integrated construction, Automation in Construction 144 (2022) 104612.

[2] R. Masood, K. Roy, V.A. González, J.B. Lim, A.R. Nasir, A supply chain perspective of prefabricated housebuilding diffusion in New Zealand, Engineering, Construction and Architectural Management 31 (2024) 2697–2720.

[3] N. Zeng, Y. Liu, M. König, 4D BIM-Enabled Look-Ahead Scheduling for Early Warning of Off-Site Supply Chain Disruptions, Journal of Construction Engineering and Management 149 (2023) 04022154.

[4] W. Pan, D. Parker, M. Pan, Problematic Interfaces and Prevention Strategies in Modular Construction, Journal of Management in Engineering 39 (2023) 05023001.

[5] Y. Liu, A. van Marrewijk, E.-J. Houwing, M. Hertogh, The co-creation of values-in-use at the front end of infrastructure development programs, International Journal of Project Management 37 (2019) 684–695.

[6] E. Papadonikolaki, H. Wamelink, Inter- and intra-organizational conditions for supply chain integration with BIM, Building Research & Information 45 (2017) 649–664.

[7] P. Hagedorn, M. König, BPMN-related ontology for modeling the construction information delivery of linked building data., in: LDAC, 2021: pp. 91–102.

[8] M. Häußler, S. Esser, A. Borrmann, Code compliance checking of railway designs by integrating BIM, BPMN and DMN, Automation in Construction 121 (2021) 103427.

[9] X. Ye, N. Zeng, X. Tao, D. Han, M. König, Smart Contract Generation and Visualization for Construction Business Process Collaboration and Automation: Upgraded Workflow Engine, Journal of Computing in Civil Engineering. 38 (2024) 04024030.

[10] S. Moreira, H.S. Mamede, A. Santos, Process automation using RPA–a literature review, Procedia Computer Science 219 (2023) 244–254.

[11] S. Afrin, S. Roksana, R. Akram, AI-Enhanced Robotic Process Automation: A Review of Intelligent Automation Innovations, IEEE Access (2024).

[12] J.X. Zhan, Z. Ling, Z. Xu, L. Guo, S. Zhuang, Driving efficiency and risk management in finance through AI and RPA, Journal of Advanced Computing Systems 4 (2024) 1–9.

[13] H.A. Reijers, Business Process Management: The evolution of a discipline, Computers in Industry 126 (2021) 103404.

[14] J.C.P. Cheng, K.H. Law, H. Bjornsson, A. Jones, R.D. Sriram, Modeling and monitoring of construction supply chains, Advanced Engineering Informatics 24 (2010) 435–455.

[15] D. Mertens, J. Kim, J. Xu, E. Kim, C. Lee, Smart flow: a provenance-supported smart contract workflow architecture, Cluster Comput 27 (2024) 8173–8187.

[16] K. Sigalov, X. Ye, M. König, P. Hagedorn, F. Blum, B. Severin, M. Hettmer, P. Hückinghaus, J. Wölkerling, D. Groß, Automated payment and contract management in the construction industry by integrating building information modeling and blockchain-based smart contracts, Applied Sciences 11 (2021) 7653.

[17] S. Agostinelli, T. Hohenadl, A. Marrella, A. Martínez-Rojas, SmartRPA: Generating software robots from user interface logs, SoftwareX 29 (2025) 101995.

[18] A. Tezel, P. Febrero, E. Papadonikolaki, I. Yitmen, Insights into Blockchain Implementation in Construction: Models for Supply Chain Management, Journal of Management in Engineering 37 (2021) 04021038.

[19] Q. Chen, D.M. Hall, B.T. Adey, C.T. Haas, Identifying enablers for coordination across construction supply chain processes: a systematic literature review, Engineering, Construction and Architectural Management 28 (2021) 1083–1113.

[20] T.O. Olawumi, D.W.M. Chan, S. Ojo, M.C.H. Yam, Automating the modular construction process: A review of digital technologies and future directions with blockchain technology, Journal of Building Engineering 46 (2022) 103720.

[21] R. Kenley, O. Seppänen, Location-based management for construction: Planning, scheduling and control, Routledge, 2006.

Modeling Cooling Tower Fan Speed Using Symbolic Regression

Mohamed S. Kandil[1], Gary Chang[2] and J.J. McArthur[1]

[1]Dept. Architectural Science, Toronto Metropolitan University, Canada
[2]H.H. Angus & Associates, Canada
mkandil@torontomu.ca, garychang@hhangus.com, jjmcarthur@torontomu.ca

Abstract

Chilled water-cooling systems are essential for maintaining optimal building temperatures in commercial applications, such as hospitals, where precise temperature control is critical for patient care. Accurately estimating cooling tower fan speed is vital for energy efficiency and operational optimization. Traditional physics-based models often require iterative solving techniques, such as Newton-Raphson, leading to high computational costs and slow performance. This paper presents an alternative approach using symbolic regression to develop a data-efficient, explicit model for predicting cooling tower fan speed. Trained on just 72 hours of data, the model was validated on three separate periods: the first week immediately after training, the following week, and another week approximately two months later, achieving R^2 values of 0.92, 0.87, and 0.83, respectively. The inference speed is nearly instantaneous, as the model eliminates iterative calculations, making it well-suited for real-time applications. Using commonly available system parameters, the model enhances both computational efficiency and energy management. Results demonstrate its potential to improve HVAC system performance, reduce energy consumption, and support sustainable building operations, offering a scalable solution for practical deployment in commercial cooling systems.

Keywords –

HVAC, Cooling Tower, Simulation, Symbolic Regression

1 Introduction

Efficient cooling is essential for maintaining optimal environmental conditions in commercial buildings, particularly in facilities like hospitals where temperature control is critical for patient care [1]. Cooling towers, as integral components of chilled water-cooling systems, play a crucial role in rejecting heat from the system. Accurate estimation of the cooling tower's fan speed is vital, as it serves as a proxy for estimating power consumption through the application of the affinity laws, which describe how fan power scales with changes in fan speed. This relationship can be leveraged in simulations to optimize control variables for the condenser water loop, thereby enhancing the cooling system's performance [2]. By balancing the power consumption between chillers and cooling towers, it is possible to minimize the overall energy consumption of the chilled water system, leading to improved operational efficiency and reduced operational costs [3, 4, 5].

Traditionally, physics-based models (also known as gray-box models), such as Merkel-based models, have been employed to simulate cooling tower performance [6]. These models rely on a combination of fundamental physical principles and empirical data, providing robust options for system analysis and power estimation. However, they do not offer a direct output for fan speed, necessitating the use of iterative search methods like the Newton-Raphson technique to solve for fan speed [2]. This requirement for iterative solving introduces complexity and computational inefficiencies, rendering these models less suitable for real-time applications where rapid and responsive control is necessary.

In contrast, data-driven models (or black-box models) present a promising alternative by utilizing machine learning techniques to learn the relationship between input parameters—such as temperature, flow rate, and environmental conditions—and output variables like fan speed directly from data. While data-driven models can achieve high accuracy, they often suffer from limitations in generalization capability, particularly when operating outside the range of the training data. Additionally, these models typically require large datasets for effective training, which can limit their applicability in scenarios where data availability is constrained or where dynamic, real-time system responses are needed [7].

The motivation for this research lies in overcoming the limitations of both gray-box and black-box approaches by developing an interpretable and data-

efficient model using symbolic regression (SR). Symbolic regression provides a unique methodology that automatically discovers mathematical expressions to describe the relationships between input and output variables [8]. Unlike traditional machine learning methods, SR generates transparent, closed-form mathematical models that are both interpretable and computationally efficient [9]. This makes SR particularly suitable for real-time control strategies, as the resulting models eliminate the need for iterative numerical solutions required by gray-box models while avoiding the black-box nature and data dependency of conventional machine learning techniques.

2 Objectives of the Study

The primary objectives of this research are as follows:

1. Develop a data-efficient explicit model for cooling tower fan speed using symbolic regression, incorporating key input variables such as wet-bulb temperature, inlet and outlet water temperatures, and water flow rate, while minimizing the need for large datasets.
2. Improve simulation speed and computational efficiency by providing an explicit, direct model for fan speed estimation, thereby eliminating the need for iterative solutions required by traditional physics-based models and enabling faster real-time decision-making.
3. Leverage symbolic regression to create an interpretable and generalizable data-driven model that performs well with limited data and can be adopted in real-time applications, offering a practical and computationally efficient alternative to black-box models that typically require large, extensive datasets.

By addressing these objectives, this research aims to present a data-efficient, computationally efficient, and interpretable approach for estimating fan speed in chilled water-cooling systems. This advancement has the potential to enhance energy optimization, improve real-time control mechanisms, and offer a practical solution for systems with limited data, thereby contributing to more sustainable and cost-effective building operations.

3 Literature Review

Cooling towers are critical components of HVAC systems, responsible for rejecting heat from processes or buildings to the atmosphere. Their operational efficiency and energy consumption depend significantly on the control of fan speed, which in turn determines the leaving water temperature. Accurately predicting this temperature is essential for improving energy efficiency and ensuring optimal performance [3, 10, 11]. To model the leaving water temperature, three primary methods have been developed over the years: the Merkel method, the Poppe method, and the enthalpy-driven effectiveness-NTU method [6]. These models have served as the foundation for cooling tower analysis and control strategies.

The Merkel method, first proposed in 1925, simplifies the cooling process by assuming a constant Lewis factor (the ratio of sensible to latent heat transfer coefficients) and neglecting the evaporation losses of water. It integrates the heat and mass transfer equations along the height of the cooling tower to estimate the leaving water temperature [6]. While this approach has been widely adopted for its simplicity and ease of implementation, it fails to capture the complex interactions between heat and mass transfer that occur in real operating conditions, particularly under varying ambient and operational parameters.

The Poppe method improves upon Merkel's assumptions by considering the simultaneous and coupled heat and mass transfer processes inside the cooling tower. Unlike Merkel's method, the Poppe approach does not assume a constant Lewis factor and accounts for variations in the air and water enthalpy. This method provides higher accuracy in predicting the leaving water temperature, particularly in scenarios where Merkel's assumptions lead to discrepancies [12]. However, the Poppe model requires solving a system of coupled nonlinear differential equations, which makes it computationally intensive and less suited for real-time control.

The enthalpy-driven effectiveness-NTU method provides a more flexible and systematic framework for predicting the leaving water temperature. By introducing the concept of heat exchanger effectiveness, this method evaluates cooling tower performance using enthalpy as the driving force. It allows for improved accuracy in capturing cooling tower dynamics under varying conditions compared to traditional methods [2]. Despite its advantages, the enthalpy-driven method still involves solving nonlinear relationships, which can pose challenges for real-time optimization and control strategies. Another, simpler model was proposed based on Merkel’s theory and the effectiveness-NTU method. This model was developed through energy balance and heat-mass transfer analysis [13].

For a closed-loop cooling tower leaving water temperature control strategy, the fan speed must be set to its optimal value to maintain the optimal leaving water temperature. However, if the desired leaving water temperature setpoint deviates from this optimal condition, determining the appropriate fan speed requires solving

the governing equation numerically. Equation (1) illustrates this relationship [2]:

$$T_{CWS} = f_T(T_{CWR}, T_{WB}, M_{CW}, SPD, K_d'A) \quad (1)$$

where T_{CWS} is the leaving water temperature, T_{CWR} is the entering water temperature, T_{WB} is the ambient air wet bulb temperature, M_{CW} is the water mass flow rate, SPD is the fan speed, and $K_d'A$ the design moisture transfer coefficient.

Solving for SPD under these conditions requires iterative numerical techniques such as the Bisection method or Newton-Raphson method, which are widely used for solving nonlinear equations. These methods, while effective, are computationally demanding and time-consuming, especially for real-time applications where fast control responses are essential [2, 3]. Similarly, for closed-loop approach temperature control, the fan speed must also be determined iteratively if the approach temperature setpoint differs from the optimal value. These computational requirements can lead to delays in control responses, limiting their feasibility for real-time applications [3]. Thus, while the Merkel, Poppe, and enthalpy-driven NTU methods provide robust frameworks for cooling tower performance prediction, their reliance on computationally intensive numerical solvers introduces significant overhead during closed-loop control.

To overcome the computational challenges associated with traditional models, data-driven approaches have gained increasing attention. Data-driven methods utilize operational data to identify complex system dynamics and provide real-time predictive models. Among these approaches, symbolic regression (SR) has emerged as a powerful technique for deriving explicit and interpretable mathematical expressions from data [8, 9]. SR identifies relationships between variables without relying on predefined model structures, enabling the discovery of closed-form equations that can directly describe the behavior of cooling towers, including fan speed and leaving water temperature.

Although symbolic regression has been applied successfully in other areas of HVAC systems, such as building thermodynamics [14] and gray-box modeling of air conditioning systems [15], its application to cooling tower modeling remains underexplored. [14] used SR to model building energy dynamics and demonstrated a 16.1% reduction in peak power through a predictive control framework. [15] applied SR for gray-box modeling of unitary air conditioning systems, achieving high accuracy in predicting cooling capacity and coefficient of performance (COP). [16] further showcased the potential of SR for indoor temperature forecasting, reducing model complexity while improving computational efficiency. These studies highlight SR's ability to capture nonlinear dynamics effectively while providing explicit and interpretable solutions, making it a suitable candidate for real-time applications.

In this research, we address the limitations of traditional cooling tower models and propose a symbolic regression-based model to predict cooling tower fan speed as a function of the leaving water temperature, entering water temperature, water mass flow rate, and wet bulb temperature. Unlike traditional methods, which require iterative numerical solutions to determine fan speed for a given setpoint, our approach formulates an explicit and interpretable model derived directly from operational data.

The proposed symbolic regression (SR) model eliminates the need for computationally intensive techniques, such as the Bisection method or Newton-Raphson iterations, which are typically employed to solve the governing nonlinear equations. By generating closed-form expressions, the SR model enables faster and more efficient inference, making it particularly well-suited for closed-loop control strategies. This capability allows for real-time adjustments of fan speed to maintain desired leaving water temperature or approach temperature setpoints. Furthermore, the resulting SR model is computationally efficient and highly interpretable, providing clear relationships between key operational variables—leaving and entering water temperatures, wet bulb temperature, and water mass flow rate. This facilitates real-time optimization of cooling tower performance, overcoming the bottlenecks inherent to traditional numerical approaches while ensuring robust and accurate control of cooling tower operations.

4 Methodology

4.1 Case study and Data Collection

The primary case study for this research is a hospital located in the Greater Vancouver area, within ASHRAE climate zone 4. This hospital is one of the busiest in the region, providing comprehensive care across multiple specialties, including trauma, cardiac, neurosurgery, high-risk obstetrics, neonatal intensive care, and acute mental health services. The diverse range of services necessitates a robust and efficient cooling system to maintain optimal environmental conditions, which are critical for patient care. The hospital's central cooling plant, comprising chillers and cooling towers, is responsible for meeting the facility's substantial cooling demands.

Data for this study were collected from the central cooling system of the hospital over a five-month period, from May 1, 2024, to September 30, 2024. However, only 72 hours of data, specifically from July 8th to July

10th, were utilized for training the models (see Figure 1). This subset consists of 288 data points, sampled at 15-minute intervals, capturing the system's dynamic behaviour. Representing approximately 2% of the total collected data, these 72 hours were carefully selected to include a wide range of operational conditions, such as fluctuations in outdoor temperature, humidity, and cooling demand. This ensures a diverse and representative dataset that effectively captures the cooling system's performance across different operating states.

The collected variables include leaving water temperature (T_{CWS}, in °C), entering water temperature (T_{CWR}, in °C), condenser water mass flow rate (M_{CW}, in L/s), dry-bulb outdoor air temperature (OAT, in °C), relative humidity (RH, in %), and cooling tower fan speed (SPD, in %). The fan speed during this period varied, with a lower bound of 33.3%, and its operational range was crucial for evaluating the system's performance.

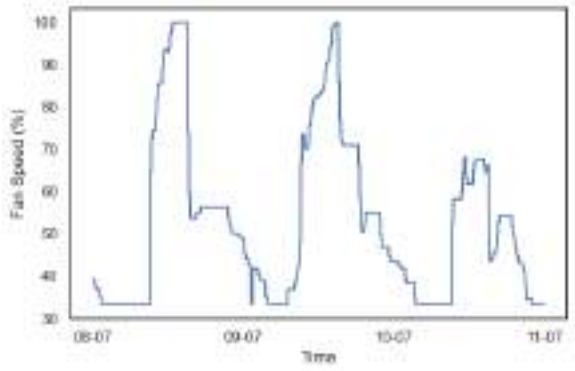


Figure 1: Cooling tower fan speed dataset used for model training

4.2 Data Pre-processing and Feature Engineering

To ensure the quality and relevance of the data for modeling, several pre-processing and feature engineering steps were undertaken. Wet-Bulb Temperature was calculated using both Relative Humidity and Air Temperature based on the formula provided by [17]. This additional feature was included to better capture the thermodynamic conditions affecting the cooling tower performance.

The raw dataset underwent a series of filtering criteria to remove unreliable or irrelevant measurements. First, instances where the fan speed (SPD) was below 33.3% were discarded to eliminate periods of suboptimal operation. Additionally, data points with a wet-bulb temperature below 10°C were excluded to focus the analysis on conditions where cooling demands are more significant. Measurements where the condenser water supply temperature (T_{CWS}) exceeded the condenser water return temperature (T_{CWR}) were also removed, as such scenarios are physically implausible and likely indicative of sensor errors or operational issues.

Further, any records where the condenser water mass flow rate (M_{CW}) fell below 50 L/s were filtered out to ensure the dataset only included stable and efficient operating conditions. Finally, rows containing NaN values in any of the input or target variables were discarded to maintain data integrity and avoid computational issues during model training.

After the pre-processing steps, a total of 283 data points remained, reflecting the impact of the filtering criteria in removing unreliable or irrelevant data. This cleaned dataset was subsequently split into training and testing subsets, with an 80-20% division, respectively. The split was performed without shuffling to preserve the temporal sequence of the data, which is critical given the time-series nature of the dataset. For modeling, the fan speed (SPD) was designated as the target variable, while the remaining variables were selected as the input features. This structured and pre-processed dataset provided a robust foundation for the subsequent symbolic regression modeling and analysis.

4.3 Symbolic Regression Modeling with PySR

The modeling process was conducted using Python, leveraging the PySR package [18] for symbolic regression. Symbolic regression is a data-driven approach that searches the space of mathematical expressions to find models that best fit the data. Unlike traditional regression methods that require predefined forms of the model, symbolic regression explores a wide variety of possible equations, offering both accuracy and interpretability [8].

PySR employs a genetic algorithm-based approach to evolve mathematical expressions, optimizing them to minimize the error between the predicted and actual values. The underlying computational engine utilized Julia, a high-performance programming language, to enhance the efficiency and scalability of the symbolic regression process. This integration allows for rapid evaluation of numerous candidate models, facilitating the discovery of optimal or near-optimal symbolic representations of the target variable.

In this study, the symbolic regression was configured with specific hyper-parameters to guide the model generation process. The binary operators included in the search space were addition (+), multiplication (*), subtraction (-), division (/), and exponentiation (^). The loss function was set to L2DistLoss(), which corresponds to the mean squared error between the predicted and actual fan speeds. Model selection was based on accuracy to ensure that the generated models achieved high predictive performance.

To accelerate the training process, several performance settings were enabled. The multithreading setting was activated to utilize multiple CPU cores, improving computational efficiency. The fast_cycle setting was set to False, which prevents the batch over population subsamples algorithm, designed to speed up the cycles by approximately 15%, but at the risk of being less algorithmically efficient. The turbo setting was also False, as it refers to the experimental use of LoopVectorization.jl, which speeds up the search evaluation but may not support certain operators and does not work with 16-bit precision floats. Finally, the bumper setting was turned off; this experimental feature uses Bumper.jl to speed up the search evaluation, though it also does not support 16-bit precision floats. These configurations were carefully chosen to strike a balance between computational efficiency and model accuracy, ensuring that the symbolic regression process remained robust, scalable, and suitable for generating interpretable models for cooling tower fan speed prediction.

4.4 Model Training and Validation

The training phase involved fitting the symbolic regression model to the training dataset using the specified hyper-parameters and performance settings. The model was trained to identify the mathematical relationship between the input features (T_{CWS}, T_{CWR}, T_{WB}, and M_{CW}) and the target variable (SPD). This relationship can be mathematically represented in the following general form:

$$SPD = f_{SPD}(T_{CWS}, T_{CWR}, T_{WB}, M_{CW}) \quad (2)$$

Once the model was generated, its performance was evaluated using two key metrics: R-squared (R^2) to measure the proportion of variance in the target variable that is predictable from the input features and Root Mean Squared Error (RMSE) to provide an absolute measure of the prediction error in the same units as the target variable.

To assess the model's generalizability, validation was performed on an independent dataset comprising two weeks of subsequent data. This validation step was crucial to ensure that the symbolic regression model maintained its accuracy and reliability when applied to new, unseen data. The performance metrics on the validation set provided an indication of the model's robustness and its potential applicability in real-world cooling system optimization.

Overall, the methodology combines rigorous data collection and preprocessing with advanced symbolic regression techniques to develop an accurate and interpretable model for predicting cooling tower fan speed. The use of PySR and its efficient computational backend facilitates the discovery of optimal mathematical models that can enhance the understanding and control of cooling system operations.

5 Results

5.1 Symbolic Regression Model

The best-performing model derived through symbolic regression is represented by the following mathematical expression:

$$SPD = 0.707M_{cw} + \frac{0.707M_{cw}}{T_{WB}} - \frac{13.2M_{cw}}{T_{cwr}} + T_{cwr}^{\frac{0.764T_{WB}}{T_{cwr}-9.71}} - 13.5 - M_{cw}^{-0.878M_{cw}} \quad (3)$$

This equation provides a direct relationship between cooling tower fan speed and the key input features: wet-bulb temperature, condenser water supply temperature, condenser water return temperature, and mass flow rate of condenser water. The model is interpretable and captures the nonlinear interactions among the input variables.

5.2 Performance Evaluation on Test Data

The model's performance was first evaluated on the designated test dataset, comprising 20% of the data selected from the initial 72-hour training period. The following metrics were obtained:

- R^2 score: 0.8881
- RMSE: 3.9325

These results indicate that the model captures a significant portion of the variability in the fan speed while maintaining low prediction error. The R^2 score of 0.89 demonstrates high predictive accuracy, while the RMSE of approximately 3.93 reflects the model's strong ability to estimate fan speed with minimal deviation from the actual values.

5.3 Performance Evaluation on Validation Data

To assess the model's generalization capability, it was validated on an independent dataset across three separate periods: the first week after training (July 11th to 18th), the following week (July 18th to 25th), and an additional week nearly two months later (September 1st to 8th), demonstrating its credibility and robustness.

For the week of July 11th to 18th, 2024, the model achieved an R^2 of 0.92 and an RMSE of 3.76, demonstrating its strong predictive accuracy and robustness in estimating fan speed. The time-series plot (Figure 2) illustrates a close alignment between actual and predicted values, while the scatter plot (Figure 3) confirms a strong correlation, with most points closely

following the diagonal reference line, indicating minimal prediction error.

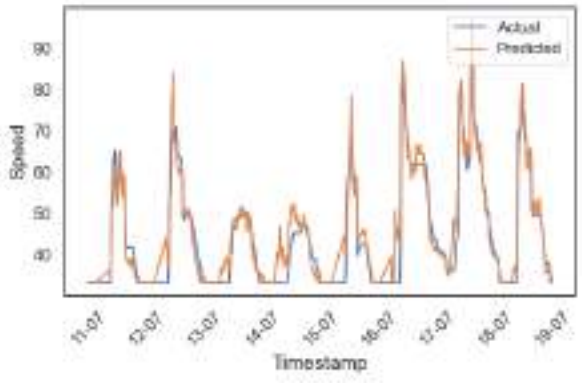


Figure 2 Time series plot of actual versus prediction values for first week of validation data

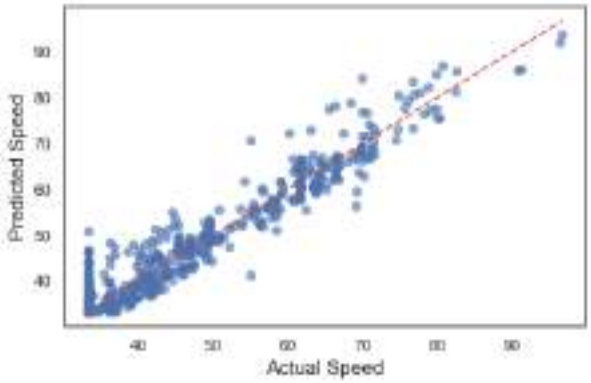


Figure 3 Scatter plot of actual versus prediction values for first week of validation data

For the week of July 18th to 25th, 2024, the model achieved an R^2 of 0.87 and an RMSE of 3.82, maintaining strong predictive performance and consistency in estimating fan speed. The time-series plot (Figure 4) illustrates a close match between actual and predicted values, while the scatter plot (Figure 5) confirms a strong correlation, with most points aligning along the diagonal reference line. The slight decline in R^2 suggests minor operational variations, yet the model continues to generalize well, reinforcing its reliability in diverse conditions.

To evaluate the performance and generalizability of the model, it was further validated on an independent dataset covering the week of September 1st to 8th, 2024, achieving an R^2 of 0.83 and an RMSE of 6.20. While the slight decline in R^2 suggests some operational variations, the model remains effective in capturing key system dynamics. The time-series plot (Figure 6) shows strong alignment between actual and predicted values, while the scatter plot (Figure 7) confirms a consistent correlation, reinforcing the model's reliability in predicting fan speed under varying conditions.

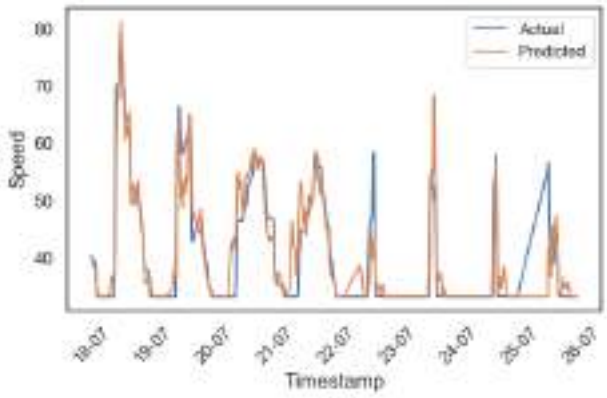


Figure 4 Time series plot of actual versus prediction values for second week of validation data

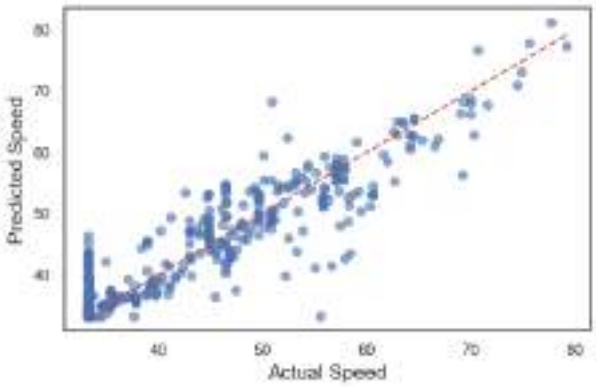


Figure 5 Scatter plot of actual versus prediction values for second week of validation data

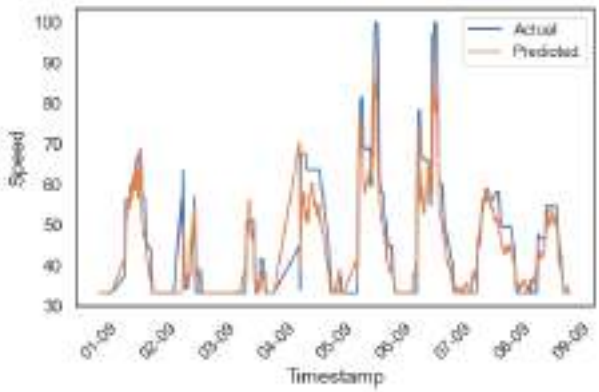


Figure 6 Time series plot of actual versus prediction values for third week of validation data

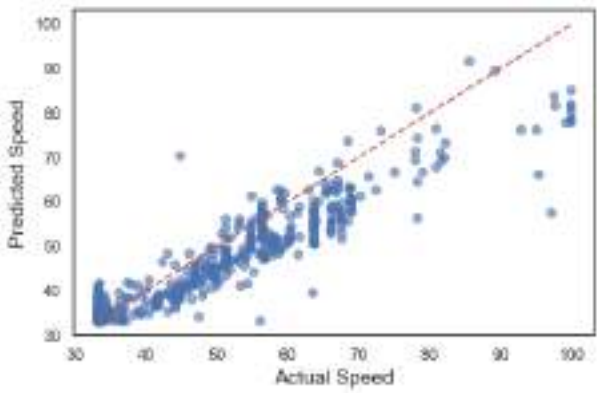


Figure 7 Scatter plot of actual versus prediction values for third week of validation data

6 Discussion

On the test dataset (20% of the training period), the model achieved an R^2 score of 0.8881 and an RMSE of 3.9325, capturing nearly 89% of the variance in fan speed with minimal prediction error. Further validation confirmed its generalizability. During the first validation week (July 11–18, 2024), the model achieved an R^2 of 0.9206 and an RMSE of 3.7566, demonstrating strong agreement between predicted and actual values. For the second validation week (July 18–25, 2024), the model maintained an R^2 of 0.8676 and an RMSE of 3.8246, despite slight variations likely caused by operational changes. A third validation nearly two months later (September 1–8, 2024) resulted in an R^2 of 0.8349 and an RMSE of 6.1975. While the R^2 declined slightly over time, the model remained effective, confirming its robustness under changing conditions.

The symbolic regression model eliminates the need for iterative numerical methods, such as Newton-Raphson, which are computationally intensive and impractical for real-time applications. By generating explicit mathematical expressions, the model enables fast inference, allowing real-time adjustments to cooling tower fan speed to maintain desired operating conditions.

Beyond computational efficiency, the model supports practical energy optimization strategies for chilled water plants, where balancing chiller and cooling tower operation is critical to minimizing total energy consumption. Cooling tower fans consume energy, but their operation influences chiller efficiency; optimizing only one component in isolation may lead to suboptimal system performance. The proposed model can be integrated into real-time optimization frameworks to determine the best trade-off between chiller and cooling tower energy use, reducing overall power consumption while ensuring adequate cooling.

With its fast inference capability, the model enables real-time control adjustments based on changing conditions, preventing excessive cooling tower fan speeds that could increase energy use unnecessarily. Its explicit equations provide a direct relationship between key operational parameters, making it suitable for predictive control applications where chiller load and cooling tower operation are jointly optimized for efficiency.

The model's interpretability also facilitates decentralized control, where cooling tower fan speed is adjusted at the local level rather than relying on a central controller. This reduces communication delays and computational overhead while improving responsiveness to system dynamics. In large-scale systems with multiple cooling towers, localized optimization enhances overall performance, leading to energy savings and improved reliability.

By integrating the model into smart HVAC control platforms, facilities can automate real-time energy management strategies to optimize cooling performance while reducing costs. This scalable approach makes the model applicable to hospitals, commercial buildings, and industrial facilities where precise control and efficiency are essential.

In summary, the symbolic regression model provides a practical, real-time solution for cooling tower control. By balancing chiller and cooling tower energy use, it enhances system-wide efficiency, offering a valuable tool for intelligent HVAC management in energy-conscious applications.

7 Conclusions

This study demonstrated the effectiveness of symbolic regression for predicting cooling tower fan speed in central chilled-water systems. The SR model achieved high accuracy, strong generalization, and computational efficiency across test and validation datasets. By producing explicit and interpretable mathematical models, it overcomes the limitations of traditional physics-based approaches and data-hungry black-box techniques.

The proposed model offers a practical and scalable solution for real-time optimization and decentralized control of cooling systems, facilitating improved energy efficiency and operational performance. Future work will focus on enhancing the model's adaptability to dynamic system changes and integrating it into broader energy management frameworks to further optimize large-scale cooling operations.

Acknowledgments

This research was funded by the Natural Science and Engineering Research Council (Canada) [ALLRP-580958-22] and H. H. Angus and Associates Consulting Engineers Ltd.

Bibliography

[1] D. Schurk, "Conditioning for the Environment of Critical Care: Hospital Operating Rooms," *ASHRAE Journal*, vol. 61, no. 10, pp. 16-24, 2019.

[2] G. Wang, S. Ghoddousi, L. Ding and L. Song, "Investigation of Different Cooling Tower Fan Control Strategies Using Actual Chiller Cop and Calibrated Cooling Tower and Fan Models," *Energy and Buildings*, vol. 277, p. 112585, 2022.

[3] J. Liao, X. Xie, H. Nemer, D. E. Claridge and C. H. Culp, "A simplified methodology to optimize the cooling tower approach temperature control schedule in a cooling system," *Energy Conversion and Management*, vol. 199, p. 111950, 2019.

[4] K. Ma, M. Liu and J. Zhang, "Online optimization method of cooling water system based on the heat transfer model for cooling tower," *Energy*, vol. 231, p. 120896, 2021.

[5] K. Zaw, K. Matsuoka and T. Poh, "A real-time operational optimization of commercial district cooling systems: A practical application," *Energy and Buildings*, vol. 297, p. 113434, 2023.

[6] J. C. Kloppers and D. G. Kröger, "Cooling Tower Performance Evaluation: Merkel, Poppe, and e-NTU Methods of Analysis," *J. Eng. Gas Turbines Power*, vol. 127, no. 1, pp. 1-7, 2005.

[7] S. Ikeda and T. Nagai, "A novel optimization method combining metaheuristics and machine learning for daily optimal operations in building energy and storage systems," *Applied Energy*, vol. 289, p. 116716, 2021.

[8] G. Kronberger, B. Burlacu, M. Kommenda, S. M. Winkler and M. Affenzeller, Symbolic Regression, CRC Press, 2024.

[9] D. Angelis, F. Sofos and T. E. Karakasidis, "Artificial Intelligence in Physical Sciences: Symbolic Regression Trends and Perspectives," *Archives of Computational Methods in Engineering*, vol. 30, p. 3845–3865, 2023.

[10] K. Singh and R. Das, "Exergy optimization of cooling tower for HGSHP and HVAC applications," *Energy Conversion and Management*, vol. 136, pp. 418-430, 2017.

[11] G. F. Cortinovis, J. L. Paiva, T. W. Song and J. M. Pinto, "A systemic approach for optimal cooling tower operation," *Energy Conversion and Management*, vol. 50, no. 9, p. 2200–2209, 2009.

[12] D. R. Baker and H. A. Shryock, "A Comprehensive Approach to the Analysis of Cooling Tower Performance," *Journal of Heat Transfer*, vol. 83, no. 3, pp. 339-349, 1961.

[13] G.-Y. Jin, W.-J. Cai, L. Lu, E. L. Lee and A. Chiang, "A simplified modeling of mechanical cooling tower for control and optimization of HVAC systems," *Energy Conversion and Management*, vol. 48, no. 2, pp. 355-365, 2007.

[14] Y. Ozawa, D. Zhao, D. Watari, I. Taniguchi, T. Suzuki, Y. Shimoda and T. Onoye, "Data-driven HVAC Control Using Symbolic Regression: Design and Implementation," in *2023 IEEE Power & Energy Society General Meeting (PESGM)*, Orlando, FL, USA, 2023.

[15] S. Yousaf, C. R. Bradshaw, R. Kamalapurkar and O. San, "A gray-box model for unitary air conditioners developed with symbolic regression," *International Journal of Refrigeration*, vol. 168, pp. 696-707, 2024.

[16] D. Zhao, Z. Chen, Z. Li, Y. Yuan and I. Taniguchi, "Improving Building Temperature Forecasting: A Data-driven Approach with System Scenario Clustering," in *2024 IEEE Power & Energy Society General Meeting (PESGM)*, Seattle, WA, USA, 2024.

[17] R. Stull, "Wet-Bulb Temperature from Relative Humidity and Air Temperature," *Journal of Applied Meteorology and Climatology*, vol. 50, no. 11, p. 2267–2269, 2011.

[18] M. Cranmer, "Interpretable Machine Learning for Science with PySR and SymbolicRegression.jl," *arXiv preprint arXiv:2305.01582*, 2023.

Developing Computer Vision-based Digital Twin for Vegetation Management Near Power Distribution Networks

F. Bahreini[1], M. Nik-Bakht[2], A. Hammad[1], and M. Gaha[3]

[1]Concordia Institute for Information Systems Engineering, Concordia University, Canada
[2]Department of Building, Civil and Environmental Engineering, Concordia University, Canada
[3]Hydro-Quebec Research Institute, Quebec, Canada
fardin.bahreini@concordia.ca, mazdak.nikbakht@concordia.ca, amin.hammad@concordia.ca, gaha.mohamed@hydroquebec.com

Abstract –

The rapid development of digital twin technology has opened new avenues for infrastructure management, particularly for addressing vegetation encroachment risks near power lines. This paper builds upon our previous work in LiDAR-based proximity detection by proposing a framework for creating digital twin for vegetation management near power distribution networks. The framework leverages the RandLA-Net model for semantic segmentation of power lines, poles, and vegetation followed by clustering and rule-based thresholding for data refinement. Detecting vegetation encroachment is achieved through KDTree-based spatial analysis, ensuring efficient identification of risk zones. The segmented and processed point cloud data is then transformed into detailed 3D models, forming the basis of the digital twin, which can be enhanced in the future by adding advanced semantic attributes and predictive tree growth models, enabling proactive vegetation management. The methodology is demonstrated through a case study, highlighting its potential to enhance operational efficiency and the resilience of power distribution networks.

Keywords –

Digital Twin, Computer Vision, 3D Point Cloud, Power Lines, Semantic Segmentation, Vegetation Management

1 Introduction

Vegetation encroachment near power lines is a critical issue that threatens the reliability and safety of power distribution networks. Approximatively 40% of the distribution outages are caused by vegetation. Power outages often happen when vegetation is exposed to severe weather conditions [1]. Traditionally, vegetation management has relied on manual inspection, which is labor-intensive, time-consuming, and often limited in accuracy and frequency. This highlights the need for automated solutions capable of delivering accurate, consistent, and cost-effective monitoring. Advancements in sensing technologies, such as Light Detection and Ranging (LiDAR), and machine learning algorithms have significantly improved vegetation management processes [2]. LiDAR's ability to capture high-resolution 3D spatial data enables precise detection of vegetation and its proximity to power lines, surpassing the capabilities of traditional 2D imaging or visual inspection. However, the vast volume of data generated by LiDAR systems presents computational challenges, particularly in efficiently segmenting and classifying power lines, poles, and vegetation. Semantic segmentation models have been pivotal in overcoming these challenges, offering high-accuracy classification of large-scale point clouds [3].

On the other hand, digital twin technology provides a novel framework for addressing these challenges by enabling the creation of dynamic virtual replicas of vegetation and power distribution networks. Digital twins integrate real-world data with advanced modeling and analytical capabilities, offering several benefits, including the potential to model the dynamic behavior of trees, such as growth patterns, movements under loads, and responses to environmental factors like wind [4, 5]. This capability is further enhanced by incorporating additional attributes such as tree type and age, allowing for predictive and adaptive management strategies. Additionally, digital twins facilitate improved asset management for utility poles and power lines by integrating attributes like material type, age, and condition, enabling comprehensive and data-driven maintenance planning.

In our previous work, we demonstrated the application of RandLA-Net for vegetation segmentation and proximity detection, achieving improved accuracy and operational efficiency [6]. Building on this work, the present study advances these efforts by developing a digital twin for vegetation management near power

distribution network. By leveraging 3D models derived from LiDAR data, this study aims to enhance the identification and management of vegetation risks near power lines. The proposed framework integrates refined segmentation, clustering methods and proximity analysis, supporting proactive maintenance strategies and real-time operational insights. The contributions include an integrated framework for accurate risk detection, a digital twin with 3D models, and a scalable solution that can be used for improving the resilience of distribution networks, as well as for managing the assets of these networks. The paper is structured as follows: Section 2 reviews existing works on digital twins based on point cloud analysis. Section 3 describes the proposed framework. Section 4 presents the implementation of the framework. Section 5 discusses findings and future directions, while Section 6 concludes the paper and outlines potential areas for further research.

2 Related Work on Digital Twins Based on Point Cloud Data

Digital twins can be used for assessing risks and managing vegetation near power lines. These models leverage point cloud data to create detailed 3D models, enabling precise monitoring of spatial relationships between trees and infrastructure.

Surface reconstruction is a critical step in translating raw point cloud data into usable 3D models. Techniques such as Delaunay triangulation and KD-tree (k-dimensional tree) indexing, as highlighted by Glumova and Filinskih [7], have proven effective for generating high-quality meshes from unstructured data. Similarly, Ismail et al. [8] explored methods like Poisson Reconstruction and Ball Pivoting Algorithm, focusing on their adaptability and simplicity for applications like building modeling in digital twins. Tagarakis et al. [9] emphasized maintaining geometric accuracy during mesh generation, which is vital for capturing intricate details like tree branches critical for proximity assessments. Chen et al. [10] proposed integrating dynamic simulation data to enhance mesh precision, offering better insights into environmental interactions and vegetation growth.

Alignment of point cloud data with pre-existing models is facilitated by advanced matching algorithms. Zhang and Li [11] utilized simulation tools such as Blender and the Grove plugin to model tree growth under various environmental conditions, ensuring accuracy through comparisons with LiDAR-derived ground truth. Gobeawan et al. [12] extended this by introducing species-specific growth models, precisely aligning these models within spatial constraints derived from point clouds to represent biological and structural tree characteristics effectively.

Digital twins offer continuous monitoring and risk assessment capabilities for power line management. Buonocore et al. [13] demonstrated their use in predicting vegetation growth and simulating environmental impacts, such as weather events, to mitigate risks.

Kim et al. [14] introduced a framework integrating semantic segmentation, geometric processing, and data export into CityGML, streamlining the generation of accurate indoor models from point clouds. Other researchers, such as La Guardia and Koeva's [15] developed interactive 3D web navigation model. Wu et al. [16] proposed a precision alignment method for CAD data to enhance digital twin applications in urban and industrial contexts. Agapaki and Brilakis [17] proposed a deep learning approach for matching point cloud clusters with CAD models. Li et al. [18] developed GRNet framework for 3D object detection and demonstrated the potential of digital twins in diverse infrastructure scenarios. While digital twins have been proposed in urban modeling, existing studies lack an integrated framework that combines LiDAR-based segmentation with digital twin modeling for vegetation management.

3 Proposed Framework

This study extends our prior work [6], where we developed a computer vision framework for managing vegetation near power lines, employing the Random Sampling in Large-scale Point Cloud Analysis Network (RandLA-Net) [19] for semantic segmentation of large-scale 3D point clouds. This framework also included post-processing techniques such as Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [20] for object isolation and rule-based thresholding to filter urban elements and mitigate noise. Furthermore, a proximity detection method was utilized to assess potential risks by analyzing spatial relationships between vegetation and power lines in urban settings. Building on this framework, the current study enhances and integrates these processes into a comprehensive digital twin framework tailored for vegetation management near power distribution networks. The extended framework advances the previous work by creating detailed, dynamic 3D models of power lines, poles, and vegetation using LiDAR data. These models facilitate continuous monitoring and risk assessment, aligning with the concept of a digital twin. Figure 1 shows the proposed integrated framework, where the segmentation and post-processing pipeline serves as the core data processing engine, while the new components enable the generation of the digital twin for enhanced decision-making and management capabilities.

(1) Semantic Segmentation: The RandLA-Net model [19] is employed for segmenting LiDAR point clouds into key classes such as power lines, poles, and

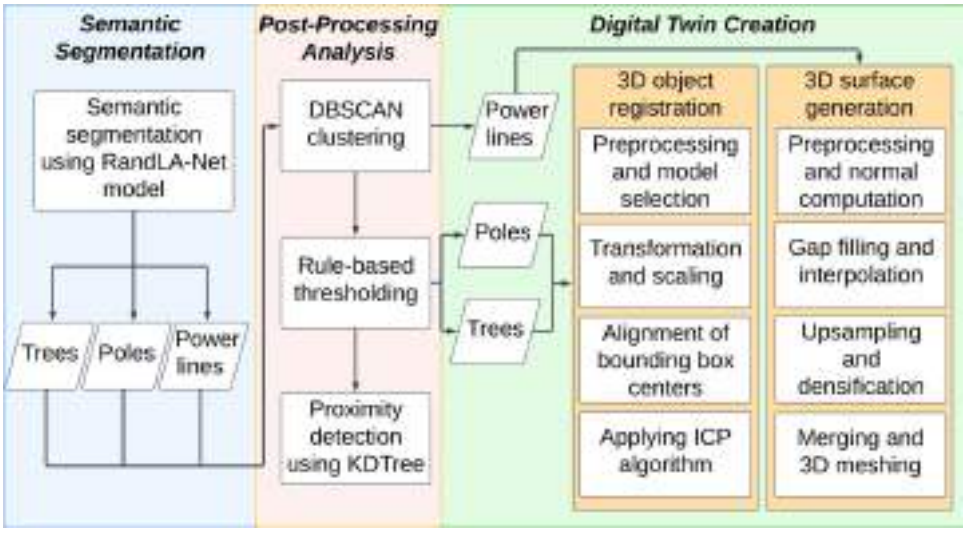


Figure 1. Proposed framework

vegetation, leveraging its efficiency in processing high-density data. This model is chosen for its ability to achieve high accuracy on large-scale datasets, including its notable performance on the Toronto-3D dataset, where it demonstrated superior overall accuracy compared to other methods. RandLA-Net employs random point sampling for downsampling, significantly reducing computational overhead while preserving critical geometric details within the point cloud. Its architecture integrates innovative components, including a local feature aggregation module, comprising local spatial encoding and attentive pooling, which effectively captures complex local structures. Additionally, its use of shared Multilayer Perceptrons (MLPs) and dilated residual blocks enhances both processing speed and scalability, enabling it to handle up to one million points per iteration efficiently. To evaluate the model's performance, the Intersection over Union (IoU) metric is utilized, providing a robust measure of the overlap between predicted and ground truth classifications for each class.

(2) Post-processing: Following the segmentation process, the output data undergoes a series of post-processing steps to enhance its accuracy and utility. DBSCAN algorithm is employed to identify clusters within the segmented point cloud, leveraging its capability to detect clusters of varying shapes and densities without requiring predefined cluster counts. Key parameters, such as epsilon and minimum samples, are carefully tuned to suit the complex structure of urban environments, ensuring sensitivity to variations in density and distribution. Complementing DBSCAN, rule-based thresholding is applied to further refine the output, using class-specific thresholds, such as height and point count, to isolate critical features like trees and poles while effectively filtering noise and irrelevant data. To evaluate proximity risks, that could cause the encroachment of vegetation on power lines, a K-Dimensional Tree (KDTree) structure is used for efficient spatial querying. This approach allows for rapid identification of the nearest tree to each power line, leveraging the KDTree's capacity for efficient multi-dimensional space queries. Proximities are then compared against safety thresholds to flag potential hazards. However, the current proximity detection and risk assessment rely on static conditions, which limits their ability to account for dynamic environmental factors such as wind, tree movement, or weather changes.

(3) Digital Twin Creation: This study employs a systematic approach to construct a seamless representation of power lines, poles, and trees from segmented point cloud data, using Python for 3D data handling and analysis. For power lines, the process begins with importing the segmented point cloud, representing discrete points corresponding to power line geometry. To estimate normal vectors and curvature at each point, a k-nearest neighbor (KNN) search algorithm (K=10) is applied, balancing efficiency and accuracy in densely populated regions. Normal vectors are calculated by fitting a plane to the nearest neighbors, with curvature derived from deviations of neighboring points from the fitted plane, enabling precise modeling of intricate power line geometries. To address data gaps in the segmented point cloud, Radial Basis Function (RBF) interpolation is used for its effectiveness in reconstructing missing points while ensuring smooth, continuous outputs. Additionally, upsampling is employed to achieve uniform point density, enhancing the resolution and accuracy of power line representation.

For poles and trees, segmented point cloud data are aligned with reference 3D models sourced from libraries like Sketchfab [21] and CGTrader [22], assuming all



(a) Examples of poles 3D models



(b) Examples of trees 3D models

Figure 2. Poles and trees libraries

trees belong to the Honey Locust species, which is common in Canada. Figures 2 (a) and (b) show examples of the poles and trees libraries, respectively). The Iterative Closest Point (ICP) algorithm is used for alignment, iteratively refining transformation parameters (translation, rotation, and scaling) to minimize discrepancies between the segmented objects and reference models. Bounding box centers of segmented objects and reference models are aligned for accurate spatial placement, while a scaling step is included to adjust for size differences by computing and applying appropriate scaling factors based on bounding box dimensions. This process involves initializing alignment parameters, applying an initial transformation matrix, and iteratively refining alignment to ensure better matching of the model, location, and size.

4 Implementation

The implementation of the proposed framework builds upon the methodologies discussed earlier, extending the capabilities established in previous work to realize a comprehensive digital twin system for vegetation management near power distribution networks. The process was executed in the following stages:

(1) Data Acquisition: This study utilized the Toronto-3D dataset [23], a high-density point cloud dataset collected via Mobile Laser Scanning (MLS) along a one-kilometer stretch of Avenue Road in Toronto, Canada. The dataset, consisting of approximately 78.3 million points with an average ground point density of 1,000 points per square meter, is particularly suited for capturing intricate details of urban environments. Each data point includes attributes such as XYZ coordinates, LiDAR intensity, RGB reflectance, GPS time, scan angle, and class labels, which classify features like roads, trees, poles, power lines, and buildings. The dataset's high density and comprehensive labeling make it ideal for semantic segmentation tasks in this study. Preprocessing steps involved grid subsampling at a 0.06-meter resolution to balance data volume and feature retention. Additionally, projection indices were generated to map processed results back to the original dense point cloud, ensuring consistency and accuracy. These preparations provided a well-structured dataset optimized for semantic segmentation and further analysis. Figure 3 shows point cloud data of a sample area.

(2) Semantic Segmentation: The RandLA-Net model was trained and optimized using the Toronto-3D dataset, focusing on the segmentation of vegetation, poles, and power lines. The training process involved dividing the dataset into sections for training and testing, with three sections designated for training and one for evaluation. The model underwent 100 epochs of training using the Adam optimizer with an initial learning rate of 0.01, which decreased by 5% per epoch to improve convergence. A batch size of 4 was selected to balance computational efficiency and training performance.

Training was conducted over 124 hours and 33 minutes on a workstation equipped with an NVIDIA RTX A6000 GPU and a 48-core CPU, ensuring high computational power for processing the dense point cloud data. However, the inference time takes about 10 minutes. The model incorporated geometric features along with RGB



Figure 3. A sample area of point cloud data

and intensity attributes to enhance segmentation accuracy. To ensure optimal model performance, various configurations of RandLA-Net were tested on the Toronto-3D dataset. This involved systematically exploring different combinations of hyperparameters to evaluate their impact on classification and segmentation accuracy. The best-performing configuration from these tests was selected and utilized in the subsequent stages. The test results demonstrated high IoU scores for individual classes, achieving 97.05% for trees, 87.44% for power lines, and 82.33% for poles, ensuring reliable classification for the target features. This robust performance underpins the digital twin framework by providing precise and detailed semantic segmentation of the urban environment.

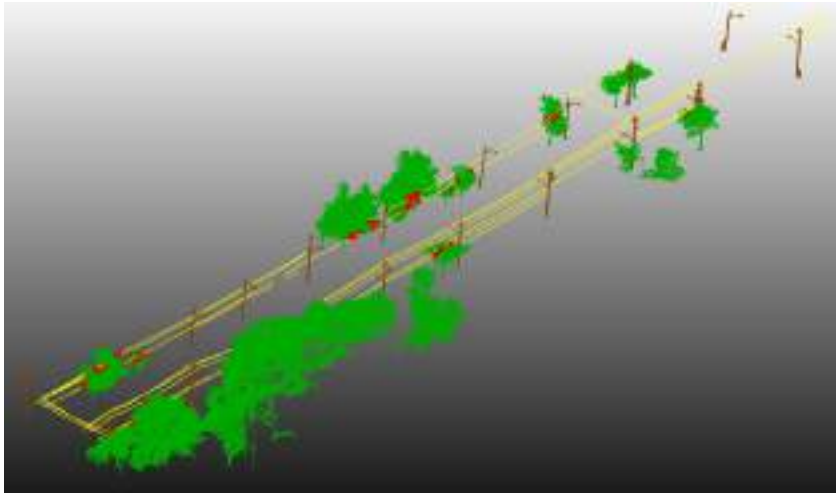
(3) Post-Processing: In the post-processing phase, DBSCAN clustering was utilized to refine the segmented point clouds by identifying and isolating critical features based on spatial density and predefined thresholds for height and point count. Parameters such as epsilon (maximum distance between samples) and the minimum number of samples were carefully calibrated to align with the attributes of the Toronto-3D dataset. Clusters exceeding specific thresholds, such as a height threshold of 8 meters for poles or a minimum point count of 500 points for vegetation, were retained, while irrelevant features and noise were filtered out. This approach ensured that significant urban elements were clearly distinguished for further analysis. To assess proximity risks, KDTree spatial queries were employed to calculate the distances between vegetation and power lines in a 3D space. These distances were compared against a safety threshold (e.g., 1 meter) to identify hazardous tree regions encroaching on power lines. These flagged areas were then prioritized for risk management, supporting effective vegetation maintenance and infrastructure safety. Figure 4(a) shows the post-processed semantic segmentation result, highlighting hazardous tree areas (red points) within the safety perimeter of power distribution lines.

(3) Digital Twin Development: For the utility lines, the implementation of the methodology was carried out using Python and its specialized libraries for point cloud processing. Segmented point cloud data was imported, and a KNN algorithm was applied to compute the 10 nearest neighbors for each point. Normal vectors were calculated by fitting a plane to these neighbors, and curvature was estimated by measuring deviations from this plane. To ensure continuity in the utility lines, RBF interpolation was employed to fill gaps in the point cloud. This process created a seamless structure by reconstructing missing data points. An upsampling step was then applied to increase point density in sparse regions, enhancing the model's resolution and uniformity. Finally, the segmented sections were integrated into a single, cohesive representation, ensuring geometric continuity.

The process for poles and trees involved aligning segmented point cloud data with reference 3D models using Python. Reference models were imported, and alignment began with the initialization of parameters based on approximate geometry and spatial data. A transformation matrix, incorporating rotation, translation, and scaling, was applied for preliminary alignment. The ICP algorithm was then used to iteratively refine these parameters, minimizing alignment errors and ensuring a precise fit. To ensure spatial and size accuracy, the bounding box centers of segmented objects and reference models were aligned, and a scaling factor was applied based on a comparison of their bounding box dimensions. Finally, the geometrical attributes of the created 3D objects are added to these objects. Figure 4(b) illustrates the digital twin model integrating the 3D models of trees, poles, and utility lines derived from the point cloud data.

5 Discussion

Digital twins offer a dynamic, continuously updated virtual representation of power lines, poles, and



(a) Post-processed result of semantic segmentation, highlighting hazardous tree areas near power lines



(b) Digital twin model

Figure 4. Digital twin model based on point cloud data

vegetation, that can be a part of a federated digital twin at the urban level. The work presented in this paper is limited by the fact that the digital twin developed so far is static. However, by integrating diverse data sources, including point clouds, GIS data, satellite imagery, and real-time weather updates, the digital twin can be further developed to provide a high-fidelity, dynamic model for vegetation management near power distribution networks. A critical advantage of this digital twin lies in its ability to simulate and predict how various factors, such as vegetation growth and adverse weather, affect overhead power lines. This predictive capability can be bolstered by deep learning (DL) algorithms, which enable accurate forecasting of vegetation dynamics, offering actionable insights for proactive management [24, 25]. Key attributes such as tree height, canopy spread, and species types are indispensable for evaluating risks to power infrastructure. Tree height helps identify potential threats from vegetation of comparable or greater height than poles, while canopy spread highlights horizontal encroachments even from distant trunks. Species type offers further context, enabling growth pattern predictions and customized maintenance strategies. Incorporating these attributes into the digital twin enhances risk assessment, providing utility managers with a robust framework for mitigating vegetation-related hazards.

Furthermore, this digital twin can integrate weather simulation to assess and address compound risks, such as the interplay of high winds and dense tree canopies, which exacerbate outage probabilities. By analyzing historical weather data and integrating live forecasts, this digital twin can dynamically update risk profiles, enabling timely interventions like proactive pruning or operational adjustments to minimize disruptions. This real-time adaptability transforms network management

from reactive to preventive, which can optimize resources [4, 5]. Moreover, this digital twin fosters organizational collaboration by serving as a centralized platform that unifies data across departments. This shared understanding of network conditions enhances communication and aligns teams toward common objectives, such as reliability and resilience. The capacity to simulate the outcomes of various vegetation management strategies, such as modified pruning schedules or low-maintenance species planting, supports sustainable practices that balance operational demands with environmental considerations.

While the creation and maintenance of digital twins require an initial investment, their unparalleled benefits justify the effort. Unlike static models, they provide not only visualization, but also actionable intelligence for strategic planning. For example, predictive growth models informed by species-specific data and environmental conditions (e.g., climate, soil) allow precise forecasts of tree encroachment on power lines. Adopting digital twins also opens avenues for sustainable vegetation management. By optimizing trimming schedules and employing selective pruning techniques, they minimize carbon emissions associated with vegetation maintenance. While traditional methods focus on clearing immediate risks, digital twins enable long-term planning that preserves larger trees while maintaining network safety. This approach aligns utility management goals with environmental stewardship, enhancing biodiversity and urban air quality [26, 27].

6 Conclusions and Future Work

This paper presents an integrated framework for computer vison-based digital twin for power distribution network management, focusing on vegetation management and risk assessment. The framework leverages the RandLA-Net model for semantic segmentation of power lines, poles, and vegetation, coupled with post-processing techniques such as DBSCAN clustering and rule-based thresholding to refine data outputs. Proximity detection using KDTree spatial analysis ensures precise identification of hazardous zones. The processed data is utilized to generate accurate 3D models, forming the foundation of a digital twin. By addressing key challenges such as data segmentation and geometric refinement, the framework enhances proactive vegetation management strategies and contributes to infrastructure resilience. The study demonstrates the potential of integrating computer vision and point cloud data processing techniques for developing digital twins for vegetation management near power distribution networks. The proposed framework effectively combines semantic segmentation, clustering, and proximity analysis to provide detailed 3D models for risk assessment and maintenance planning. The results highlight the scalability and reliability of the approach, offering a practical solution for improving vegetation management and infrastructure safety. These findings represent a significant step toward leveraging digital twin technology for utility management, with potential applications across a wide range of urban and industrial systems.

While the proposed framework offers significant advancements, several limitations remain. The framework has been tested primarily on the Toronto-3D dataset, which may not capture the diversity of urban and rural environments. Future work should involve testing on datasets with varying complexities and characteristics to enhance generalizability and robustness. Additionally, the RandLA-Net model and post-processing techniques, such as DBSCAN clustering and KDTree queries, require substantial computational resources. Optimization efforts are necessary to enable real-time processing and scalability for larger networks, especially in resource-constrained settings.

Beyond these points, as explained in Section 5, additional limitations include incorporating real-time and multi-source data streams, predictive vegetation growth modeling, and sustainability practices. Addressing these topics will be essential to fully realize the potential of the proposed framework for comprehensive risk assessment and more resilient power distribution systems.

Acknowledgments

This research is supported by a grant from Canada First Research Excellence Fund and additional funding from Hydro-Quebec.

References

[1] Hydro Quebec, Trees and power lines - Power-outages | Hydro-Québec, (2024). https://poweroutages.hydroquebec.com/poweroutages/understand-and-prevent/vegetation.html (accessed February 2, 2024).

[2] Cao, W., Wu, J., Shi, Y. and Chen, D., 2022. Restoration of individual tree missing point cloud based on local features of point cloud. *Remote Sensing*, 14(6), pp. 1346/1-20.

[3] Mongus, D., Brumen, M., Žlaus, D., Kohek, Š., Tomažič, R., Kerin, U. and Kolmanič, S., 2021. A Complete Environmental Intelligence System for LiDAR-Based Vegetation Management in Power-Line Corridors. *Remote Sensing*, *13*(24), pp. 5159/1-15.

[4] Dehghanian, P., Zhang, B., Dokic, T. and Kezunovic, M., 2018. Predictive risk analytics for weather-resilient operation of electric power systems. *IEEE Transactions on Sustainable Energy*, 10(1), pp. 3-15.

[5] Gao, P., Du, W., Lei, Q., Li, J., Zhang, S. and Li, N., 2023. NDVI forecasting model based on the combination of time series decomposition and CNN–LSTM. *Water Resources Management*, 37(4), pp. 1481-1497.

[6] Bahreini, F., Hammad, A. and Nik-Bakht, M., 2024. Point Cloud-based Computer Vision Framework for Detecting Proximity of Trees to Power Distribution Lines, *In Proceedings of the 41st International Symposium on Automation and Robotics in Construction, International Association on Automation and Robotics in Construction*, Lille, France, pp. 729-737.

[7] Glumova, E.S. and Filinskikh, A.D., 2020. Investigation of algorithms for generating surfaces of 3D models based on an unstructured point cloud. *In CPT2020 The 8th International Scientific Conference on Computing in Physics and Technology Proceedings*, Pushchino, Moscow region, pp. 287-293.

[8] Ismail, F.A., Shukor, S.A., Rahim, N.A. and Wong, R., 2023. Surface Reconstruction from Unstructured Point Cloud Data for Building Digital Twin. *International Journal of Advanced Computer Science and Applications*, 14(10), pp. 1-10.

[9] Tagarakis, A.C., Benos, L., Kyriakarakos, G., Pearson, S., Sørensen, C.G. and Bochtis, D., 2024. Digital Twins in Agriculture and Forestry: A Review. *Sensors*, 24(10), pp. 3117/1-26.

[10] Chen, C., Wang, H., Wang, D. and Wang, D., 2024. Towards the digital twin of urban forest: 3D modeling and parameterization of large-scale urban trees from close-range laser scanning. *International Journal of Applied Earth Observation and Geoinformation*, 127(1), pp. 103695/1-11.

[11] Zhang, W. and Li, W., 2024. Construction of Environment-Sensitive Digital Twin Plant Model for Ecological Indicators Analysis. *Journal of Digital Landscape Architecture*, 1(1), pp. 18-28.

[12] Gobeawan, L., Wise, D.J., Wong, S.T., Yee, A.T., Lim, C.W. and Su, Y., 2021. Tree species modelling for digital twin cities. *Transactions on Computational Science XXXVIII*, 12620(1), pp. 17-35.

[13] Buonocore, L., Yates, J. and Valentini, R., 2022. A proposal for a forest digital twin framework and its perspectives. *Forests*, 13(4), pp. 498.

[14] Kim, T., Cho, W., Matono, A. and Kim, K.S., 2020, November. PinSout: Automatic 3D indoor space construction from point clouds with deep learning. *In Proceedings of the 28th International Conference on Advances in Geographic Information Systems*, Seattle WA USA, pp. 211-214.

[15] La Guardia, M. and Koeva, M., 2023. Towards Digital Twinning on the Web: Heterogeneous 3D Data Fusion Based on Open-Source Structure. *Remote Sensing*, 15(3), pp. 721/1-13.

[16] Wu, Y., Shang, J. and Xue, F., 2021. Regard: Symmetry-based coarse registration of smartphone's colorful point clouds with CAD drawings for low-cost digital twin buildings. *Remote Sensing*, 13(10), pp. 1882/1-22.

[17] Agapaki, E. and Brilakis, I., 2022. Geometric digital twinning of industrial facilities: Retrieval of industrial shapes. *arXiv preprint arXiv:2202.04834*.

[18] Li, Y., Ma, L., Tan, W., Sun, C., Cao, D. and Li, J., 2020. GRNet: Geometric relation network for 3D object detection from point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, 165, pp. 43-53.

[19] Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z., Trigoni, N. and Markham, A., 2020. Randla-net: Efficient semantic segmentation of large-scale point clouds. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, Seattle*, WA, USA, pp. 11108-11117.

[20] Ahmed, K.N. and Razak, T.A., 2014. A comparative study of different density based spatial clustering algorithms. *International Journal of Computer Applications*, 975(1), pp. 8887/1-8.

[21] Sketchfab, Inc., 2024. https://sketchfab.com/ (accessed September 20, 2024).

[22] CGTrader, 2024. https://www.cgtrader.com/ (accessed September 20, 2024).

[23] Tan, W., Qin, N., Ma, L., Li, Y., Du, J., Cai, G., Yang, K. and Li, J., 2020. Toronto-3D: A large-scale mobile LiDAR dataset for semantic segmentation of urban roadways. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, Seattle, WA, USA, pp. 202-203.

[24] Heluany, J.B. and Gkioulos, V., 2024. A review on digital twins for power generation and distribution. *International Journal of Information Security*, 23(2), pp. 1171-1195.

[25] Qi, D., Xi, X., Tang, Y., Zheng, Y. and Guo, Z., 2024. Real-time scheduling of power grid digital twin tasks in cloud via deep reinforcement learning. *Journal of Cloud Computing*, 13(1), p. 121/1-12.

[26] Stephenson, N.L., Das, A.J., Condit, R., Russo, S.E., Baker, P.J., Beckman, N.G., Coomes, D.A., Lines, E.R., Morris, W.K., Rüger, N. and Álvarez, E., 2014. Rate of tree carbon accumulation increases continuously with tree size. *Nature*, 507(7490), pp. 90-93.

[27] Petri, A.C., Koeser, A.K., Lovell, S.T. and Ingram, D., 2016. How green are trees? Using life cycle assessment methods to assess net environmental benefits. *Journal of Environmental Horticulture*, 34(4), pp. 101-110.

Simulation-based Integrated Transportation Planning in Panelized Construction

Ahmed Zaalouk [1], Mohammed Sadiq Altaf [2] and SangHyeok Han [3]

[1] Ph.D. Candidate, Department of Building, Civil and Environmental Engineering, Concordia University, Montréal, QC, Canada.
[2] Technical Supervisor, ACQBUILT Inc., 4303 55 Ave NW, Edmonton, AB, T6B 3S8, Canada.
[3] Associate Professor and Co-Director, Centre for Innovation in Construction and Infrastructure Engineering and Management (CICIEM), Department of Building, Civil and Environmental Engineering, Concordia University, Montréal, QC, Canada.
Email: ahmed.zaalouk@concordia.ca, sadiq@promiserobotics.com, sanghyeok.han@concordia.ca

Abstract –

Panelized construction is an offsite construction approach that offers enhanced design flexibility and cost-effective assembly. Despite these advantages, it faces challenges due to a fragmented supply chain (SC) and transportation coordination issues that can lead to cost overruns and schedule delays. To address these challenges, this study introduces a transportation planning method that leverages Just-In-Time principles and hybrid multi-agent simulation to align factory and onsite operations across multiple projects through integrated transportation procedures. These procedures categorize SC flows into: (1) a dispatching flow, which manages trailer loading and dispatch readiness; and (2) a reverse flow, which ensures the timely retrieval of empty trailers to maintain continuous production. These flows form the core of an innovative resource coordination strategy that dynamically allocates resources, including trailers and trucks, to ensure an uninterrupted supply of panels from factory to sites. The validation results using operational data from a residential panel manufacturer illustrate the model's effectiveness in coordinating SC operations and resources. The model successfully adapts to fluctuating onsite demands and aligns transportation operations with delivery dates, matching real operational timelines. The model also manages limited production capacities efficiently and ensures precise trailer dispatch, effective truck assignments, and timely retrievals of empty trailers from different sites. The proposed simulation-based planning model provides an integrated and automated solution to improve SC performance, supporting efficient multi-production lines and multi-project operations in panelized construction.

Keywords –

Offsite Construction; Panelized Construction; Supply Chain Management; Integrated Transportation Planning; Hybrid Simulation Model; Agent-Based Modeling; Discrete-event Simulation; Just-In-Time Delivery.

1 Introduction

In panelized construction, floor, wall, and roof panels are produced in a factory-controlled environment and then transported to project sites for final assembly [1]. This approach provides several advantages compared to volumetric modular construction, such as easier handling of prefabricated components (since the components are smaller and lighter), a higher degree design flexibility, and reduced crane capacity required for onsite assembly [2]. However, the adoption of panelized construction faces challenges with respect to supply chain management (SCM), integrated planning, transportation logistics, and storage constraints (both at the factory and on site). These issues can disrupt the coordination between production, transportation, and onsite assembly, leading to increased project costs and extended schedules [1–3]. To address these challenges, efficient transportation planning is essential. Transportation planning facilitates the integration of factory production lines with project sites and manages the delivery of loaded trailers to sites and the return of empty trailers to the factory (i.e., distribution and reverse logistics). The transportation process bridges factory and onsite operations by ensuring the timely delivery of panels to project sites—a vital function for maintaining the continuous progress of onsite activities. Disruptions in transportation operations, such as delays in trailer distribution or inefficiencies in reverse logistics, can lead to idle resources and delays in onsite schedules. Moreover, factory production may be interrupted due to the unavailability of empty trailers (which are needed for loading manufactured panels). Thus, effective transportation planning is crucial for minimizing supply chain (SC) operational time and costs while enhancing the performance of both factory and project site operations.

Studies on offsite construction (OSC) transportation planning and logistics have primarily focused on individual aspects of the SC, often addressing either factory operations or construction site requirements [4,5]. While some studies

have considered integrated planning [6], these have not accounted for the unique transportation challenges in panelized construction, particularly the coordination of individual panels produced in multiple production lines, the various trailer types, operational-level planning of SC tasks, and just-in-time (JIT) delivery (as opposed to using intermediate storage space between factory and site).

Existing studies have explored methods for optimizing transportation in OSC contexts such as ready-mixed concrete, precast components, and modular construction [7–10]. These works have addressed diverse topics related to transportation planning in OSC, including optimal vehicle routing and the development of stacking plans [11,12]. However, the unique needs of panelized construction, such as integrating distribution and reverse logistics and the treatment of trucks and trailers as separate units performing distinct tasks, have not been accommodated in these studies.

These limitations highlight the following critical gaps in the literature: (1) existing studies have not considered the needs of factory and project sites simultaneously, which is essential for integrated SC operational performance; (2) reverse logistics, a key component of panelized construction, has not received adequate attention (with this gap resulting in inefficiencies in trailer utilization and delays in production schedules, ultimately disrupting the SC flow); and (3) existing studies have typically assumed the availability of sufficient onsite storage, which is unrealistic for panelized construction projects (in which, ideally, JIT delivery is employed to address limited space and assembly sequencing requirements). Additionally, prior studies, such as Hsu et al. (2018) [6], have focused on truck-based transportation planning while overlooking the need to manage trailers separately from trucks (which, in turn, introduces additional coordination challenges, as trailers have distinct dispatching and returning operations that directly affect SC efficiency).

To address these gaps, the present study develops an integrated transportation planning framework tailored to panelized construction. The aim underlying the framework development is to coordinate factory production with onsite requirements, ensuring efficient trailer distribution and return operations while minimizing idle time for both transportation fleets and construction resources. Unlike previous approaches, this framework encompasses trailer prioritization based on onsite assembly sequencing and crane availability, ensuring efficient resource allocation and uninterrupted SC flow. By aligning transportation operations with factory production and onsite requirements, the proposed framework enhances SC performance, reduces project delays, and improves overall SCM efficiency.

2 Methodology

As illustrated in Figure 1, the proposed framework consists of the following four steps: (1) define SC operational objectives and constraints; (2) develop integrated transportation procedures; (3) design and build the simulation model; and (4) conduct model validation and output evaluation. The first step identifies goals such as synchronizing factory production with onsite schedules and minimizing delays, inventory, and inefficiencies. In the second step, a dispatching flow is established to ensure timely trailer loading and dispatch based on onsite requirements, along with a reverse flow for the efficient retrieval of empty trailers as a means of maintaining production continuity. In the third step, Agent-based Modeling (ABM) is implemented in the simulation model to represent factory production lines, trailers, trucks, and onsite operations. Moreover, discrete-event simulation (DES) is used to model workflows inside the agents, and GIS integration is used as a shared agent environment, facilitating resource agent movement (e.g., truck movement) within the SC. The model incorporates a GIS city map, enabling agents to navigate SC locations—factory and project sites—positioned using geographic coordinates. Agents move along transportation routes, with travel times determined based on real distances, speed limits, and road network constraints. This GIS functionality supports route selection and scheduling by incorporating distance-based travel times and automatically determining the shortest route for transportation resources, ensuring a realistic representation of transportation planning. Finally, the robustness and accuracy of the model are ensured through verification using artificial datasets and validation using real operational data to evaluate JIT adherence, resource utilization and efficiency, operations timelines, and synchronized SC performance.

To efficiently integrate transportation operations with other SC activities involved in the production and onsite assembly stages, integrated transportation procedures for SCM in panelized construction are first developed. These procedures constitute a systematic approach to managing a multi-unit fleet of trucks and trailers during a typical working day. This approach also defines critical decision points such as trailer prioritization, onsite sequence checks, and operational constraints (e.g., panel delivery dates, truck availability). Unlike traditional truck-based approaches, this method treats trailers as independent transportation units requiring distinct dispatching and returning operations. By dynamically prioritizing trailer dispatch based on onsite assembly sequencing and crane availability, the method enhances resource utilization and minimizes SC waste (i.e., resource idle time). By aligning transportation with production and onsite demands, these procedures ensure adherence to JIT delivery principles.

Figure 2 presents the developed pseudocode, which captures these transportation procedures in detail through two primary flows-dispatching flow and reverse flow.

(1) Dispatching flow—managing the fleet by checking loaded trailer availability and panel delivery dates, prioritizing trailers that contain panels required earlier in the onsite sequence, and ensuring trailers are assigned to sites with idle cranes to reduce waiting times. This trailer prioritization strategy prevents bottlenecks and ensures cranes remain productive, improving overall SC performance. Dispatching flow also assigns trucks for dispatch, and verifies sequence readiness at the site before delivery. Once dispatched, trucks deliver the loaded trailers to the project site and return to the factory for subsequent tasks.

(2) Reverse flow—retrieving empty trailers during a truck's return from dispatch or through an independent retrieval process triggered when the factory's trailer buffer falls below a threshold. This mechanism prevents production disruptions from trailer shortages and maintains fleet circulation, enhancing SC efficiency.

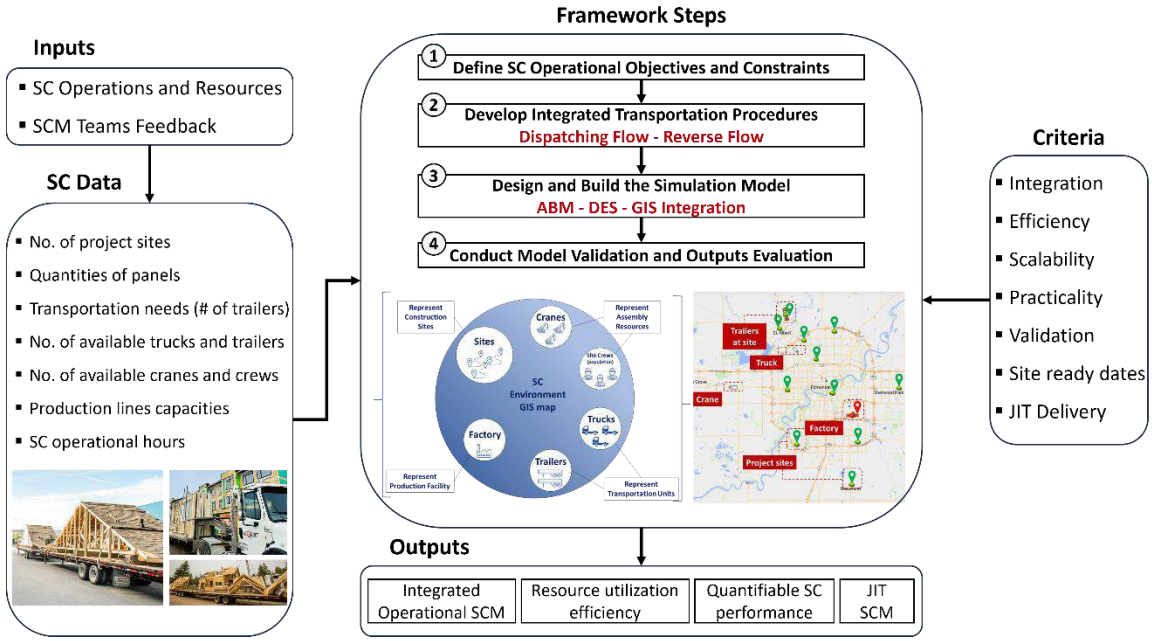


Figure 1. Overview of the research framework

The developed transportation procedures serve as the foundation for the development of the simulation's governing logic, ensuring seamless coordination between factory, transportation, and onsite operations. The simulation model integrates SC operations by leveraging ABM and DES to plan transportation-oriented panelized construction processes. This simulation models SC resources and entities as agents communicating and executing tasks based on specific production and logistics requirements within a predefined planning period. Figure 3 presents a sequence diagram illustrating the interactions among key agents: (1) the factory, which processes production and logistics requests, assigns production lines, manages trailers, trucks, cranes, and crews, and dynamically coordinates both dispatching and reverse flows; (2) the project site, which generates logistics requests, monitors deliveries, completes the onsite assembly, and manages the marking and release of empty trailers for retrieval; (3) trailer and truck, which facilitate the transportation of panels to project sites and the retrieval of empty trailers to the factory; and (4) crane and crew, which execute onsite tasks synchronized with production and transportation progress.

The simulation model outlines two primary transportation operations: (i) dispatching operations, initiated when production is completed, panels are loaded into trailers, and the panel delivery date comes due (based on the simulation time); and (ii) returning operations, activated post-onsite assembly, where trailers are marked as empty, and the factory assigns trucks for retrieval. In the developed simulation model, the factory agent plays a pivotal role in managing logistics operations to ensure the completion of all SC activities. The factory agent receives production requests (i.e., information about the quantities of panels and the location of the delivery site) and logistics requests (i.e., information about the crane type required and the number of crews) from project site agents. Based on these requests, the factory agent ensures the assignment of cranes and crews before establishing any dispatching to sites.

The factory agent processes the requests received using DES models that depict production lines and logistics workflows. These DES models actively communicate with truck and trailer agents by monitoring the behavior charts (i.e., statecharts) that show their operational status within the SC. Figure 4 illustrates the

statecharts of trucks and trailers, depicting their state transitions during key operations such as loading, dispatching, and returning. These charts detail the interactions and transitions for trucks and trailers, and they are dynamically updated as the simulation progresses. Trailers progress through three primary states: (i) (AtFactory), representing empty trailers available for loading; (ii) (TrailerLoaded), indicating trailers loaded and delivered to construction sites; and (iii) (TrailerUnLoaded), referring to trailers that have completed their deliveries and are awaiting return to the factory. To minimize delays and support uninterrupted production, the factory agent ensures the prompt return of trailers, following the reverse flow procedures outlined in Figure 2.

Similarly, trucks transition between two states: (i) (TruckAvailable), indicating readiness for new transport assignments; and (ii) (TruckUnavailable), signifying active dispatch or return operations. By incorporating the transportation procedures, agent requirements, and statecharts reflecting resource operational conditions, the simulation achieves seamless and efficient transportation cycles throughout the planning horizon.

```
# Pseudocode: Integrated Transportation Procedures

# Dispatching Flow
START
WHILE working_day_active:
    IF loaded_trailers_available AND panel_delivery_date_is_due:  # Check if trailers and dates are aligned
        IF several_trailers_in_yard:  # Decision point for prioritizing trailers
            prioritize_loaded_trailers

        IF truck_available:  # Availability check for truck
            capture_truck
            select_trailer_for_dispatch_based_on_priority_rules
            IF previous_trailers_in_sequence_on_site:  # Ensure sequence is complete
                connect_trailer_and_move_to_site
            ELSE:
                wait_for_sequence

            disconnect_truck_at_site
            return_truck_to_factory  # Reverse flow logic applies during return
        ELSE:
            wait_for_truck
    ELSE:
        trailers_wait_in_yard

# Reverse Flow
    # During Dispatch
    IF empty_trailers_needed_at_factory AND truck_is_returning:  # During truck return, check for empty trailers
        IF empty_trailer_at_site after delivery:
            connect_trailer_to_truck_and_move_to_factory
            release_truck_and_trailer_at_factory
        ELSE:
            check_other_sites_for_empty_trailers
            IF empty_trailer_found:
                move_truck_to_nearest_site_and_connect_trailer_to_truck
                move_truck_to_factory
                release_truck_and_trailer_at_factory

    # If factory runs out of trailers
    # For when dispatching is done, and trailers are still needed at the factory.
    IF factory_trailer_buffer_below_threshold:  # Independent retrieval when factory is short on trailers
        check_sites_for_empty_trailers
        IF empty_trailer_found:
            capture_truck
            move_truck_to_nearest_site_and_connect_trailer_to_truck
            move_truck_to_factory
            release_truck_and_trailer_at_factory

    # End conditions
    IF all_trailers_returned OR end_of_working_day:  # Conclude operations if no further tasks remain
        END
```

Figure 2. Pseudocode for integrated transportation procedures in panelized construction

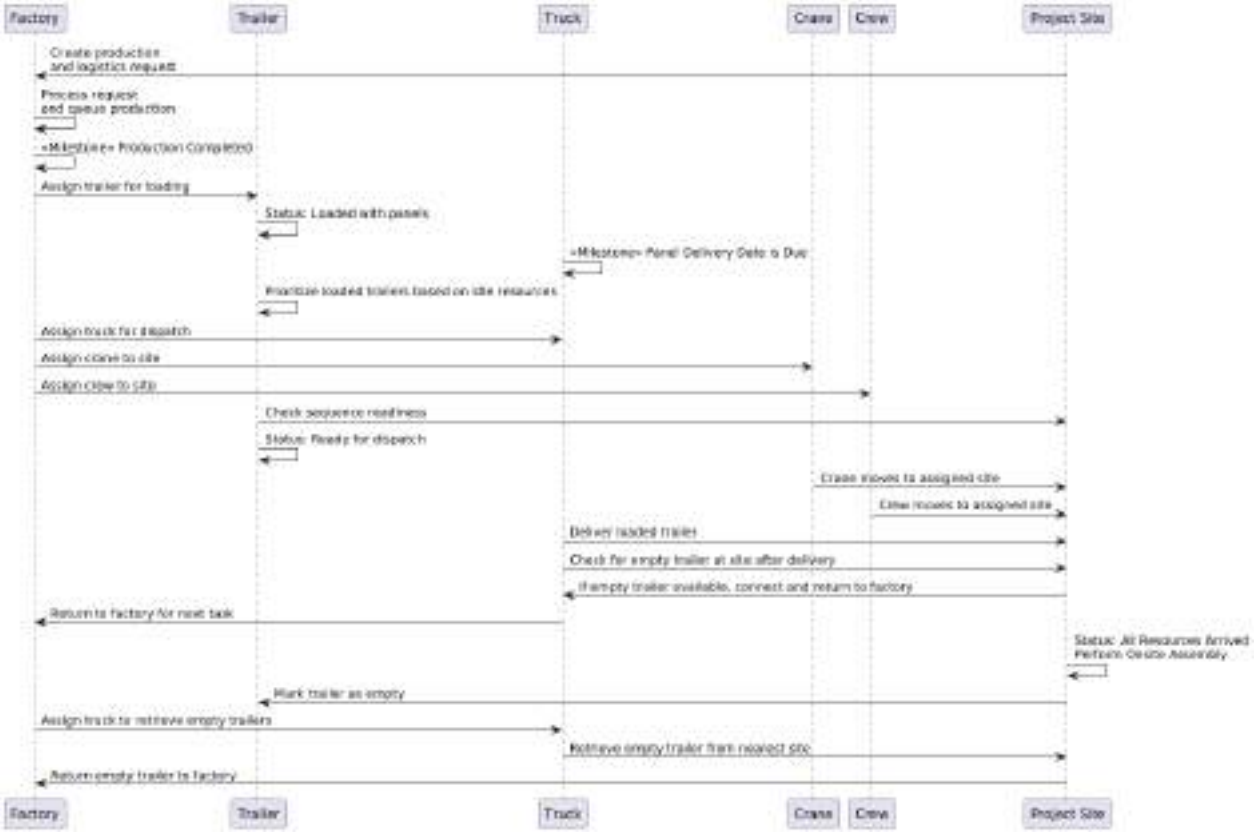


Figure 3. Sequence diagram of agent interactions in integrated SC operations

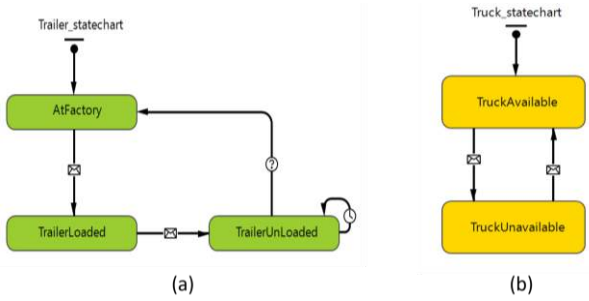


Figure 4. Statechart of the transportation agents: (a) trailer agent; (b) truck agent

3 Implementation and Results

To evaluate the proposed integrated transportation procedures, a case study was conducted using operational data from ACQBUILT, an Alberta-based prefabricated panel manufacturer. The case company produces floor, wall, and roof panels in a factory-controlled environment and transports them to project sites for assembly. Despite advancements in production automation, the case company's SCM remains manual, lacking integration between production and onsite operations. It relies on push planning, producing panels early regardless of onsite delivery dates, resulting in excess inventory and inefficiencies. The proposed transportation procedures ensure timely delivery of panels by bridging factory and onsite operations. A simulation model was developed to test the proposed procedures under realistic conditions in order to demonstrate their ability to connect factory production, transportation, and onsite assembly for efficient SC performance.

To assess the effectiveness of the proposed framework, several verification and validation steps were conducted. Test cases using artificial datasets were developed to confirm agent interactions and workflows, ensuring the procedures functioned as intended. Additionally, the movement of transportation resources on a GIS map was visually checked to verify that trucks were traveling to the required destinations while adhering to operational hours and taking the shortest available route. Real operational data and project details from April, 2021, was also used to validate the ability of the proposed framework to align with actual production progress and meet onsite requirements. During April, 2021, panels for 35 distinct home construction projects needed to be produced, transported and assembled on site. These projects varied in design complexity, panel sizes, and panel quantities, leading to differences in transportation scheduling, trailer allocation, and onsite

sequencing. The proposed framework manages variations through independent project site agents, each autonomously handling its panel orders, trailer arrivals, and resource coordination based on project-specific parameters. This structure ensures that each site dynamically adapts its logistics based on actual SC conditions, supporting multi-project scalability. With respect to the case implementation, the simulation model used the same production capacities and SC resources as those established by the partner in reality for the same period (April, 2021). While predefined factory capacities, working hours, and resource availabilities serve as operational constraints in the proposed framework, the scheduling process itself is automated. The framework dynamically extracts actual job details and available resources from the enterprise resource planning system (ERP) and updates scheduling parameters accordingly. A dedicated user interface (UI) extracts data and integrates it into a built-in database (DB) within the simulation that structures SC configurations and job details. These automated updates ensure that each scheduling horizon reflects the latest available data without requiring manual intervention beyond predefined constraints. To ensure data reliability, extracted datasets undergo a structured verification process before integration into the simulation model. Before integration, extracted data is validated in structured Excel files to flag inconsistencies. Missing values either are corrected through cross-validation with ERP records or are replaced with default estimates based on historical data to maintain scheduling feasibility.

The purpose of the validation was to evaluate the performance of the proposed framework against operational data for manufacturing, delivery, and onsite assembly. Metrics such as delivery times and adherence to JIT delivery were used to assess the successful coordination of SC processes. Expert feedback further confirmed the practicality of the generated SC operational plans and the effectiveness of the developed model in minimizing delays in dispatching and returning operations and enhancing resource allocation, thus achieving integrated SC performance. It should be noted that, given its structured database architecture and ERP-based data extraction, the framework can be adapted to other OSC companies with similar SC structures, ensuring flexibility in different operational environments.

The first verification step was to ensure that the model accurately manages production operations in response to onsite demand. The factory agent should be able to receive and record all required panel quantities and then manage production according to factory procedures. Moreover, the factory agent should adhere to internal constraints (e.g., production line capacities and operational days). For instance, as shown in Figure 5, on April 15, 2021, a peak daily request of approximately 20,000 sq ft was recorded, exceeding the production line's capacity of 12,000 sq ft/day. Despite the high demand, the model ensured that production remained within capacity by deferring excess quantities to subsequent operational days. Non-operational days, such as April 11, 18, and 25, were accurately represented with zero production, reflecting compliance with predefined factory capacities. Another verification aspect was to ensure model accuracy in terms of the total quantities received from the sites being equal to the total amounts produced by the factory at the end of the planning period. As shown in Figure 5, cumulative quantities confirmed that the wall production line successfully fulfilled all orders received, with the cumulative received and produced quantities reaching the same number, 168,680 sq ft, by April 28, 2021. These results underscore the model's ability to dynamically adjust production to meet varying demands while maintaining efficiency and strict adherence to operational constraints.

After verifying the production performance, the focus shifted to evaluating the transportation procedures to confirm their effectiveness in seamlessly linking production and onsite activities. The factory uses two types of trailers: flatbed trailers for loading floor and roof panels and vertical trailers for wall panels. In the model, "TA 1-TA 35" represents flatbed trailers while "TA36-TA55" refers to vertical trailers. As shown in Table 1, the model assigns trailers based on panel type and initiates dispatching operations efficiently. For instance, Trailer 38 was loaded with Wall0 on April 12 at 10:51. However, Truck 1 was assigned (and dispatching occurred) on April 15 to ensure delivery on the same day panels were required on site, with arrival at the site on April 15 at 5:39 and assembly completed by 8:30. As this example demonstrates, the model was found to adhere to JIT principles, ensuring trailers and trucks arrive onsite before assembly begins and promptly return empty trailers after assembly. Regarding returning operations, there are two notable patterns: daytime returns to meet immediate factory needs and end-of-day returns for flexibility. Dispatching operations are prioritized to ensure timely delivery to sites and to maintain continuity in onsite assembly activities by avoiding crane idle times.

Finally, the model's ability to manage the assembly of the same number of jobs completed in reality was validated by comparing assembly quantities between actual SC performance and simulated results. Although daily assembled quantities were found to fluctuate, the model completed all assembly operations on the same day as the actual schedule (April 30, 2021). Additionally, the total cumulative quantities (58,600 sq ft) in reality were the same as in the simulation, demonstrating the model's accuracy in completing all necessary onsite tasks. These results validated the model's ability to synchronize dispatching, onsite activities, and trailer returns, ensuring JIT delivery and maintaining alignment across the SC.

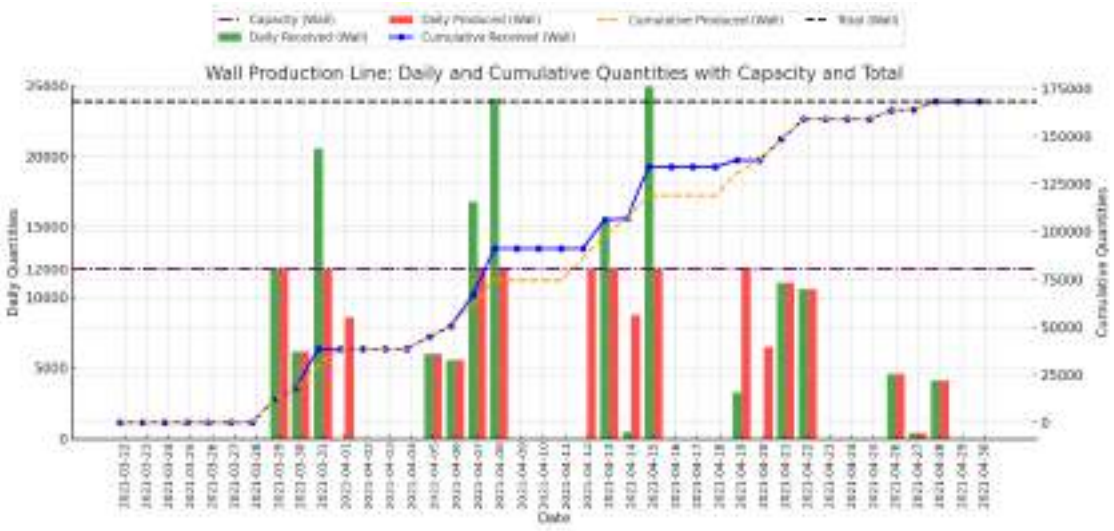


Figure 5. Simulation results for wall production line received and produced quantities

Table 1. Example of trailers and truck assignments during dispatching and returning operations

Job details			Dispatch				Onsite	Return	
#	stage	PT	LC	TA #	TU #	AS	AC	TU #	RT
1	Capping	Wall0	12-04-10:51	TA 38	TU 1	15-04-5:39	15-04-8:30	TU 3	15-04-19:30
1	Capping	Floor1	12-04-13:35	TA 22	TU 3	15-04-5:51	15-04-11:01	TU 3	15-04-20:29
1	Lifting	Wall1	20-04-10:33	TA 38	TU 3	21-04-5:51	21-04-7:12	TU 2	21-04-10:14
1	Lifting	Wall1	20-04-11:55	TA 41	TU 1	21-04-6:27	21-04-8:20	TU 3	21-04-10:32
1	Lifting	Roof	19-04-13:00	TA 25	TU 2	21-04-9:39	21-04-11:01	TU 2	21-04-11:24
1	Lifting	Roof	19-04-13:30	TA 11	TU 3	21-04-9:57	21-04-12:08	TU 2	21-04-12:35
1	Lifting	Roof	19-04-14:00	TA 13	TU 1	21-04-10:31	21-04-13:16	TU 3	21-04-19:17
2	Capping	Floor1	13-04-14:05	TA 10	TU 1	16-04-5:29	16-04-11:00	TU 3	16-04-19:44
2	Lifting	Wall1	22-04-13:02	TA 53	TU 2	27-04-5:35	27-04-7:00	TU 3	27-04-7:20
2	Lifting	Floor2	23-04-12:38	TA 23	TU 2	27-04-6:25	27-04-8:01	TU 3	27-04-8:10
2	Lifting	Wall2	22-04-15:19	TA 50	TU 3	27-04-6:55	27-04-9:04	TU 1	27-04-13:20
2	Lifting	Roof	26-04-9:45	TA 26	TU 3	27-04-7:45	27-04-10:05	TU 1	27-04-14:05
2	Lifting	Roof	26-04-10:15	TA 9	TU 1	27-04-12:55	27-04-14:10	TU 1	27-04-14:55
2	Lifting	Roof	26-04-10:45	TA 16	TU 1	27-04-13:40	27-04-15:11	TU 1	27-04-18:17

*panel_type (PT); load_complete (LC); arrival_site (AS); assembly_complete (AC); return_trailer (RT); trailer (TA); truck (TU)

4 Conclusion

This study introduces a simulation-based framework designed to address the complex challenges of SCM in panelized construction. It integrates production, transportation, and onsite assembly using JIT delivery principles and hybrid multi-agent simulation, enabling synchronized SC operations. The proposed framework develops integrated transportation procedures that plan dispatching and reverse flows in the SC, emphasizing coordinated operations and resource allocation.

The contributions of this study are as follows. First, a synchronized transportation planning framework tailored to panelized construction is proposed. This framework uniquely addresses the dual objectives of improving factory and onsite operations by considering trucks and trailers as distinct entities with different tasks. By integrating distribution and reverse logistics with JIT, it resolves inefficiencies in trailer utilization, mitigates delays, and aligns transportation flows with onsite needs, enhancing overall SC performance. Second, a detailed approach to simulation of SC workflows, including dispatching and reverse logistics, is developed. The model captures interactions among trailers, trucks, cranes, and crews to address limited onsite storage and improve resource allocation. By focusing on actionable strategies, such as adhering to onsite delivery dates, the framework provides practical insights for mitigating SC disruptions. Third, a practical simulation-based SCM approach is proposed that enables the automated generation of SC plans, ensuring dynamic coordination between production, transportation, and onsite assembly. It also addresses SCM challenges such as onsite storage constraints by applying JIT, ensuring minimal waste and efficient resource utilization. Moreover, the proposed transportation planning procedures introduce an algorithmic innovation by integrating rule-based

heuristics for trailer prioritization, dynamically aligning dispatching and retrieval decisions with onsite SC conditions to improve resource allocation and minimize idle time. Additionally, the framework supports operational decision-making, empowering stakeholders to test various SCM scenarios, thereby ensuring adaptability to onsite-specific challenges and constraints.

While the case study presented herein demonstrated the framework's ability to coordinate multi-project operations, handling logistics for 35 distinct sites with varying resource requirements, future research could be undertaken to further validate the model's scalability by applying it to a more diverse project environment, such as multi-manufacturer settings or projects with different types (e.g., commercial projects). Moreover, the model's application could be expanded to assess its adaptability to additional logistical constraints and SC configurations.

Furthermore, while the proposed simulation model effectively coordinates transportation and production using predefined decision rules and heuristics, further enhancements could be made by incorporating optimization-driven decision-making. The integration of metaheuristic algorithms (e.g., genetic algorithms) could further improve resource allocation by, for instance, optimizing the balance between owned and rented transportation resources, ensuring cost-effective fleet management while maintaining operational flexibility. Additionally, stochastic factors could be incorporated improve the model's ability to account for uncertainties in SC performance, particularly fluctuations in transportation availability, onsite demand variations, and operational constraints. These advancements could contribute to further reductions in operational costs and scheduling delays while improving overall SC resilience. The framework could also be extended to evaluate alternative transportation methods, such as rail, ship, or autonomous vehicles, by introducing corresponding agents into the simulation model. This extension would allow for the impact of different transportation strategies on the efficiency and cost-effectiveness of the SC to be analyzed. Furthermore, sustainability metrics could be integrated to evaluate the environmental impact of different transportation strategies, providing insights into reducing carbon emissions and improving the sustainable performance of SC operations.

References

[1] M. S. Altaf, A. Bouferguene, H. Liu, M. Al-Hussein, and H. Yu. Integrated production planning and control system for a panelized home prefabrication facility using simulation and RFID. *Automation in Construction*, 85:369–383, 2018.

[2] S. J. Ahn, S. Han, M. S. Altaf, and M. Al-Hussein. Integrating off-site and on-site panelized construction schedules using fleet dispatching. *Automation in Construction*, 137:104201, 2022.

[3] A. Zaalouk, S. Moon, and S. H. Han. Operations planning and scheduling in off-site construction supply chain management: Scope definition and future directions. *Automation in Construction*, 153:104952, 2023.

[4] J.-Q. Li, Y.-Q. Han, P.-Y. Duan, Y.-Y. Han, B. Niu, C.-D. Li, Z.-X. Zheng, and Y.-P. Liu. Meta-heuristic algorithm for solving vehicle routing problems with time windows and synchronized visit constraints in prefabricated systems. *Journal of Cleaner Production*, 250:119464, 2020.

[5] C. Zou, J. Zhu, S. Ma, K. Lou, N. Lu, and L. Li. Optimal transportation scheduling of prefabricated components based on improved hybrid differential firefly algorithm. *Mathematical Problems in Engineering*, 2022(1):3302983, 2022.

[6] P. Y. Hsu, P. Angeloudis, and M. Aurisicchio. Optimal logistics planning for modular construction using two-stage stochastic programming. *Automation in Construction*, 94:47–61, 2018.

[7] P. V. H. Son and H. T. Hieu. Logistics model for precast concrete components using novel hybrid Ant Lion Optimizer (ALO) algorithm. *International Journal of Construction Management*, 23(9):1560–1570, 2023.

[8] Z. Liu, Y. Zhang, M. Yu, and X. Zhou. Heuristic algorithm for ready-mixed concrete plant scheduling with multiple mixers. *Automation in Construction*, 84:1–13, 2017.

[9] Z. Liu, Y. Zhang, and M. Li. Integrated scheduling of ready-mixed concrete production and delivery. *Automation in Construction*, 48:31–43, 2014.

[10] M. Hussein, A. Darko, A. E. E. Eltoukhy, and T. Zayed. Sustainable logistics planning in modular integrated construction using multimethod simulation and Taguchi approach. *Journal of Construction Engineering and Management*, 148(6):04022022, 2022.

[11] W. Yi, H. Wang, L. Zhen, and Y. Liu. Automated generation of horizontal precast slab stacking plans. *Journal of Construction Engineering and Management*, 149(12):04023121, 2023.

[12] A. E. E. Eltoukhy, M. Hussein, M. Xu, and F. T. S. Chan. Data-driven game-theoretic model based on blockchain for managing resource allocation and vehicle routing in modular integrated construction. *International Journal of Production Research*, 61(13):4472–4502, 2023.

Building Component Segmentation-oriented Indoor Localization using BIM-based Synthetic Data Generation and Deep Learning

Haomiao Zhao[1], Yifeng Sun[2], Chi Chiu Lam[2], Junxiang Zhu[3], Mi Pan[2], Mun On Wong[2,*]

[1] School of Computer Science and Engineering, University of New South Wales, Sydney, Australia
[2] Department of Civil and Environmental Engineering, University of Macau, Macao SAR, China
[3]Department of Engineering, University of Cambridge, Cambridge, United Kingdom
[*]Corresponding author.
z5521569@ad.unsw.edu.au, mc35239@um.edu.mo, fstccl@um.edu.mo, jz652@cam.ac.uk, mipan@um.edu.mo, mowong@um.edu.mo

Abstract

This paper presents a novel image-based indoor localization approach that leverages semantic segmentation techniques by combining synthetic images generated from building information modelling (BIM) and deep learning (DL) methods. Unlike traditional image-based methods that compare overall visual features, our approach focuses on stable building components—such as floors, ceilings, walls, doors, and windows—that remain largely consistent over time. To enhance localization accuracy, we introduce a new similarity metric based on the intersection-over-union (IoU) ratio for each building component. This metric accounts for the spatial distribution of building elements, allowing for more robust comparisons between semantic masks derived from real-captured images and those generated from synthetic BIM models. The method reduces the dependence on detailed BIM models, enhancing reliability by focusing on the geometric and semantic consistency of key building elements. The feasibility of the approach is demonstrated through a case study with an institutional building, showcasing promising localization performance despite low-fidelity BIM models.

Keywords –

Indoor Localization; Semantic Segmentation; Building Information Modeling; Deep Learning; Image Retrieval; Synthetic Data Generation

1 Introduction and Related Works

Accurate indoor localization plays a crucial role in enabling a wide range of applications such as facility management, disaster response, and navigational assistance in complex indoor environments [1]. In recent years, various indoor localization methods have been developed, such as dead reckoning approaches with inertial sensors and piloting approaches using range-based signals [2-3]. However, these methods have their own constraints. For instance, range-based signals such as Wi-Fi, Bluetooth, and ultra-wideband signals are highly susceptible to environmental interference (e.g., walls) and require additional deployment of hardware infrastructure within buildings. Inertial measurement unit (IMU)-based methods suffer from cumulative drift errors, leading to decreased positioning accuracy over time [4]. In contrast, image-based localization methods offer a promising alternative, as they have a lower reliance on specialized hardware and can provide discrete, location-referenced measurements [5].

Traditional image-based indoor localization methods typically incorporate an offline database collection and an online image retrieval process [6]. During offline collection, various views of indoor environments need to be captured to establish an image database along with the associated locations and orientations. Subsequently, the online image retrieval is performed to leverage a real-captured query image to search the database for the most visually similar match, from which the position can be inferred. Nevertheless, the database collection process demands extensive preparatory work, and the effectiveness of the database may be affected over time when the indoor environment evolves, and the visual characteristics of certain areas change significantly [7].

To reduce the manual effort in data collection, recent research has explored building information modelling (BIM) to generate synthetic images in an automated and controllable manner. For instance, Ha et al. [8] generated the BIM-rendered images and then utilized a convolutional neural network (i.e., VGG 16 [9]) to measure the similarity between the real-captured images and the BIM-rendered images for indoor localization.

Similarly, in the study by Acharya et al. [10], the authors used 3D indoor models to generate rendered synthetic images and perform similarity comparison on a modified GoogLeNet-based deep convolutional neural network architecture (i.e., PoseNet [11]). These studies showcase the benefits of BIMs for image database collection, such as rapid image generation and customizable camera configurations like positions, orientation, and focal lengths. Despite these advantages, these BIM-derived methods have their limitations. To ensure the effectiveness of the similarity calculation between the real-captured and BIM-rendered images, these methods rely heavily on the high fidelity of BIM models, such as the inclusion of furniture, fixtures, and decorations. Such a prerequisite, in turn, requires certain human efforts to maintain the similarity of visual appearance continuously and meticulously. Any discrepancy, caused by insufficient detail in modelling or significant changes in actual environments, can severely degrade the positioning accuracy.

To address these challenges, this study proposes a semantic segmentation-oriented indoor localization approach by leveraging synthetic BIM image generation and deep learning (DL) techniques. Instead of comparing the holistic visual similarity of environments, the proposed approach concentrates on fundamental building components, i.e., floors, ceilings, walls, doors, and windows, since these components usually remain stable regardless of the interior changes and thus provide robust referencing patterns for similarity computation. In this way, the approach reduces the dependency on the highly detailed BIM models. The objectives of this study are: (1) to create an automated approach to generate the semantic masks of BIM-rendered images and real-captured images for similarity computation; (2) to develop a semantic segmentation-oriented similarity metric to implement image retrieval; and (3) to investigate the feasibility and quantify the performance of the proposed approach using a real-world case study. The detailed method development and validation are discussed in the following sections.

2 Methodology

The proposed approach consists of three key stages, including BIM-based synthetic image generation, DL-based image segmentation for real captures, and similarity-based image retrieval. Fig. 1 illustrates the architecture of the proposed approach.

2.1 BIM-based Synthetic Mask

In this study, fundamental building components (i.e., floors, ceilings, walls, doors, and windows) are prioritized since they generally maintain stable geometric and spatial relationships over time. In contrast, furniture and decorative elements may sometimes change and thus offer less reliable features for image matching. Specifically, the relative positions and proportions of these building elements can provide persistent features to enhance similarity measurements. Therefore, this study intends to generate semantic masks to exploit both the semantic and geometric information contained in each captured view.

To construct a database of semantic masks, BIM models are selected as they offer accurate geometries of the target environment and allow flexible configurations for camera deployment. A custom plugin is developed via the Revit API [12] to automate the placement of virtual cameras within the BIM environment for capturing various interior views. Specifically, the cameras are positioned at predefined intervals along two

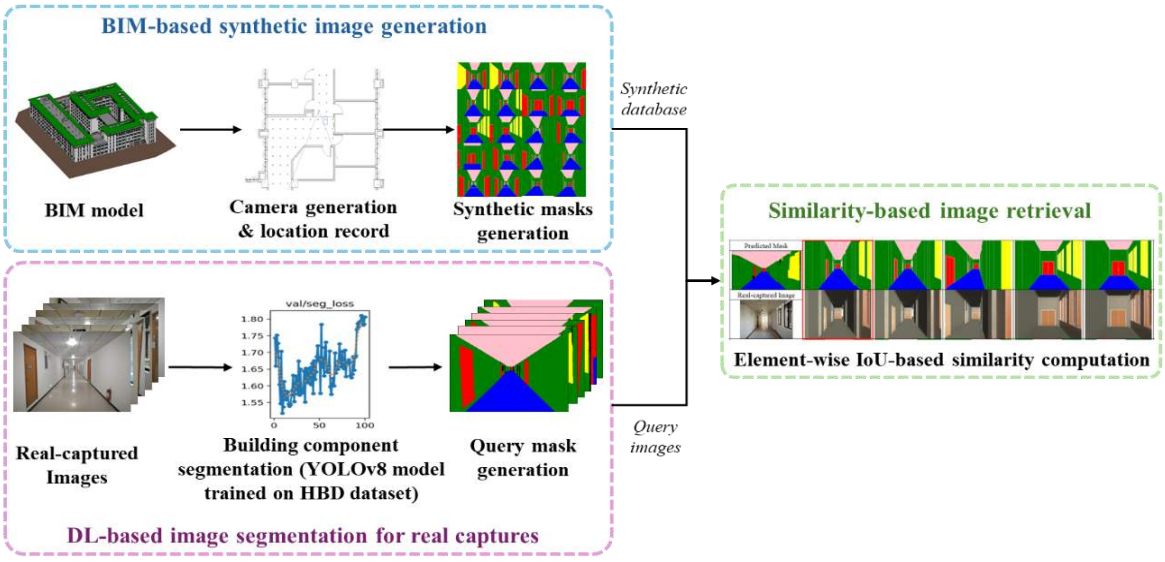


Figure 1. The architecture of the proposed approach

axes, and the boundaries of interior spaces are used to exclude any out-of-bound camera placements. From each camera's position, four axis-aligned orientations are recorded to ensure comprehensive interior coverage. Moreover, camera parameters such as height, image resolution, and dots-per-inch (DPI) can be further customized to produce images with suitable sizes and qualities.

Once the virtual cameras are configured, each building component type within the BIM model is assigned a distinct rendered color to generate component-specific semantic masks. This process involves filtering each target component into a collection and then overriding its graphical properties to establish consistent color coding. Through these steps, synthetic semantic masks from various camera views are generated, along with the corresponding positions and orientations, together constituting the reference database for image retrieval.

During the large-scale semantic mask generation, certain images may lack sufficient spatial features (such as the cases where cameras are placed too close to walls) and provide limited values for image matching. To address this, a filtering process is introduced to enhance the quality of the synthetic BIM-based mask database. The filtering criterion is to retain images that simultaneously include both the ceiling and the floor classes to ensure sufficient semantic and geometric patterns are covered in the masks across various capturing views. In this way, the less distinguishing semantic masks are removed by filtering to enhance the referential quality of the database.

2.2 DL-based Image Segmentation for Real Captures

After preparing the offline database, it is also necessary to process real-captured images and create corresponding masks that have consistent classes with the synthetic BIM masks. To do so, this study employs supervised learning techniques to classify and delineate the building components within images. Specifically, an off-the-shelf image segmentation model, YOLOv8 [13], is adopted as it enables rapid inference speed while maintaining good recognition performance. To align with the classes in the synthetic BIM masks, this study selects the HKU building image dataset (HBD) as the training material since it comprises large-scale annotated images of both building interiors and exteriors, covering all the interested building components [14]. Given the scope of this study, a subset of HBD containing only the interior scenes is utilized, yielding 1,780, 224, and 231 images for training, validation, and testing.

The training process is performed using a computer with configurations of Intel i9-14900HX CPU, NVIDIA GeForce RTX 4060 laptop GPU, and 8 GB GPU memory. After 300 epochs, the training and validation losses plateau, and the training process is terminated. As a result, the best-performing model with the highest validation is subsequently deployed for real-world predictions. The segmentation performance on the test set is summarized in Table 1, which indicates the model obtains moderately high segmentation accuracy. The segmentation of ceilings and floors performs relatively well, probably due to their fixed positions and shapes within images. In contrast, the segmentation result of windows is less effective because of the limited training materials (731 instances in total in the interior scenes of the HBD dataset [14]) and their various appearances.

Table 1. Building component segmentation results on the test set

Class	Images	Instances	mAP^{50}	mAP^{50-95}
Wall	228	554	0.829	0.670
Door	183	315	0.938	0.868
Window	34	94	0.741	0.540
Ceiling	190	213	0.915	0.776
Floor	215	237	0.945	0.866

2.3 Similarity-based Image Retrieval

Last, similarity-based image retrieval is implemented to match the real query images and the BIM-based image database to calculate the most similar pair and acquire the camera positions for indoor localization. Specifically, both the BIM-derived synthetic masks and the real-captured semantic mask are utilized for comparison. These semantic masks effectively capture the composition of the architectural elements, where their spatial arrangement and proportions help characterize the interior layouts. In general, two images deemed similar would exhibit analogous distribution and counts of building elements, occupying comparable proportions and locations within each image. Based on this pattern, this study designs a similarity metric using the element-wise intersection-over-union (IoU) ratio to reflect the overlaps between the real and synthetic masks. The similarity metric focuses on the five fundamental building components to compute the class-balanced IoU values. For each class c_i in the class set C, the pre-class binary masks are generated for the real and synthetic masks, as shown in Equations (1) – (2). Next, a pre-class IoU (IoU_{c_i}) is calculated to quantify the overlap between these corresponding binary masks. Given that the spatial distribution of door and window elements usually provides more distinguishable cues to describe the location-aware context than the other classes, this study assigns different weights w_{c_i} to each class. As a result, the overall similarity metric s is derived by aggregating these weighted IoU values into a single score, as shown

in Equations (3) – (4).

$$R_{c_i} = \{(x,y)|R(x,y) = c_i\}, \quad c_i \in C \tag{1}$$

$$S_{c_i} = \{(x,y)|S(x,y) = c_i\}, \quad c_i \in C \tag{2}$$

where $C = \{c_1, c_2, \ldots, c_n\}$ is the set of semantic classes, R and S is the real mask and the synthetic mask, and R_{c_i} and S_{c_i} are the pre-class binary real and synthetic masks that include the pixel locations labelled as a class c_i.

$$IoU_{c_i} = |R_{c_i} \cap S_{c_i}|/|R_{c_i} \cup S_{c_i}| \tag{3}$$

$$s = \sum_{c_i \in C} w_{c_i} IoU_{c_i} \Big/ \sum_{c_i \in C} w_{c_i} \tag{4}$$

where the default weights of the door, window, wall, floor, and ceiling classes are empirically set to 0.325, 0.325, 0.15, 0.1, and 0.1.

By comparing the real and synthetic masks, the similarity metric identifies the image pairs whose semantic layouts most closely resemble one another, thereby determining the best candidate for indoor localization.

3 Case Study

3.1 Experiment Setup

To validate the feasibility and localization performance of the proposed approach, a real-world case study was carried out. The experiment was conducted in an institutional building at the University of Macau, China, where a publicly accessible corridor area was selected as the testbed. The corridor area is an L-shaped space with a length of approximately 175 meters. Both the fundamental building elements and the furniture and decorative elements (e.g., garbage bin, water dispenser, and poster board) exist inside the corridor area. The floorplan layout of the testbed is displayed in Fig. 2.

The proposed approach was applied to the corridor area for indoor localization. First, the LOD 200 BIM model of the institutional buildings was employed to construct the offline database. This BIM model contained fundamental building components, while other furniture and decorative components were not created due to the considerable efforts required. This study then leveraged the custom plugin to generate a set of virtual cameras, where the camera interval was set to 0.914 meters (3 feet), resolution was set to 640×480 pixels, and the DPI was configured at 600. The camera interval was determined through an iterative testing process. Initial experiments with smaller intervals yielded excessive redundancy in the image database, as numerous virtual cameras generated highly similar images with negligible variation. In contrast, larger intervals substantially reduced the size of the reference database, but this came at the cost of increased localization errors due to inadequate image coverage. Therefore, a moderate interval was determined via trial-and-errors to balance the database size and the localization accuracy.

Subsequently, 616 virtual cameras were generated, covering 2,464 capturing views, with cameras oriented

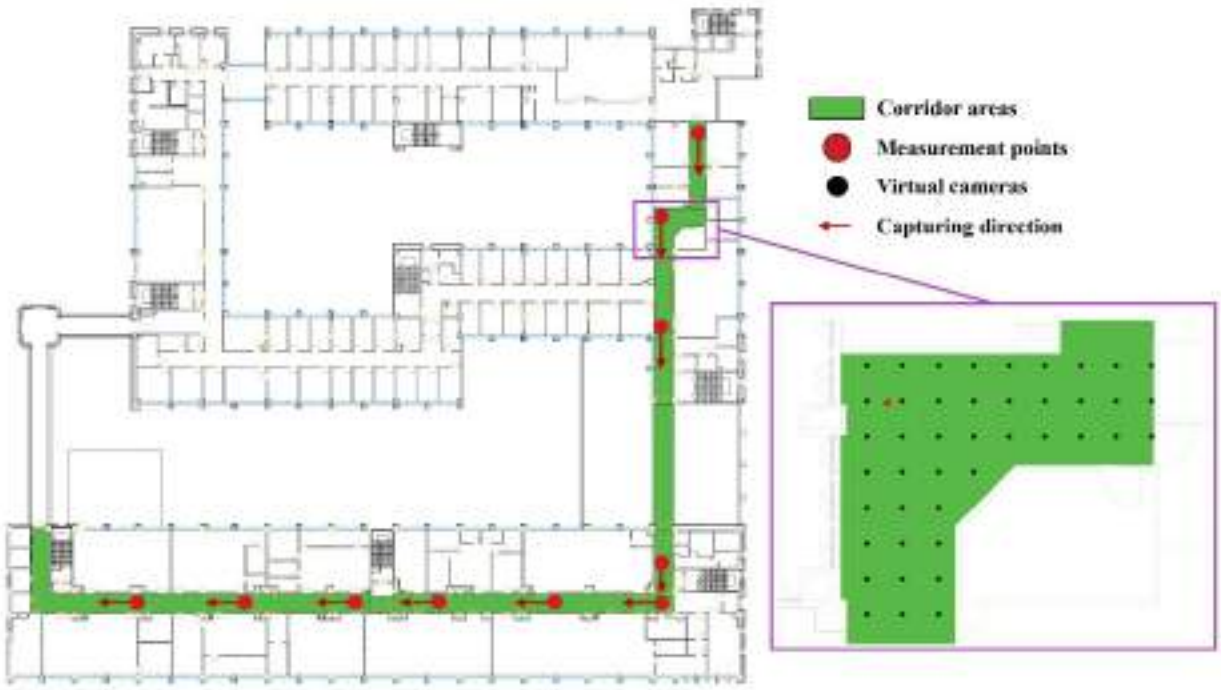


Figure 2. Virtual camera deployment and the measurement points in the testbed BIM model

along the four directions of the X and Y axes. After filtering, 1,005 semantic masks remained as the reference data in the offline database. Fig. 2 shows the deployment of the virtual cameras in the BIM model.

Next, this study conducted a pedestrian walk inside the indoor environment. During the walk, ten measurement points were selected to capture real photos along the walking direction. Additionally, their relative locations were recorded manually using a tape to form the ground truth data. In this study, a digital camera equipped with a mobile device (Xiaomi Mi 13 [15]) was utilized to capture the real photos. Fig. 2 shows the locations of the measurement points in the BIM model. These photos were then performed segmentation predictions to generate the masks of building components. Last, the mask of each measurement point severed as the query images to match with the offline database. The most similar pair was calculated based on the class-balanced IoU metric.

For comparison, this study selected the method proposed by Ha et al. [8] as the baseline approach. The baseline approach leveraged BIM models to generate realistic rendered images and then utilized convolutional neural networks to extract visual features from images for similarity computation. Accordingly, we generated the BIM-rendered images of the corridor areas and took the fourth convolutional block of the ResNet50 model [16] built on the Keras framework [17] to characterize the image features. Subsequently, the cosine similarity metric was employed for image retrieval.

To perform quantitative analysis, this study introduced two indicators to compare the accuracy of the proposed and baseline approaches. The first indicator is the system error, which measures the deviation between the actual positions and its nearest reference points in the offline database. This system error reveals the inherent impact of the interval spacing of the virtual cameras in BIM on localization accuracy. The second indicator is the top10 range accuracy, which comprehensively quantifies the image retrieval results in terms of both location proximity and mask similarity. The traditional top10 accuracy only considers whether the nearest reference point of the query image exists in the first ten search results or not, without accounting for other reference points that also lie near the actual location. Hence, the top10 range accuracy addresses this limitation by incorporating two parts, i.e., a positioning score and a sequencing score.

The positioning score (p_i) gauges whether the matched reference lies within an acceptable distance (i.e., a specified radius) from the actual position and further quantifies their proximity, as shown in Equation (5). The positioning score ranges from 0 to 1, where a value closer to 1 indicates that the matched reference point is closer to the actual location. The sequencing score (q_i) calculates the relative similarity of each matched result using the IoU metric, as shown in Equation (6). Consequently, the top10 range accuracy (t_{10}Acc) is determined by selecting the maximum product between the position score and the sequencing score among the first ten retrieved results (Equation (7)).

$$p_i = \begin{cases} (r - d_i)/r, & \text{when } d_i \le r \\ 0, & \text{when } d_i > r \end{cases} \quad (5)$$

$$q_i = s_i / s_1 \quad (6)$$

$$\mathrm{t_{10}Acc} = \max_{i \in [1,10]} \{p_i \times q_i\} \quad (7)$$

where d_i is the distance between the i-th matched reference point and the true location, r is the radius of the acceptable range (i.e., 2 meters in this study), and s_i is the IoU metric obtained from Equation (4).

3.2 Experimental Results and Performance Evaluation

After the experiment, the results of the case study are calculated accordingly. For the system error, the results indicate that an average deviation of 0.38 meters exists between the actual locations and their nearest reference points. This error indicates that due to the camera interval configuration, the proposed approach reaches approximately decimetre-level accuracy in the best scenarios, which should meet the requirements of most pedestrian indoor navigation applications.

Next, the results of the top10-range accuracy for the two approaches are summarized in Table 2.

Table 2. Top10 range accuracy of the two approaches

Measurement point	t_{10}Acc	
	The proposed approach	The baseline approach
1	0.88	0
2	1	0
3	0.77	0.95
4	0	0
5	0.54	0.54
6	0.54	0
7	0.51	0
8	0	0
9	0.33	0.34
10	1	0
Average	0.57	0.18

Specifically, the results demonstrate that the proposed approach achieves an average t_{10}Acc of 0.57, showing a moderate level of performance for indoor localization. Fig. 3 illustrates more details about the image retrieval results, revealing that the IoU metric can identify highly similar mask images for all measurement points. However, the range accuracy is sometimes

constrained when multiple images exhibit high resemblance across different locations. Comparatively, the first three measurement points which contain all five building components would result in relatively higher accuracy, given that these images provide richer semantic and spatial details. On the other hand, the rest measurement points only incorporate a reduced set of components (i.e., floors, ceilings, walls, and doors), which may result in similar patterns across various locations and thus sometimes yield suboptimal matches.

By contrast, the baseline approach produces less satisfactory results, leading to an average t_{10}Acc of 0.18. The relatively low accuracy could be attributed to the inconsistent visual features between the BIM-rendered images and the real-captured images. Particularly, the BIM models include only fundamental building components painted in default colours, which significantly differ from the actual indoor environment in terms of visual fidelity. Therefore, the baseline approach may become ineffective when applied to low LOD BIM requirements.

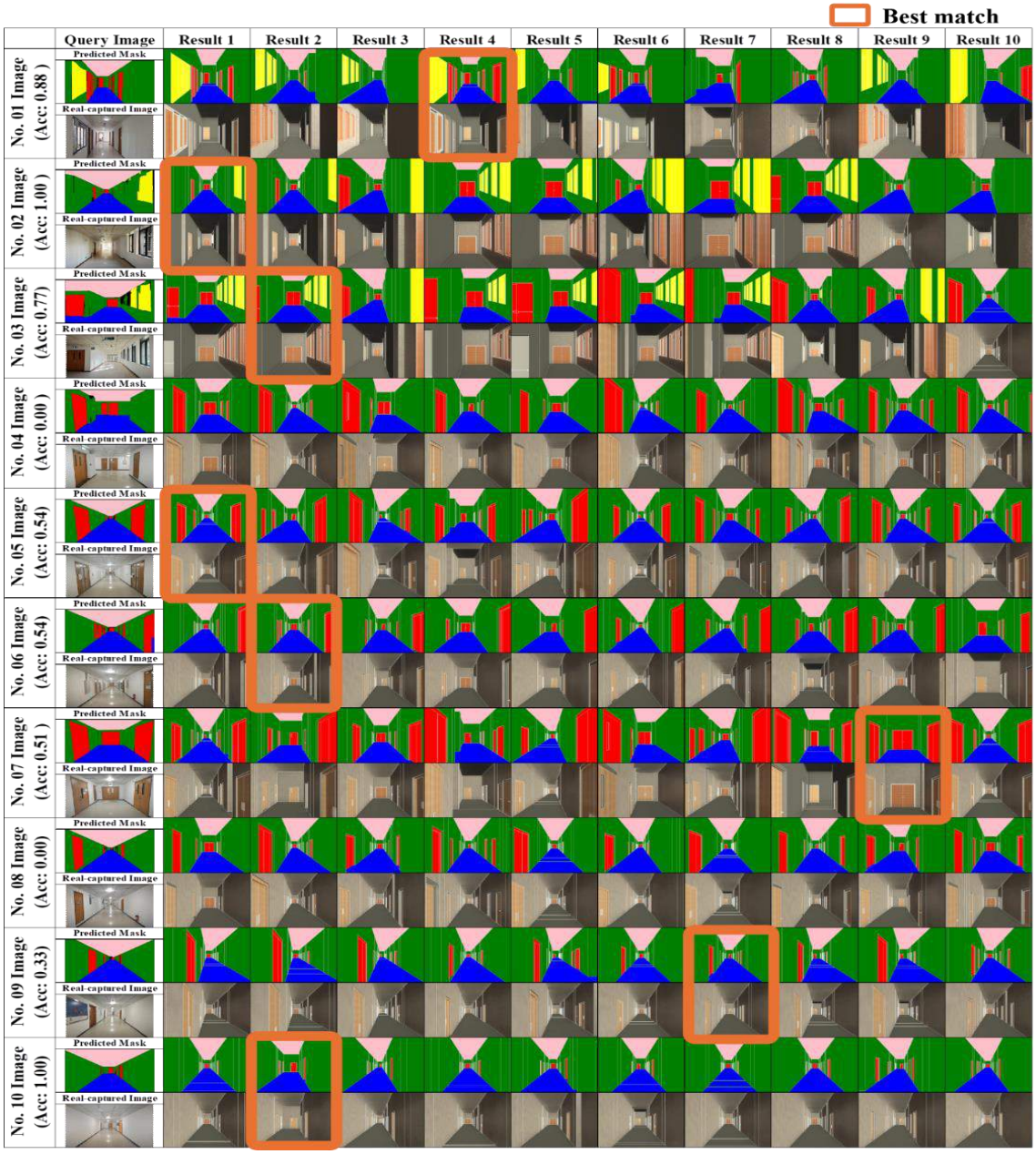


Figure 3. Image retrieval results of the proposed approach.
Note: Queries with no box-selected images indicate that the best match is absent from the top10 results

4 Discussion and Conclusion

This study presents a novel image-based indoor localization approach by integrating BIM synthetic data generation and building component semantic segmentation. Through the development and evaluation of the proposed approach, the benefits and limitations are identified. Specifically, the approach concentrates on the fundamental building components and generates semantic masks for both offline BIM database construction and online image querying. The benefits of this approach are multi-fold. First, it operates effectively with relatively low-fidelity BIM models, thereby minimizing the manual effort required for detailed modelling or continuous updates. Only essential building components are required, which enables greater applicability for real-world uses. Second, it provides a more robust performance against environmental changes. Prior methods often rely heavily on visual fidelity between BIM models and real-world environments, making them vulnerable to discrepancies in actual conditions. By contrast, the proposed approach makes use of the spatial and semantic information of those static building components, which offers consistent patterns for feature matching over time.

Compared to traditional localization methods such as IMU-based tracking or Wi-Fi fingerprinting, the proposed image-based approach offers advantages in measurement mechanisms and ease of deployment. The accuracy of IMU-based methods is gradually affected by cumulative drift errors, which require additional recalibration. Range-based approaches (e.g., Wi-Fi signals) may suffer from environmental interference and often rely on additional infrastructure, such as multiple access points, beacons, or specialized sensors, which may not be readily available in all environments. By contrast, the proposed approach determines the user's locations using independent capturing and similarity matching, which does not incur any accumulated errors. In addition, the approach provides position references using stable architectural features extracted from images without the need for extensive sensor deployments. This makes the approach scalable to large-scale indoor environments.

The result of the case study indicates that the top 10-range accuracy of the proposed approach achieves moderate performance (0.57), which significantly outperforms the conventional baseline approach (0.18). Moreover, the result also discloses that the top 10-range accuracy varies across different measurement points. When positioning with various surrounding features like distinct combinations of windows and doors, the proposed approach results in a better performance. It is because these divergent environmental features provide a more distinguishable image pattern in the segmentation mask to separate it from others in the image database. However, the proposed approach may struggle in common corridor scenes, where repetitive features like similar doors and walls occur and confuse the image retrieval results.

Several limitations are also identified and require further investigation. First, the parameters of virtual camera deployment, such as orientation directions, camera spacing intervals, and focal lengths, are set empirically in the present study, which needs to be optimized by systematic experiments to better understand how they affect localization accuracy and computational efficiency. Second, the current approach only considers discrete measurements using individual image captures, which makes localization accuracy highly susceptible to segmentation errors. Incorporating continuous measurements with probabilistic models could mitigate this issue and offer a more robust solution. Third, the case study reveals that the proposed approach may become less effective when applied to locations with similar spatial patterns of building components. To address this, additional essential elements (e.g., mechanical, electrical, and plumbing components) could be included in the segmentation process to enhance feature distinguishability. Finally, future work should validate the method across diverse interior environments to confirm its generalizability and practical applicability. Specifically, the present study is still in the proof-of-concept stage to investigate the feasibility and performance of the proposed approach for BIM-based indoor localization. In the future, the approach should be further integrated into self-developed prototypes or existing software to empower different applications such as facility management and emergency response. Such integration has the potential to assist various practitioners, including facility managers, construction engineers, and emergency responders, in enhancing operational efficiency and decision-making. Besides, systematic comparison to existing approaches using open datasets is needed to quantify its performance and identify its utilization and potential challenges.

To conclude, this study demonstrates the potential of leveraging spatial and semantic details of fundamental building components in capturing views to develop an efficient and reliable approach for indoor localization. At the same time, follow-up research is still needed to fully exploit the advantages of the approach and broaden its adoption in more complex indoor environments.

Acknowledgements

The authors acknowledge funding support from The Science and Technology Development Fund, Macao S.A.R (FDCT project no. 0034/2024/RIB1).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant

agreement No 101034337.

References

[1] P. Roy and C. Chowdhury. A survey of machine learning techniques for indoor localization and navigation systems. *Journal of Intelligent & Robotic Systems*, 101(63), 2021. https://doi.org/10.1007/s10846-021-01327-z.

[2] M.O. Wong, H. Zhou, H. Ying, and S. Lee. A voice-driven IMU-enabled BIM-based multi-user system for indoor navigation in fire emergencies. *Automation in Construction*, 135, 104137, 2022. https://doi.org/10.1016/j.autcon.2022.104137.

[3] M. O. Wong and S. Lee. Indoor navigation and information sharing for collaborative fire emergency response with BIM and multi-user networking. *Automation in Construction*, 148:104781, 2023. https://doi.org/10.1016/j.autcon.2023.104781.

[4] J. Kunhoth, A.G. Karkar, S. Al-Maadeed, and A. Al-Ali. Indoor positioning and wayfinding systems: a survey. *Human-centric Computing and Information Sciences*, 10(1):1–10, 2020. https://doi.org/10.1186/s13673-020-00222-0.

[5] I. Ha, H. Kim, S. Park, and H. Kim. Image retrieval using BIM and features from pretrained VGG network for indoor localization. *Building and Environment*, 140:23–31, 2018. https://doi.org/10.1016/j.buildenv.2018.05.026.

[6] H. Kawaji, K. Hatada, T. Yamasaki, and K. Aizawa. Image-based indoor positioning system: fast image matching using omnidirectional panoramic images. *In Proceedings of the 1st ACM International Workshop on Multimodal Pervasive Video Analysis*, pages 1–4, October 2010. https://doi.org/10.1145/1878039.1878041.

[7] X. Sun, Y. Xie, P. Luo, and L. Wang. A dataset for benchmarking image-based localization. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7436–7444, 2017.https://doi.org/10.1109/CVPR.2017.598

[8] I. Ha, H. Kim, S. Park, and H. Kim. Image-based indoor localization using BIM and features of CNN. *In Proceedings of the 35th International Symposium on Automation and Robotics in Construction (ISARC)*, pages 1–4, 2018. https://doi.org/10.22260/ISARC2018/0107.

[9] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014. https://doi.org/10.48550/arXiv.1409.1556.

[10] D. Acharya, K. Khoshelham, and S. Winter. BIM-PoseNet: indoor camera localisation using a 3D indoor model and deep learning from synthetic images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 150:245–258, 2019. https://doi.org/10.1016/j.isprsjprs.2019.02.020

[11] A. Kendall, M. Grimes, and R. Cipolla. Posenet: A convolutional network for real-time 6-DOF camera relocalization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2938–2946, 2015. https://doi.org/10.1109/iccv.2015.336.

[12] Autodesk. Revit 2024 API Documentation. On-line: https://www.revitapidocs.com/, Accessed: 24/12/2024.

[13] G. Jocher, J. Qiu, and A. Chaurasia. Ultralytics YOLO (Version 8.0.0) [Computer software]. On-line: https://github.com/ultralytics/ultralytics, Accessed: 24/12/2024.

[14] M. O. Wong, H. Ying, M. Yin, X. Yi, L. Xiao, W. Duan, and L. Tang. Semantic 3D reconstruction-oriented image dataset for building component segmentation. *Automation in Construction*, 165:105558, 2024. https://doi.org/10.1016/j.autcon.2024.105558.

[15] Xiaomi. Xiaomi 13 - Specifications. On-line: https://www.mi.com/global/product/xiaomi-13/specs, Accessed: 01/13/2025.

[16] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016. https://doi.org/10.1109/cvpr.2016.90.

[17] Chollet, F., & others. (2015). *Keras*. Retrieved from https://keras.io

Assessing the Efficacy of Virtual Reality Site Visits to Enhance Remote Learning Experiences for Civil and Construction Engineering Students

Samira Mirhasani[1], Jared Bosch[1] and Joseph Louis[1]

[1]School of Civil and Construction Engineering, Oregon State University, Corvallis, Oregon, United States
mirhasas@oregonstate.edu, boschja@oregonstate.edu, joseph.louis@oregonstate.edu

Abstract –

Remote construction engineering education can provide immense benefits to students and colleges by removing time and cost hindrances that currently prevent enrolment. However, replicating the hands-on experiential learning provided by traditional in-person site visits poses significant challenges for remote students due to accessibility barriers. These barriers include limited access to physical construction sites and the inability to engage in peer-to-peer interactions common in in-person settings. This research aims to bridge this gap between remote and in-person students of construction engineering by studying the potential of virtual reality site visits on undergraduate education for civil and construction engineering students. Thus, the study developed and deployed a virtual site visit using the discrete event simulation software ConStrobe and game engine Unity, accessible through WebGL on smartphones. Participants experienced both in-person and virtual site visits and engaged with the simulation to complete classwork tasks such as calculating equipment productivity for the Heavy Civil Construction Management course. Data was collected using surveys and analysed to assess their understanding and engagement. Initial findings suggest that virtual simulations effectively complement traditional site visits by improving students' understanding of construction concepts, particularly productivity and equipment interaction. However, some students reported challenges such as feeling overwhelmed, emphasizing the need for enhanced user guidance. These insights provide guidance for construction instructors and curriculum designers aiming to integrate virtual simulations into the classes, offering an accessible and effective tool to support experiential learning for remote learners.

Keywords – Virtual Reality Site Visits, Remote Education, Experiential Learning, Construction Education

1 Introduction

The evolving demands of construction management education necessitate innovative approaches to deliver experiential learning, especially in remote and hybrid settings. The shift toward remote education, accelerated by the COVID-19 pandemic, further highlights the need to address issues of student engagement and experiential learning. Studies, such as those by (Neisler 2020), indicate that remote students often struggle with motivation due to a lack of hands-on experiences and peer interactions. In response, (Park et al. 2020) recommend the use of virtual spaces and tools to replicate the serendipitous interactions and moments of engagement that occur in in-person settings. On the other hand, traditional in-person site visits are invaluable for providing hands-on exposure but pose challenges related to accessibility for remote students.

Thus, the goal of this research is to explore the use of virtual reality to enhance remote learning by providing in-person student experiences of construction site visits and peer interaction through virtual reality platforms. This research will implement virtual reality site visits for students in the Heavy Civil Construction Management course and collect student feedback and data about its use to answer the following research questions: (1) Can virtual reality site visits improve student experience for remote students? (2) Can virtual reality site visits measurably improve student understanding compared to site visits?

By addressing these questions, the study assesses whether and to what extent virtual simulations can replicate in-person construction site visits and serve as a robust learning tool. The findings contribute to the development of scalable, sustainable, and inclusive learning tools in construction management education, aligning with broader goals of accessibility.

2 Literature Review

Remote learning technologies, such as virtual site visits, are increasingly used in construction education, providing accessible and cost-effective alternatives to traditional methods (Jaselskis *et al.* 2012). During the COVID-19 pandemic, remote teaching broadened accessibility but highlighted challenges in maintaining educational quality and hands-on skills acquisition (Martínez-Muñoz *et al*. 2021). These challenges underscore the need for innovative solutions like virtual simulations to bridge gaps in accessibility, engagement, and comprehension while addressing technological and pedagogical challenges. This section explores the use and approaches of virtual simulations in construction education.

2.1 Use of Virtual Simulations in Construction Education

Virtual simulations are transforming construction education by bridging theoretical knowledge and practical application. They have been effectively applied in safety training (Ismail *et al.* 2021), immersive learning for specific construction phases (Salman 2023), and onboarding programs (Belaroussi *et al.* 2023). Studies demonstrate their role in enhancing comprehension, engagement, and problem-solving skills across topics like highway engineering (Seifan *et al.* 2020), railway transportation (Eiris *et al.*, 2022), and steel structure assembly (Salman 2023).

Virtual scenarios also support construction management education by reinforcing principles such as project planning and safety compliance (Shojaei *et al.* 2023) and enabling hands-on practice in scheduling *(Ilbeigi et al*. 2024) and post-disaster assessment (Humayoun *et al.* 2024). Despite these advances, research on their application in customized heavy civil construction management topics remains limited, which this study aims to address.

2.2 Variance of Simulation Approaches

The degree of immersion and type of simulation significantly impact virtual learning outcomes. Visualization methods range from 2D displays to immersive 3D environments (Shojaei *et al.* 2022), allowing educators to tailor experiences based on subject complexity and learner proficiency. Custom VR platforms, like those by Eiris et al. (Eiris *et al.* 2022) and Elgewely et al. (Elgewely *et al.* 2021), explore innovative teaching methods that enhance student engagement. However, immersive tools like head-mounted displays (HMDs) and CAVE systems (Wang *et al.* 2024) offer high levels of engagement while presenting challenges related to comfort, accessibility, and cost, particularly in large or remote learning settings (Lucas 2018). In contrast, web-based platforms such as Mozilla Hubs® (Eiris et al. 2022) offer greater accessibility for devices like smartphones and tablets. However, allowing diverse hardware use, ranging from phones to laptops and HMDs, can result in inconsistent user experiences and technical challenges, such as varying device capabilities and a lack of experimental control. Additionally, cybersickness with immersive systems can be another issue (Eiris et al. 2022), whereas WebGL-based platforms mitigate this problem and ensure consistent performance across devices.

This study adopts a simpler approach, avoiding immersive technologies like 360-degree images, which can sometimes overwhelm students. Instead, the simpler visualization was tailored to meet specific course learning outcomes. The implementation relied on a web-based simulation accessed primarily through smartphones, ensuring affordability and consistency across devices. This standardization minimized variability caused by diverse hardware setups and provided an equitable learning experience for all students. By prioritizing accessibility, simplicity, and alignment with educational objectives, this approach demonstrates the potential of virtual simulations to effectively enhance remote learning in construction education.

3 Research Methodology

This research examines differences in student perceptions of experience and understanding between real-world and virtual simulation environments. Key steps include 1. Development of Survey Questions 2. Real-World site visit, 3. Development of virtual site visit, 4. Data collection, and 5. Data analysis. Figure 1 shows research methodology workflow.

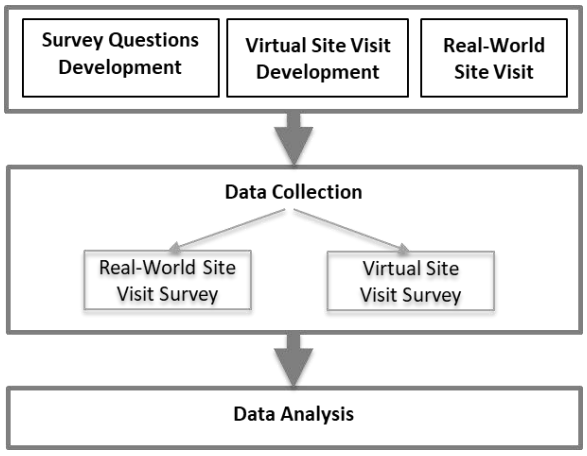


Figure 1. Research Methodology Workflow

3.1 Development of Survey Questions

The development of survey questions to evaluate the impact of virtual simulations is grounded in a diverse range of scholarly sources. For instance, (Yuan Sun et al. 2022), Kamińska et al. (2022), and (Wen et al. 2023) provide frameworks and insights that inform the design of assessment tools. These studies emphasize the importance of crafting questions that capture both qualitative and quantitative aspects of the learning experience, such as engagement, comprehension, and ease of use.

To ensure relevance, survey statements were derived from major topics covered in the CEM441/541 class, such as equipment productivity (both individual and collaborative performance), factors influencing productivity, cost implications of equipment usage and maintenance, and the importance of selecting appropriate machinery to optimize efficiency and minimize costs. The survey comprised two main parts: the first included statements derived from literature-based questions to evaluate students' learning experience. The second focused on CLO #6 (“Develop an estimate for a project using production and costs from a real-world project”) and CLO #7 (“Critically analyze a heavy civil operation and suggest alternatives”), assessing students' understanding of operational and productivity concepts. The post-virtual simulation survey also included five open-ended questions for deeper insights. Tables 1 and 2 categorize statements related to experience and understanding, respectively.

Table 1. Real Site Visit and Virtual Simulation Experience Statements

No.	**Experience Statements**
1	I have a clearer sense of the complexity of a worksite
2	Seeing the activity clarified concepts discussed in class
3	I learned something about site processes that was not discussed in class
4	Questions I had, or unclear concepts, were clarified.
5	The atmosphere/environment was overwhelming and made it difficult to learn
6	The experience was more engaging than I expected
7	The experience captured and held my attention
8	The experience was a valuable part of my learning process
9	I am more comfortable identifying interdependence of site activities
10	The experience provided me enough time to feel fully engaged with material presented
11	I felt like I could ask questions and receive satisfactory answers
12	I felt present and involved in the learning process
13	The experience aligned with my expectations
14	Peer interaction was an important component of the experience
15	I felt like my presence was disruptive to the work process

Table 2. Real Site Visit and Virtual Simulation Understanding Statements

No.	**Understanding**
10	Understanding factors that can improve equipment production (Improving
4	Identifying improvements to maximize equipment production
12	Ability to describe what swell and shrinkage factors are, how they apply to soil volume, and the effect of swell and shrinkage on site production
1	Estimating material to be excavated (either cut of fill)
11	Understanding of how soil is converted from a bank state to a loose state and the effect
14	What a fill factor is, and how it relates to the type of soil being moved or loaded, and its effect on production
3	Estimating cost of equipment
6	Estimating cost per unit of material moved
8	Optimizing equipment pathways for maximum productivity
9	Causes for delays, long lines, and long wait times
19	Understanding how cycle times relate to matching of equipment, and causes of delays, and bunching/queueing
13	Understanding of grade and rolling resistance, and their influence on production
7	Calculating ideal ratio of pieces of equipment (trucks: excavators or pushers: scrapers)
15	Understanding the importance of selecting the proper equipment
16	Describing different pieces of equipment and what they do
17	Understanding how efficiency relates to heavy equipment production
18	Describing how different factors affect efficiency (operator skill, visibility, blade adjustment factor, etc.)

3.2 Real Site Visit

The in-person site visit was a critical component of this study, providing students with firsthand exposure to real-world applications of concepts discussed in class. This visit allowed participants to observe an active construction site, engage with ongoing activities, and experience the complexities of heavy civil engineering projects in a controlled environment. Figure 2 illustrates multiple pieces of equipment working together to excavate and haul earth. Also, it highlights the impact of an unbalanced ratio of excavators to trucks, resulting in idle equipment.



Figure 2. Left Image: Excavator idle, and right image: Excavator and truck working in tandem

3.3 Virtual Site Visit

The virtual environment, developed using simulation tools and visualization techniques, enables students to observe and analyze critical concepts in heavy civil construction, including equipment productivity, operational delays, and system efficiency. The details of the simulation and visualization are explained in the following sections.

3.3.1 Simulation

The discrete event simulation (DES) was conducted using ConStrobe (Construction Operations Network Simulation for Time and Resource-Based Evaluation) (Louis 2025) to simulate real-world construction activities. The simulation focused on earthmoving operations, including activities such as loading, hauling, dumping, and returning. Fixed timings were used for loading and dumping activities, while uniform timings, representing variable durations, were applied to haul and return activities. The simulation involved one loader and four trucks per cycle. Figure 3 illustrates the simulation model, with various shapes representing activities like loading and hauling, along with their respective timings. Links within the model define system functions and the sequence of operations.

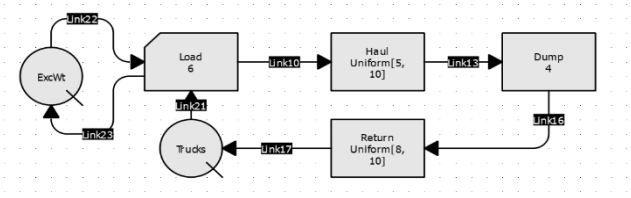


Figure 3. DES in ConStrobe Software

As seen in Figure 3, a model is created to represent earthmoving operations on a construction site. From this model, a text file is created which is exported from ConStrobe and used as input to create an animation in Unity, a game engine.

3.3.2 Visualization

To make the simulation more comprehensible for students, it was visualized using the Unity game engine. The simulation exported a trace file containing commands such as CREATE, TIME, MOVE, SETOBJCELL, and DYNUPDATECELL, which were used to generate animated visualizations.

Unity's versatility allows for exporting simulations in various formats, including desktop applications (.exe), mobile apps, and WebGL. For this study, WebGL was chosen due to its compatibility, ease of use, and ability to run in any web browser without requiring additional software installations or powerful hardware, such as head-mounted displays (HMDs) or high-performance computers. Most students accessed the visualization on their smartphones, while a few opted to use laptops. For homework, students accessed the simulation via a QR code linked to the WebGL platform, where they calculated cycle times for trucks and loaders.

The animations illustrated key earthmoving activities, including loading, hauling, dumping, and returning. Students could identify trucks by their assigned colors and interact with the simulation using features such as pause/play controls for timing calculations, touch navigation for exploration, and a reset button to restart the simulation from the initial camera angle. The visualization provided an opportunity for students to observe various scenarios in earthmoving operations as the simulation's time clock progressed. These scenarios demonstrated essential course concepts, such as bunching and queuing, variable timing, idle loaders, and waiting trucks, offering a dynamic and engaging learning experience. Figures 4–7 showcase examples of the visualization environment encountered by students.

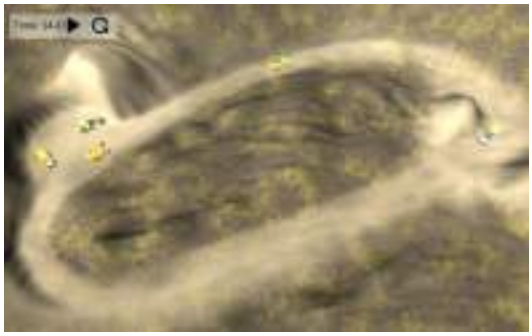


Figure 4. Scenario 1: Blue truck dumped, green hauling, red waiting to load, yellow is idle.

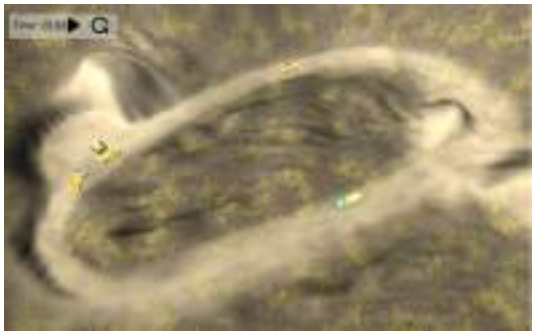


Figure 5. Scenario 2: Blue truck waiting, green returning, red hauling, yellow loading.

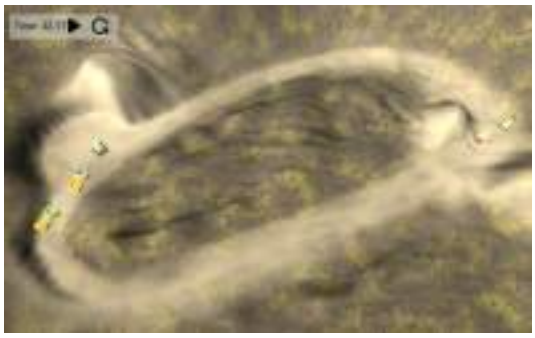


Figure 6. Scenario 3: Bunching in both loading and unloading areas.

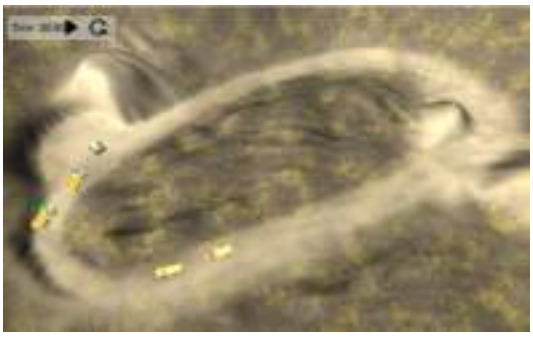


Figure 7. Scenario 4: Yellow truck overtaking red one, showing different timing.

3.4 Data Collection and Analysis

This study compared the educational impact of in-person construction site visits and virtual simulations on students' experience and understanding. Surveys were administered digitally via Qualtrics to all participants. A baseline survey was collected after the in-person site visit, followed by a post-virtual simulation survey. For post-virtual survey, participants engaged with the virtual simulation via a QR code through a WebGL platform on smartphones or laptops, completing tasks such as analysing equipment productivity and calculating cycle times(Figure 8). The surveys were identical except for 5 open-ended questions included at the end of the post-virtual simulation survey. Response options for questions evaluating experience were scored on a 5-point Likert-type scale: 1 = Strongly disagree, 2 = Somewhat disagree, 3 = Neither agree nor disagree, 4 = Somewhat agree, 5 = Strongly agree. For questions assessing understanding, the scale ranged from 1 = Not knowledgeable at all, 2 = Slightly knowledgeable, 3 = Moderately knowledgeable, 4 = Very knowledgeable, to 5 = Extremely knowledgeable. Responses were automatically recorded to the Qualtrics web portal. A total of 46 responses were collected, providing valuable insights into the educational effectiveness of virtual site visits. It is important to note that the analysis is based on a single course with 46 responses, which limits the generalizability of the findings.

Initial feedback indicates that virtual simulations improved students' understanding of key construction concepts, including equipment productivity, optimal ratios, and delay management. The simulation effectively enhanced comprehension of dynamic and complex topics compared to in-person visits. However, students' experiences showed mixed results, with some finding the virtual environment overwhelming or misaligned with expectations, emphasizing the need for improved user guidance and interface design.

Future work will involve detailed statistical analysis to compare paired ordinal responses and further evaluate the educational impact of the two methods.



Figure 8. Students working on homework using visualization

4 Findings and Discussion

This study successfully developed and deployed a virtual site visit using Unity and WebGL to enhance students' understanding and experience of heavy civil construction concepts. The visualization allowed students to explore key topics, including equipment productivity, cycle times, queuing, and delays, in a structured and accessible environment. Students accessed the simulation via a QR code, enabling them to interact with the visualization and apply course concepts through homework tasks, such as calculating cycle times and equipment productivity by observing earthmoving operations. Initial feedback from surveys indicates that the virtual simulation provided valuable insights into construction operations and effectively complemented traditional in-person site visits. While most students found the simulation engaging and informative, some reported challenges, such as feeling overwhelmed or unsure of what to expect, highlighting areas for improvement in interface design and user guidance. Future work will analyze the survey data in detail to assess the educational impact of the virtual simulation compared to in-person site visits and to identify further opportunities to enhance the virtual learning experience. Additionally, future research will incorporate objective performance metrics, such as pre- and post-simulation assessments, to provide a more concrete measure of learning effectiveness.

5 Conclusion

This study demonstrates the potential of virtual simulations as an effective and accessible alternative to traditional in-person site visits in construction management education. Virtual simulations enhanced students' understanding of key concepts such as equipment productivity, optimal ratios, causes of delays, and overall production efficiency. Compared to traditional in-person site visits, virtual simulations provide a controlled, tailored learning environment that simplifies dynamic, abstract processes and allows for deeper conceptual comprehension.

These advantages make virtual simulations particularly valuable for remote learners who may lack access to physical construction sites, bridging gaps in hands-on learning opportunities. While virtual simulations successfully replicated and, in some areas, surpassed the learning outcomes of in-person visits, areas for improvement remain. Students reported challenges, such as feeling overwhelmed, or uncertain about what to expect, emphasizing the importance of pre-simulation preparation, such as tutorials or demonstrations, to improve alignment between expectations and experiences. Furthermore, virtual simulations proved particularly beneficial in teaching concepts reliant on dynamic visualizations, such as cycle times and delay analysis, highlighting their potential as a complementary tool to traditional methods.

As remote and hybrid education continues to grow, virtual simulations offer a promising pathway to enhance accessibility and inclusivity while maintaining the quality of experiential learning. Future research should focus on tailoring simulations to specific educational objectives and optimizing learning environments for better accessibility and usability. With thoughtful implementation and continued refinement, virtual simulations hold the potential to transform construction management education, ensuring an optimized balance between virtual and in-person learning experiences. Despite these promising results, the study's scope was limited by its single-course design and sample size. Future research should validate these findings across multiple courses and larger, more diverse student populations, explore the scalability and practical implications of virtual simulations.

References

[1] J. Neisler, B. Means. *Suddenly online: A national survey of undergraduates during the COVID-19 pandemic*. Digital Promise, 2020.

[2] J. J. Park, M. Park, K. Jackson, G. Vanhoy. Remote Engineering Education under COVID-19 Pandemic Environment. *International Journal of Multidisciplinary Perspectives in Higher Education*, *5*(1), 160-166, 2020.

[3] E.J. Jaselskis, C.T. Jahren, P.G. Jewell, E. Floyd, T.C. Becker. Virtual Construction Project Field Trips Using Remote Classroom Technology. pp. 236–245, 2012.

[4] D. Martínez-Muñoz, J.V. Martí, V. Yepes. Remote Teaching in Construction Engineering Management During Covid-19. Proceedings, pp. 879-887, 2021.

[5] R. Belaroussi, H. Dai, E. González, J. González. Designing a Large-Scale Immersive Visit in Architecture, Engineering, and Construction. *Applied Sciences*, 13(5), p. 3044, 2023.

[6] R. Eiris, Y. Sun, M. Gheisari, B. Marsh, P. Lautala. VR-OnSite—Online Site Visits Using Web-Based Virtual Environments. pp. 100–109, 2022.

[7] M.H. Elgewely, W. Nadim, A. ElKassed, M. Yehiah, M. Talaat, S. Abdennadher. Immersive construction detailing education: building information modeling (BIM)–based virtual reality (VR). *Open House International*, 46(3), pp. 359–375, 2021.

[8] S.R. Humayoun, J. Wong, Kh. Nguyen, P. Zinjarde, P. Ramesh. VREE: A 3D Environment to Train

Engineers about Post-Earthquake Damages in Building Structure. *IEEE*, pp. 931–932. 2024.

[9] M. Ilbeigi, D. Bairaktarova, R. Ehsani. Gamified Method for Construction Engineering Education: Learning through Guided Active Exploration. *Journal of Civil Engineering Education*, 150(2), p. 05023011, 2024.

[10] H.B. Ismail, A. Ismail, A.A. Abu Bakar, A.R. Abu Tali, N.A. Mohd Noor. The Effectiveness of Site Visit Approach in Teaching and Learning of Construction Site Safety: A Case Study in Civil Engineering Faculty, Universiti Teknologi MARA, Johor Branch, Pasir Gudang Campus. *Asian Journal of University Education*, 17(4), pp. 212–221, 2021.

[11] D. Kamińska, G. Zwoliński, A. Laska-Leśniewicz. Usability Testing of Virtual Reality Applications—The Pilot Study. *Sensors*, 22(4), p. 1342, 2022.

[12] J. Lucas. Student Perceptions and Initial Response to using Virtual Reality for Construction Education, 2018.

[13] A. Salman. Field Trips and Their Effect on Student Learning: A Comparison of Knowledge Assessment for Physical versus Virtual Field Trips in a Construction Management Course. *Virtual Worlds*, 2(3), pp. 290–302, 2023.

[14] M. Seifan, O.D. Dada, A. Berenjian. The Effect of Real and Virtual Construction Field Trips on Students' Perception and Career Aspiration', *Sustainability*, 12(3), p. 1200, 2020.

[15] A. Shojaei, A. Mahdavian, S. Rokooei, A. Al-Bayati. Use of Virtual Site Visits in Construction Management Distance Education. pp. 182–191, 2022.

[16] A. Shojaei, R. Ly, S. Rokooei, A. Mahdavian, A. Al-Bayati. Virtual Site Visits in Construction Management Education: A Practical Alternative to Physical Site Visits. *Journal of Information Technology in Construction (ITcon)*, 28(36), pp. 692–710, 2023.

[17] X. Wang, M. Chou, X. Lai, J. Tang, J. Chen. Examining the Effects of an Immersive Learning Environment in Tertiary AEC Education: CAVE-VR System for Students' Perception and Technology Acceptance. *Journal of Civil Engineering Education*, 150(2), p. 05023012, 2024.

[18] J. Wen, M. Gheisari, J. Jain, Y. Zhang, R.E. Minchin. *iVisit-Communicate* for AEC Education: Using Virtual Humans to Practice Communication Skills in 360-Degree Virtual Field Trips. *Journal of Computing in Civil Engineering*, 37(3), p. 04023008, 2023.

[19] S. Yuan, G. Albeaino, M. Gheisari, R. Eiris. Online site visits using virtual collaborative spaces: A plan-reading activity on a digital building site. *Advanced Engineering Informatics*, 53, p. 101667, 2022.

[20] J. Louis. ConStrobe;Construction Operations Simulation for Time and Resource Based Evaluations. On-line: https://josephlouis.org/ConStrobe/, Accessed: 1/09/2025.

Efficient Camera Pose Estimation Approach for Infrastructure Inspection

Anas Alsharo[1], Max Midwinter[1], and C.M. Yeum[1]

[1]Civil and Environmental Engineering Department, University of Waterloo, Canda
aaalshar@uwaterloo.ca, mxxmidwi@uwaterloo.ca, cmyeum@uwaterloo.ca

Abstract

Visual inspection of civil infrastructure assisted by Unmanned Aerial Vehicles (UAVs) witnessed significant improvements due to the rapid development of drone-mounted cameras and sensors. In visual inspection, accurate pose estimation of the collected images is a pivotal task that enables registering images to pre-existing 3D scenes of the structure to identify the geometric relationship between the scene and the image. This allows for localizing regions of interest or defects in these images. In this work a novel image pose estimation technique is proposed to improve images pose estimation and registration of drone-collected images. The proposed method utilizes a 3D base Structure from Motion (SfM) model and pre-calibrated poses of base images to facilitate the registration and pose estimation of new query images via feature-based pose estimation framework. The method leverages deep-based feature matching to generate dense 2D correspondences to simultaneously generate the 2D and 3D correspondences needed to execute the perspective-n-point solver and accurately estimate the pose. The proposed approach was tested against drone-collected images of cell tower and the image registration efficiency was evaluated through ROI localization in the registered images.

Keywords –

Visual Inspection; Asset Management; Image Localization; Telecommunication Cell Tower

1 Introduction

Condition assessment of large and hard-to-access infrastructure is a complex task and often requires trained technicians to employ time- and resource-extensive techniques for physical inspection [1]. The rapid development in UAVs and their associated sensors has led to a wide range of applications in condition assessment of civil infrastructures [2]. However, UAVs (e.g., drones) can be considered a multifunctional data acquisition system. Hence, robust inspection algorithms are required to leverage and fuse the various data types that drones can acquire to make sure of the acquired data.

A widely adopted application of drones in condition assessment of civil infrastructure involves utilizing them to capture high-resolution images, which are then used to construct SfM models of the structure. For example, in bridge inspection, SfM models enhanced damage quantification [3], crack identification and visualization [4], and risk assessment and management [5]. Xiao et. al. utilized a deep learning-based point cloud segmentation algorithm to assist ROI extraction and crack identification in bridges [6]. For large highway truss structures, Yeum et. al. proposed autonomous ROI localization algorithm, and ROI classification algorithm [7]. Visual inspection using drone images was also implemented for crack and rust identification within ROIs extracted from monopole tower images [8] and produce a health index of power transmission tower [9].

In visual inspection, determining the 6 degrees of freedom (DoF) pose of the image (3 for translation and 3 for rotation) enables establishing a geometric relationship between a 3D scene of the structure and the images. In the SfM framework, the simultaneous 3D reconstruction of a scene and pose estimation of images are core processes. This is typically achieved by extracting correspondences between 2D keypoints across multiple images and solving for camera poses and 3D points simultaneously. This implies that in case of visual inspection that involves constructing SfM model using large volume of images, the images are registered automatically as a natural product of the 3D reconstruction.

On the other hand, the inspection process for civil infrastructure typically does not involve constructing an SfM model; instead, it primarily relies on visually analyzing high-resolution images of the structure, collected from specific targeted regions of interest (TRIs). Hence, SfM-enabled pose estimation is not possible and in case the inspector needs to estimate the pose of these images, then a pose estimation technique should be employed that works on single image, or image sequences [10].

Image pose estimation is an extensively explored

topic in the field of computer vision. A critical review by Xu et al. categorized pose estimation methods into two main categories: structure feature-based localization methods and regression-based pose estimation methods [11]. In structure feature-based methods, a 3D scene of the structure is employed to enable building correspondences between 2D points in a query image (image to be localized) and 3D points from the scene to enable estimating the pose of the image. Regression-based methods employ convolutional neural networks and deep neural networks to regress the pose of RGB images (e.g., PoseNet [12])

Structure feature-based methods are widely used due to their robustness and accuracy [11]. The 2D-3D correspondences in this method are typically established by matching distinctive features from a 2D query image with a database of 3D points, each associated with corresponding feature descriptors. Establishing these 2D-3D correspondences enables the estimation of the camera's position and orientation relative to the scene using a geometry constrain solver (e.g., Perspective-n-Point [13]).

A major drawback of this approach is the assumption that a database of descriptors is available and associated with the 3D points of the scene. However, the outputs of commercial 3D reconstruction and SfM platform does not typically involve image feature descriptors. Hence, a computationally intensive step is needed to establish the descriptors database for all base images. Another drawback related to the database of features descriptors is that the same feature extraction and description method should be used for both the base and query image to ensure consistency and accuracy [14]. Accordingly, if the feature extraction and description algorithm used in building descriptors database exhibits poor performance in matching image pairs with significant change in image environment (e.g., perspective, scale, or illumination), then finding proper number of matches between database and query images might not be possible.

On the other hand, assuming that a database of descriptors is established, if the 3D point cloud (SfM model) of the structure is not dense enough, then establishing high-quality correspondences between the database of descriptors and the query image's descriptors cannot be achieved.

To overcome the aforementioned drawbacks, an efficient structure feature-based image pose estimation approach is proposed to enable registering query images to the 3D scene without the need for database of descriptors and that can perform well in registering query images captured with significant variations in scale, perspective, zoom level, and background visibility compared to the database images.

The proposed algorithm leverages deep-learning-based local feature matchers to build dense 2D-to-2D matches between database images with known calibrated image poses and the query image. Then, the dense matches are used to simultaneously determine the 2D features and the 3D corresponding points, needed to estimate the pose. To estimate the image pose, a perspective-n-point solver within a Random Sample Consensus (RANSAC) framework was implemented.

The proposed approach was implemented on drone-collected cell tower images to test the accuracy of the approach in estimating the position and rotation of query images of the cell tower. The reliability of the proposed method was evaluated by utilizing the projection matrix calculated from the estimated pose in localizing regions of interest (ROIs) in the query images.

2 Methodology

2.1 Overview of the Technique

UAV-based inspection of civil infrastructure involves building SfM model from a large set of sequentially captured images of the structure. The process of 3D reconstruction outputs 3D map of the structure (hereafter, base model) along with calibrated camera poses (hereafter, base images).

In many cases, inspectors collect few images (hereafter, new images) around a specific TRI of high importance and the number of images might not be enough to build an SfM model. Hence, a reliable method is needed to enable registering these images and extract their poses without the need for a 3D reconstruction step. Registering the new images enable building the geometric relationship between the 3D scene and the new images which, in return, enable retrieving information from these images. To allow registering any new image to the base model, a pose estimation approach is proposed.

The approach, shown in Figure 1, starts with a base image that clearly shows the TRI where the inspector needs to perform a focused inspection by collecting new images around that region with varying camera angle, distance from the region, zoom level, or perspective. Prior to registering the new images, a vital step is to enable the inspector to retrieve the closest image relative to the base image the inspector started with. A geometry-based closest image retrieval algorithm is proposed, and it retrieves all the closest images within an applicable pre-defined criteria set by the inspector. From the set of retrieved images, the inspector can select the image with the clearest view and highest details of the all the ROIs that needs to be inspected.

The main enhanced step of this paper is to propose a registration approach that addresses the limitations of existing methods discussed in Section 1. Unlike conventional algorithms, the proposed method enables registering new images of hard-to-access infrastructure

with a pre-existing SfM model, even if created by commercial SfM platforms that lack 3D-associated feature descriptors or depth maps, without relying on such data. This broadens the applicability and utility of the registration process.

For registering the new image and in addition to the base image the inspection started with and the new image that was selected from the pool of retrieved closest images, the concept of shadow image is proposed. Shadow image is another image from the set of base images that is adjacent to the original base image. To automate the process of selecting the shadow image, the proposed algorithm extracts the shadow image by bringing the image next in sequence relative to the original base image (base images are collected sequentially). The poses of the original base image and the shadow image are known from the SfM output. What keeps missing is the pose of the new image.

The foundational step for registering the new image in the proposed approach is extracting dense 2D-to-2D correspondences between the three images: the original base image, the shadow image, and the new image. From the dense correspondences, the common extracted matching features in the three images are retained. Since the poses of the original new image and the shadow image are known, then a direct triangulation is performed on the common matched features between these two images and a 3D point cloud is generated. Moreover, from the common correspondences of the three images, the 2D-to-2D matched features between the original base image and the new image is extracted. The aforementioned steps simultaneously generate the 3D points and their corresponding 2D projections in the new image and these 2D-to-3D correspondences are then fed to a PnP solver to solve the geometric constrain relationship and estimate the 6 DoF pose of the image.

An important outcome of registering the new image to the SfM model is the ability to calculate the new image's projection matrix, which, in turn, facilitates the retrieval of ROIs from the new image. In this study, retrieving ROIs means allowing the inspector to select an ROI from the original base image, and the algorithm autonomously identifies and retrieves the corresponding ROI from the registered new image. The selection of the ROI is achieved in this study by enabling the inspector to simply draw a bounding box around the ROI in the original base image and to retrieve the same ROI form the new image. Hence, interacting with the 3D model to select the ROI points and project the ROI on the new image is not needed. More details provided in the subsequent sections explaining the details of the approach.

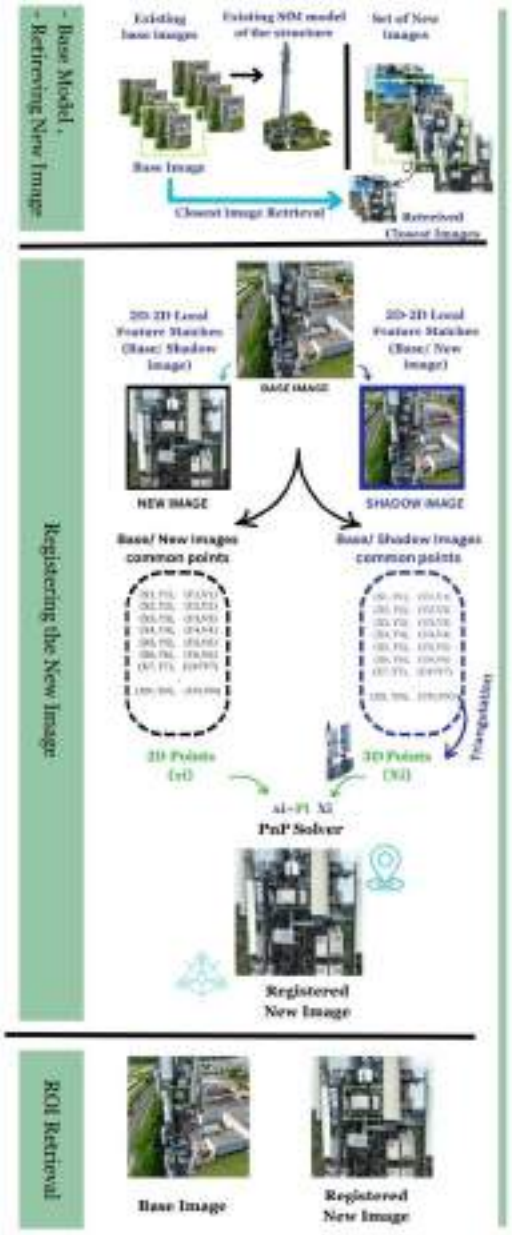


Figure 1. Proposed methodology overview

2.2 Base SfM Model

The SfM technique has been widely adopted for its ability to create accurate 3D scene reconstructions. SfM establishes correspondences from overlapping images to concurrently calibrate unknown camera parameters (intrinsic and extrinsic) while reconstructing the 3D scene. Since SfM can construct the scene and calibrate the extrinsic (position and rotation) of each camera (image), SfM technique was utilized to extract the poses of the base images.

Building high-quality SfM model requires feeding

large number of properly overlapping images to the 3D reconstruction algorithm. SfM quality depends on the quality of features matches across the sequenced images. Hence, employing high-resolution camera to collect the images is necessary to ensure reconstruction quality.

In this work, since the algorithm is implemented on a hard-to-access structure (cell tower) a commercial drone equipped with high-resolution camera was used to collect the images needed for the 3D reconstruction. The images were collected along a pre-defined path designed using a drone mission planning software to ensure the systematic overlapping of the collected images.

2.3 Retrieving Closest New Image(s)

Traditional image retrieval algorithms rely on a database of precomputed image descriptors linked to set of database images. Distance between descriptors in the new and database images are computed and the images with the highest correspondences are retrieved. However, these approaches are computationally demanding and depend on predefined descriptors databases, which may not be available and also, cause potential mismatch.

For UAV-based inspections, drones equipped with sensors like GPS, IMU, and compass typically record and associate metadata or exchangeable image file format (EXIF) data such as camera position, altitude, and heading angle with each collected image. Accordingly, A fast trigonometry-based retrieval algorithm leverages images metadata was developed to efficiently retrieve closest images without relying on complex feature-based methods.

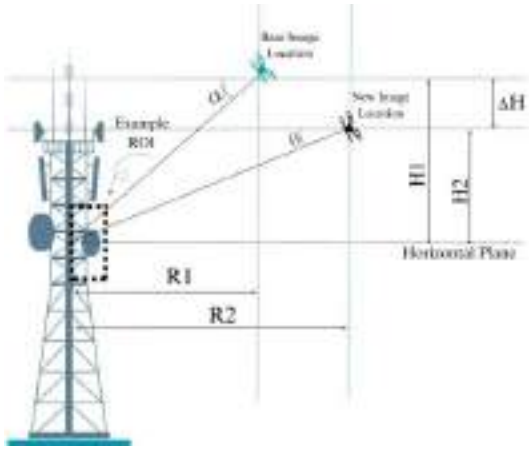


Figure 2. Image Retrieval Illustration

Figure 2 illustrates the proposed technique for retrieving closest new image(s) using a single base image. In the shown illustration, for an arbitrary ROI, a base image is assumed to be captured at a distance (R1) from the ROI with camera pitch angle (α). For the same ROI, a new image is assumed to be captured at distance (R2) from the ROI with camera pitch angle (θ) and looking at the ROI (the main region to be inspected is expected to be near the centre of the image). Subsequently, relative altitudes between base camera and ROI horizontal plane (H1) and between new camera and ROI horizontal plane (H2) and the calculated difference in altitude (ΔH_{calc}) can be calculated as follows:

$$H1 = R1 * \tan(\alpha) \quad (1)$$

$$H2 = R2 * \tan(\theta) \quad (2)$$

$$\Delta H_{calc} = R1 * \tan(\alpha) - R2 * \tan(\theta) \quad (3)$$

The trigonometric relationships in Equations (1-3) form the foundation for EXIF-based image retrieval algorithm. To retrieve the new image(s) closest to the base image, the algorithm starts by reading the heading angle for the base image and all the new images. New images with similar heading to the base image are retained and remaining images are ignored. Then, for the base image and all retained new images, the algorithm will read the longitude, latitude, altitude, and camera pitch angle of each image and calculate (R1), (R2), (H1), (H2), and subsequently, the actual altitude difference (ΔH_{actual}) between the base camera and new camera. Simply, closest images can be filtered by selecting new images with (ΔH_{actual}) close to (ΔH_{calc}). To filter images, either by heading or by ΔH similarity, the inspector can set threshold for the tolerance of the heading or the altitude differences. With higher tolerance, images with larger change in perspective might be retrieved, and vice versa.

This algorithm aims to enhance the speed and quality of UAV image retrieval for images with GPS information. In the case of operating the drone in GPS-denied zones, the conventional content-based image retrieval algorithm can be used to retrieve the closest image.

2.4 New Image Registration

This study introduces improved approach for image registration that leverages state-of-the-art deep learning-based feature matching algorithms to simultaneously construct the 3D scene points and the 2D corresponding points of the new image without the need for pre-defined database of feature descriptors.

As discussed in the "Overview of The Technique" section, three images are utilized to enable registering the new image: base image, its associated shadow image, and a retrieved new image. Robust feature matchers are required to ensure extracting dense matches between the base, shadow, and new image even if the new image was captured under challenging conditions compared to the base image. Deep learning-based feature matching algorithms are employed to ensure accurate and dense feature extraction, even in challenging conditions like varying lighting, perspective, and scale [15].

Accurately estimating pose of the new image enable

calculating the projection matrix. Then, the algorithm can retrieve ROIs selected from the base images and localize the same ROIs in the query images.

3 Experimental Validation

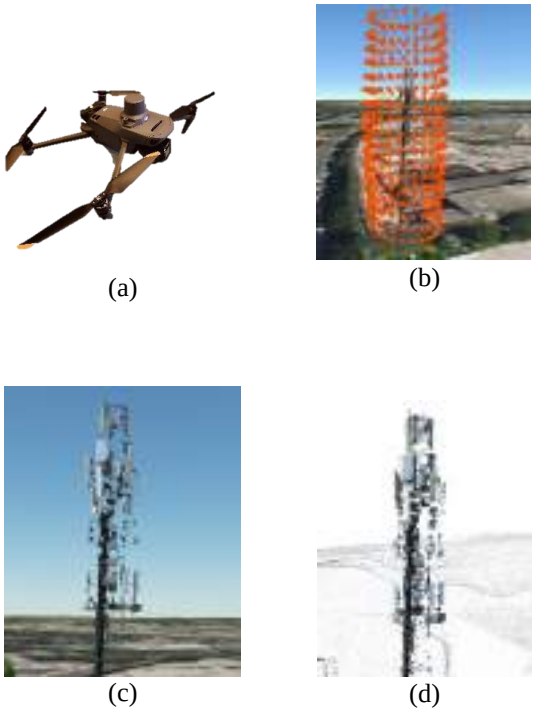
To validate the proposed approach, a lattice telecommunication cell tower in Waterloo (Ontario, Canada) was selected for testing the proposed algorithm. A cell tower was selected for validation as it is type of large and hard-to-access structures that requires systematic and frequent inspections due to its significant and importance as a vital telecommunication infrastructure. The selected tower contains multiple components with highly overlapping wiring and equipment. The cell tower height is approximately 47 meters, with a base width of around 5.5 meters and a width of 4 meters at the location of the tower components.

3.1 Base Cell Tower SfM Model

To construct the base model, we used a commercial DJI Mavic 3 Enterprise drone to capture high-resolution images of the cell tower for the purposes of building high quality SfM model. The drone comprises dual cameras: wide-view camera for regular images, and tele camera with 7x optical zoom. Moreover, the drone is equipped with RTK-GPS. Accordingly, with each collected image, the set of extracted EXIF data contain information needed for executing the various steps of the proposed algorithm. The open-source tool, ExifTool [16], was used to retrieve EXIF data for each image.

To build high quality SfM model, a set containing large number of high-quality and overlapping images of the cell tower was collected. For better results, the collected images should be sequenced and capturing the different components of the cell tower at different perspectives. To ensure systematic collection of images, a commercial software, Drone Harmony [17], for flight path planning and autonomous drone mission was utilized. The planned flight path for image collection was a helical path with a proper pitch distance between helix turns that ensures proper vertical overlap of images. The horizontal overlap of images is guaranteed as the drone is always directed towards the centre of the cell tower. The result of image collection step was a set of 409 images fully covering all the regions of the cell tower. The collected images were then fed to an SfM platform to 3D reconstruct the cell tower.

Different open-source and commercial SfM platforms were tested to create a 3D model of the tower and the platform with the highest quality was selected. In our study, a commercial SfM platform, One3D [18], was utilized to build the base SfM model. The software yielded a 3D reconstruction with relatively high quality compared to other tested platforms. The platform outputs the position and rotation of each camera which allows constructing the projection matrices of all base images.



(a) (b) (c) (d)

Figure 3: Base model 3D reconstruction (a) DJI Mavic 3D used in this study, (b) collected images and helical flight path (c) generated textured mesh (d) generated point cloud

3.2 Retrieving Closest Cell Tower Image

To validate the proposed image retrieval algorithm, a set of new images were collected at a different day and with different perspectives. New images were collected using wide-view camera and tele camera (7x zoom). In the post processing stage, the images were placed in the same folder directory of the base images to prepare for autonomous image retrieval.

As shown in Figure 4, the algorithm was able to successfully retrieve set of closest images relative to the base image. The algorithm retrieved the images taken with the wide-view camera and the tele camera. As shown in Figure 4, the base image was selected on the basis that the region to inspect located approximately at the center of the image, as shown within a bounding box in (a), the retrieved images clearly show that the main region of the base image is clearly shown in the retrieved images shown in (b). The inspector can retrieve more images by increasing the tolerance of heading and difference in altitude discussed in section “2.3 Retrieving Closest New Image(s)”.



Figure 4: Image Retrieval Algorithm Outputs. (a) base image, (b) example retrieved closest images.

3.3 Image Registration Output

The first step in this approach is to select local feature matching algorithms that can build dense 2D-to-2D common correspondences between base and shadow image and between base and new image. By common correspondences we mean features that are shown in the base, shadow image, and new image, simultaneously.

Three state-of-the-art deep learning-based detector-free local feature matching algorithms were employed in this study to build the dense correspondences: Efficient LoFTR [19], RoMa [20], and AspanFormer [21]. The RoMa model creates feature pyramids by integrating coarse-level features with fine-tuned specialized features, using classification-based loss for global matches and robust regression loss for fine-tuning to improve correspondence accuracy. Efficient LoFTR is an extension to LoFTR (Local Feature Matching with Transformers) algorithm. It enhances LoFTR by refining attention mechanisms and incorporating a two-stage correlation layer to enhance efficiency and accuracy in feature matching. AspanFormer is another transformer-based algorithm that employs a hierarchical attention structure and adaptive attention spans to dynamically focus on relevant regions. This adaptive design ensures accurate feature matching across varying scales and challenging scenarios. Each algorithm approaches the feature matching problem from a different perspective. Hence, in this study, the three algorithms were cascaded together to build dense correspondences that are spread across the TRI. Figure 5 shows the results of common features between the base and shadow and between base and new image. It worth noting that all features captured in the non-overlapping regions in the three images were discarded to reduce outliers and that feature that are not common in the three images were excluded.

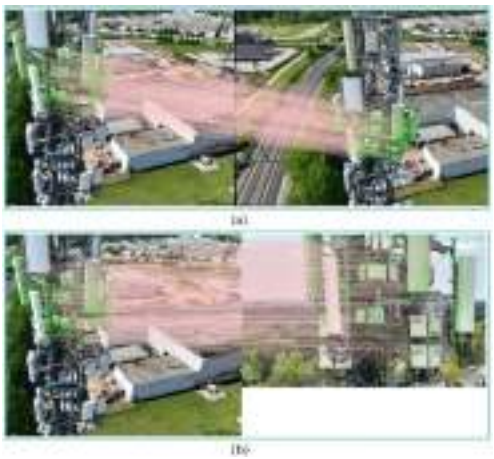


Figure 5: Feature Matching Output. (a) base image with shadow image, and (b) base image with the new image.

To build the 2D-to-3D correspondences required to localize the new image, the common features between the base and shadow images, show in Figure 5 (a), were triangulated to generate the 3D scene points and the common features between the base and new image, shown in Figure 5 (b) were extracted as the 2D projections.

The PnP algorithm was employed to minimize the reprojection error of a geometric constrain relationship. For PnP solver, and in addition to the 2D-3D correspondences, the intrinsic matrix **K** for the camera is needed and was acquired by calibrating the drone camera in the lab using camera calibration board. Additionally, a PnP solver type should be specified. For this study, the Efficient PnP solver (EPnP [22]) was implemented within a Random Sample Consensus (RANSAC) framework, in order to handle outliers.

To test the proposed algorithm, the estimated camera position of the new image should be compared to ground truth values. However, since new images have unknown ground truth values and base images have known calibrated poses, 10 base images were selected from the pool of base images and were treated as new images to evaluate the localization approach. Table 1 shows the absolute distance between actual and estimated camera positions.

Table 1. Evaluation of Image Localization Approach

Images ID	Error in Camera Position	Error in Camera Rotation (degrees)
1	0.066	0.739
2	0.022	0.180
3	0.090	0.078
4	0.075	0.544
5	0.078	0.094
6	0.061	0.685
7	0.017	0.314
8	0.041	0.089
9	0.094	0.371
10	0.064	0.620

It can be shown from Table 1 that the average error in camera position is around 0.06 m and the average error in rotation vector is 0.37°. This implies that the proposed dense feature matching-based camera pose estimation approach can accurately and reliably register new images.

In addition to the 10 base images, the whole dataset was tested to investigate the ability of the methodology to register images with different view angles and perspectives. The average mean error in camera position was found to be 0.069 m and the mean error in rotation vector was found to be 0.41°.

For the case of cell tower, and due to the fine and complex details on the structure, the image resolution was kept high during the feature matching process. This is needed to ensure building robust correspondences that guarantees high registration accuracy. High resolution images increase the inference time for deep learning-based matchers. In our experiment, an NVIDIA GEFORCE RTX 3080 GPU with 12 GB VRAM was used. The inference time for each pair takes ~2.5 seconds using AspanFormer, ~7 seconds using EfficientLOFTR, and ~9 seconds using RoMA. This inference time makes this method not compatible with real-time applications. However, the method is designed to enhance offline visual inspection and ROI retrieval from new images.

3.4 Localizing ROIs in New Cell Tower Images

Once the new image is localized, the inspector can map ROIs from base (old) images to new images to track and assess the condition changes of the ROI. In this paper, and to avoid the need to interact with the 3D scene of the structure to identify the ROI, the algorithm is designed so that the inspector selects an ROI in the base image and the features within that ROI and their corresponding features in the shadow image will be selected and triangulated using both images' projection matrices. The 3D points of the ROI are then projected back in the new image using the calculated projection matrix estimated using during image registration. Figure 6 presents different examples of ROIs selected in the base image (with the letter B in the corner) and the corresponding ROI retrieved from the new image (with the letter N). The last column in Figure 6 (with letter Z) is for zoomed retrieved ROI.

It worth mentioning that for this example the new image was selected from the set captured by the tele camera with 7x zoom. This reflects the robustness of the algorithm in localizing images and retrieving ROIs even in challenging scenarios and severe change in perspective, illumination, and scale. Figure 6 illustrates two main advantages of the proposed approach: 1) select ROIs within the region of focused inspection and retrieve the same ROI with clear details to assess condition inspection and temporal changes (as shown in first three rows of Figure 6), and 2) Inspect changes in asset dismantling or installation on the cell tower as shown in the last row where the retrieved ROI shows the change in asset installation on the selected region.



Figure 6: ROI retrieval results. First column for selected ROI in base image (B). Second column for retrieved ROI from the new image (N). Third column for zoomed in view of the ROI (Z).

4 Conclusion

This study introduces and validates an innovative method for enhancing image localization and pose estimation by leveraging state-of-the-art feature matching algorithms. Key contributions in this study include:

1. An efficient image registration method that overcomes the limitations of existing methods that require database of feature descriptors associated with 3D scene points.
2. Simplified metadata-based image retrieval algorithm for drone-collected images.
3. Accurate ROI retrieval, allowing inspectors to select ROIs on a base image, and automatically retrieve and localize the same ROI in the new image, without the need to interact with the 3D reconstruction for ROI selection.

The image localization algorithm achieved an absolute error in camera position of around 6 cm. Tests on new images from a 7x optical zoom camera shows the performance of the algorithm even in challenging scenarios of image environment. This approach significantly improves UAV-based inspections of large and hard-to-access infrastructure, offering broad applicability for diverse structures.

References

[1] B. F. Spencer, V. Hoskere, and Y. Narazaki. Advances in computer vision-based civil infrastructure inspection and monitoring. *Engineering*, 5(2):199–222, 2019.

[2] W. W. Greenwood, J. P. Lynch, and D. Zekkos. Applications of UAVs in civil infrastructure. *J. Infrastruct. Syst.*, 25(2):04019002, 2019.

[3] S. Chen, D. F. Laefer, E. Mangina, S. M. I. Zolanvari, and J. Byrne. UAV bridge inspection through evaluated 3D reconstructions. *J. Bridge Eng.*, 24(4):05019001, 2019.

[4] L. Zhou et al. UAV vision-based crack quantification and visualization of bridges: system design and engineering application. *Structural Health Monitoring*, 2024.

[5] M. Mandirola, C. Casarotti, S. Peloso, I. Lanese, E. Brunesi, and I. Senaldi. Use of UAS for damage inspection and assessment of bridge infrastructures. *International Journal of Disaster Risk Reduction*, 72:102824, 2022.

[6] J.-L. Xiao, J.-S. Fan, Y.-F. Liu, B.-L. Li, and J.-G. Nie. Region of interest (ROI) extraction and crack detection for UAV-based bridge inspection using point cloud segmentation and 3D-to-2D projection. *Automation in Construction*, 158:105226, 2024.

[7] C. M. Yeum, J. Choi, and S. J. Dyke. Automated region-of-interest localization and classification for vision-based visual assessment of civil infrastructure. *Structural Health Monitoring*, 18(3):675–689, 2019.

[8] N. M. Shajahan, T. Kuruvila, A. S. Kumar, and D. Davis. Automated inspection of monopole tower using drones and computer vision. In *Proceedings of the 2019 2nd International Conference on Intelligent Autonomous Systems (ICoIAS)*, pages 187–192, Singapore, Singapore, 2019.

[9] H. Manninen, C. J. Ramlal, A. Singh, J. Kilter, and M. Landsberg. Multi-stage deep learning networks for automated assessment of electricity transmission infrastructure using fly-by images. *Electric Power Systems Research*, 209:107948, 2022.

[10] Z. A. Al-Sabbag, C. M. Yeum, and S. Narasimhan. Enabling human–machine collaboration in infrastructure inspections through mixed reality. *Advanced Engineering Informatics*, 53:101709, 2022.

[11] M. Xu et al. A critical analysis of image-based camera pose estimation techniques. *Neurocomputing*, 570:127125, Feb. 2024.

[12] A. Kendall, M. Grimes, and R. Cipolla. PoseNet: A convolutional network for real-time 6-DOF camera relocalization. *arXiv preprint arXiv:1505.07427*, 2015.

[13] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, 1981.

[14] E. Deretey, M. T. Ahmed, J. A. Marshall, and M. Greenspan. Visual indoor positioning with a single camera using PnP. In *Proceedings of the 2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pages 1–9, Banff, AB, Canada, 2015.

[15] S. Xu, S. Chen, R. Xu, C. Wang, P. Lu, and L. Guo. Local feature matching using deep learning: A survey. *Information Fusion*, 107:102344, 2024.

[16] *ExifTool by Phil Harvey*. On-line: http://exiftool.org, Accessed: Dec. 26, 2024.

[17] *Drone Harmony*. On-line: https://app.droneharmony.com, Accessed: Dec. 22, 2024.

[18] *One3D - guide*. On-line: https://app.one3d.ai/dashboard/guide, Accessed: Dec. 22, 2024.

[19] Y. Wang, X. He, S. Peng, D. Tan, and X. Zhou. Efficient LoFTR: Semi-dense local feature matching with sparse-like speed. *arXiv preprint arXiv:2305.09016*, 2024.

[20] J. Edstedt, Q. Sun, G. Bökman, M. Wadenbäck, and M. Felsberg. RoMa: Robust dense feature matching. *arXiv preprint arXiv:2305.15404*, 2023.

[21] H. Chen et al. ASpanFormer: Detector-free image matching with adaptive span transformer. *arXiv preprint arXiv:2208.14201*, 2022.

[22] V. Lepetit, F. Moreno-Noguer, and P. Fua. EPnP: An accurate O(n) solution to the PnP problem. *Int. J. Comput. Vis.*, 81(2):155–166, 2009.

Bayesian Networks for data contextualization in Digital Twins of complex civil infrastructures

Leonardo Binni and Anna Brunetti and Fabrizio Gara and Berardo Naticchia

Polytechnic University of Marche, Italy
l.binni@pm.univpm.it, a.brunetti@pm.univpm.it, f.gara@univpm.it, b.naticchia@univpm.it

Abstract –

The management of complex civil infrastructures must adopt the principles of complex systems, focusing on identifying and modeling emergent behaviors for effective decision-making and predictive insights. Current methods face limitations in cross-scale and cross-domain analyses: process-based models are confined to single domains and fail under uncertainty, while data-driven approaches lack explainability, risking misaligned human-machine decisions. This work addresses this challenge by employing expert-elicited Bayesian Networks (BNs) to enhance the interpretative capabilities of Digital Twins (DTs) of civil infrastructures. BNs enable the modeling of expert knowledge regarding significant cross-domain and cross-scale relationships within complex systems, contextualizing data to develop knowledge and explicitly identify emergent system configurations. A case study focusing on the behavior of a reinforced concrete bridge pier under varying scenarios demonstrates the methodology. Qualitative testing validated the BN model's ability to reflect potential system behaviors and provide predictive insights. This study serves as a foundational step toward the development of civil infrastructure DTs with interpretative and decision-support capabilities. The presented knowledge development process facilitate the recognition of patterns that would otherwise remain undetected or be recognized only following time-consuming expert knowledge integration processes missing operational and emergency response needs.

Keywords –

Bayesian Networks; Data Contextualization; Digital Twin; Civil Infrastructure; Bridges;

1 Introduction

Large-scale civil infrastructures such as road infrastructures are complex systems [1] featuring nonlinear, interdependent interactions among numerous components across different domains and scales. Each component, though autonomous, influences the behavior of others, often resulting in dynamic equilibrium or instabilities that cannot be deduced by merely combining the rules governing individual components. This is not only due to the mathematical intricacy of the problem but also because of the extreme variability in outcomes that can arise from minimal changes or insufficient definition in input data.

Based on the principles governing complex systems, it is evident that the ability to identify and model their emergent behaviors must serve as the foundational basis for any decision-support tools aimed at forecasting within these systems. The contextualization of all the available heterogeneous information on the system (e.g., structural analyses and traffic flow) through cross-domain and cross-scales agent-based modeling becomes essential to develop knowledge [2] about the system and support meaningful high-level diagnosis [3]. It concerns the integration of observations (data) into a model in which both their consequences and potential causes are propagated.

Digital Twins (DTs) of complex civil infrastructure systems must include data contextualization procedures. A DT operates as a virtual shared representation of the physical system and its heterogeneous parts, aligned with it through both virtual and human agents, also exploiting real-time acquired data. This synchronization permits the perception of the system's true state, highlighting aspects that may be difficult to measure directly in the physical context. A DT of a complex system should allow for the observation of its behaviors and, crucially, the early detection of impending changes through the recognition of emergent behavior, potentially allowing for the anticipation of effects via forward simulations across time horizons that are commensurate with the system's inherent instability, typically of a short duration.

With that in mind, in this work we adopt Bayesian Networks (BNs) as a core modeling approach for data contextualization, also in accordance with the concept of human-centered digital systems. This paper therefore introduces the critical transition from information to knowledge through the use of expert-elicited BNs within

a civil infrastructure DT framework. BNs are thereby intended as a superstructure of agent modeling that allows data contextualization to explicitly identify emergent system configurations for diagnostic and prognostic purposes.

2 Background

Probabilistic inference involves making conclusions about uncertain events using existing probabilities. This approach is based on probability theory, where results are not certain but rather defined by different levels of probability. Probabilistic inference enables the representation of uncertainty and the forecasting of outcomes or choices based on the likelihood of specific results. Bayesian inference is a method of probabilistic inference that involves using Bayes' theorem to adjust the probability estimate of a hypothesis as additional evidence or data is obtained. Bayes' theorem is rooted in Bayesian inference as it explains the connection between conditional probabilities. It is asserted that the probability $P(H|D)$ of the hypothesis H given the data D, is directly related to the prior probability of the hypothesis, $P(H)$, and the likelihood of the data given the hypothesis, $P(D|H)$:

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)} \quad (1)$$

In this context, the prior indicates our assumptions about the hypothesis before observing the data, while the posterior shows a revised belief following consideration of the evidence.

BNs are a class of probabilistic graphical models representing a set of variables and their conditional dependencies through directed acyclic graphs [4]. BNs are well-suited for modeling expert knowledge in situations where uncertainty is prevalent. The concept of subjective probability, which underpins Bayesian theory [5], is rooted in modern probability theory. Unlike classical probabilistic approaches, which rely on frequentist assumptions, Bayesian probability incorporates beliefs and subjective interpretations (e.g., decision-making under uncertainty). BNs therefore excel at interpreting micro-level evidences at a macro scale by identifying patterns within system's state transitions. BNs can also enable abductive inferences at a macro level, while offering the whole range of inference capabilities. Therefore, BNs present a versatile and robust approach to modeling the behaviors of complex systems [3] overcoming many limitations of traditional process-based algorithms, and purely data-driven methods as shown in Table 1. Expert knowledge-based BNs effectiveness lies in a set of key strengths that allow them to fully invert the traditional inferences logic and adapt to complex scenarios:

- explainable inferences: BNs are explicit models enabling human-computer shared awareness,
- expert knowledge-based: BNs are uniquely positioned to leverage expert knowledge, transforming experience and intuition (subjective probability – beliefs) into structured conditional probabilities,
- flexibility with input data: BNs can process partial or incomplete data,
- multi-domain and multi-scale variables,
- bidirectional inferences: BNs can perform causal (deductive) and diagnostic (abductive) inferences,
- data-driven compatibility, which helps prevent the introduction of biases caused by the subjectivity of expert knowledge

Building on the strengths of BNs this study integrates them into a DT framework for civil infrastructure diagnostics.

Table 1. Analysis of different modeling methodologies for diagnostics of civil infrastructures.

Models	Features
Knowledge-based	Multi-domain and scale variables Explainable inferences Handle uncertainty
Process-driven (physics-based)	Explainable inferences Single domain Cannot handle uncertainty
Data-driven	Multi-domain and scale feature Non-explainable inferences

3 Methodology

It is essential to conceptualize the DT not merely as a conventional simulation tool or as a simplistic virtualization of physical systems, although these aspects are certainly included. The DT possesses significant interpretative and decision-making implications, both from the service (operation and maintenance) and proactively resilience (disasters) management perspectives. The interpretive and decision-making features of a DT, as occurs in the human world, require high-level knowledge about the system under observation. This is necessary to produce meaningful and effective diagnoses. However, high-level knowledge is typically not readily accessible without substantial computational efforts or through human experts. Therefore, DTs must develop knowledge (Figure 1) starting from low-level data and information [2]. AI models can contribute to knowledge development on multiple levels. It must be said, however, that while at a low level data-driven and process-driven models prove useful, at a higher level of the pyramid there is a risk in

the former of parallelized human-computer decision-making and in the latter a lack of multi-domain and multi-scale contextualization. Therefore, within the broader framework of a civil infrastructure DT, BN models are hereby intended to serve as high-level multi-scale and multi-domain data contextualization systems.

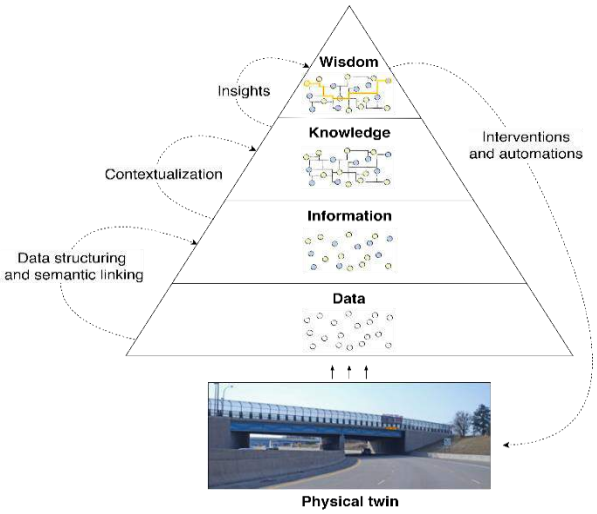


Figure 1. DT's knowledge development processes.

The proposed DT framework builds on a web-based Information Management Framework (IFM) integrating BN models that, by contextualizing a pool of significant evidences and observations through the elicitation of expert knowledge, is used as a bottom-up diagnostic framework for discovering patterns and supporting decisions. Data is collected and can describe events, observations, output from analytical models (e.g., system dynamics, discrete-events simulations, or further agent-based sub-models), real-time sensor data and design and survey data.The IFM taken as reference, and hereafter also generally referred to as DT platform, has been built to support the management processes of assets or systems of heterogeneous assets. It integrates knowledge graphs and unstructured data through a microservice-based architecture with backend and frontend services. The DT platform's backend service acts as an orchestrator of microservices, including a graph database, a document management system, data extraction, integration and query services, a REST API interface and a message broker. The BN models can be integrated with the DT platform as web-based microservices, enabling real-time inferences (Figure 2).

A Bayesian inference service was therefore developed using the Hugin Expert API REST web service [6]. It provides a Bayesian inference process server that can be queried via REST API calls to provide evidence to the network or retrieve the beliefs of nodes of interest after propagating the evidence. BN models can either be loaded or created directly within the web service by instantiating nodes and relationships through the Hugin REST API.

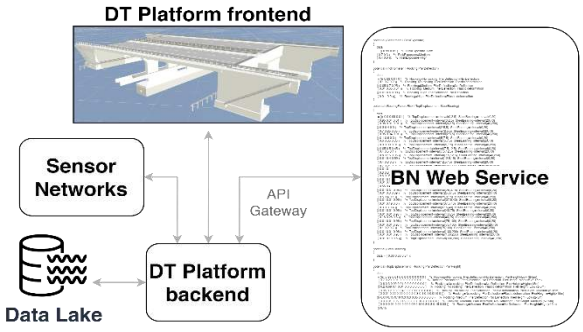


Figure 2. Schematization of the DT platform integrating the BN web service as a virtual sensor.

3.1 Case study implementation

To demonstrate the methodology, this section presents the development of a BN model representing the behaviors of a Reinforced Concrete (RC) bridge. The bridge is considered to have a simply-supported span scheme (Figure 3a). Specifically, the model focuses on the representation of high-level reasoning about the behavior of a single bridge pier, which under different actions is subject to different amplitudes of displacement at its top (Figure 3b), and may experience varying levels of bearing failure. The case study is not properly a large-scale complex infrastructure system by itself, but rather a significant part of it, given the number of multi-domain variables that influence its behavior. This case study is selected for its comprehensibility, facilitating the discussion and giving an understanding of the proposed approach along with its applicability and scalability to wider civil infrastructure systems. Key considerations include the following:

- expert knowledge was elicited to identify key variables, states, and interactions across scales and domains through five iterative 3-hour consultation session among the authors of this paper. .
- the model includes both static data (e.g., design information or data from surveys) and dynamic inputs (e.g., data from inspections, such as newly identified damages, analytical models like acceleration data analysis, or sensor data such as tilt values from inclinometers). This flexibility extends to real-time data integration,
- the BN can be updated with new evidence as it becomes available, ensuring its significance and accuracy over time,
- the BN model was integrated with the DT platform as explained in Section 3 and its output can directly

feed virtual agents as well as trigger actions from human experts,

- scalability and modularity: the network lies on a "plug-and-play" architecture, allowing for expansion and integration with other networks through loose connections, where significant. This modularity enables the development of multiple networks tailored to address further multi-domain and multi-scale interactions,
- for testing purposes, the BN was modeled using Hugin Expert [6] and exported as a HKB file.

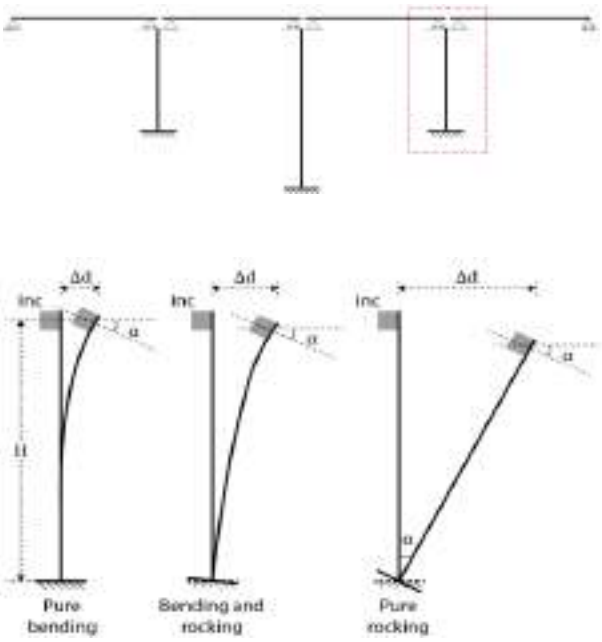


Figure 3. Structural scheme of the bridge pier taken as reference for the case study. The same inclination (α) causes different top displacements (Δd) depending on the pier's behavior.

3.1.1 Analysis of significant variables

An analysis of the significant variables was performed following the *Italian guidelines for existing bridges - classification and management of risk, safety assessment and monitoring* [7]. Given the quantity of the variables involved, a study has been performed to consider only those most appropriate for the purpose of the model presented: variables describing dynamic evidences (e.g., sensor readings or field inspection findings), context conditions (e.g., design data), and behaviors or phenomena (e.g., bending and rocking of piers, or bearing failure). The list of variables considered in the proposed model is shown in Table 2. It is not exhaustive of all possible variables, whose integration within the present model would go beyond the scope of this work. Therefore, the following key assumptions have been made:

- the model schematizes the behavior of a single circular pier of diameter $\boldsymbol{\emptyset}$ and height $\boldsymbol{H}$ in the longitudinal direction of the bridge. The pier is modelled as a fixed-base beam element free to longitudinally move at the top, involving its bending deformability (Figure 3b),
- the bearing can be either a steel or a neoprene support device (Figure 3a),
- the bridge is equipped with accelerometers, and a tilt sensor is positioned at the pier top to measure inclination α (Figure 3b),
- measurements of the relative displacement between the deck and the bearing (Δd) could be measured with displacement transducers. In some case, it might not be available,
- top displacement is caused by bending and/or rocking of the pier in the longitudinal direction.

Since the objective of this research is to validate the digital twinning methodology, these assumptions were necessary to define a specific use case. However, they do not constrain the model's updating with additional variables, limit its application scope and scalability, or compromise the network's effectiveness as a framework for contextualizing data and perform system diagnoses. Additional nodes (variables) and edges (relationships) can be modeled to capture currently omitted variables, interactions and system behaviors.

3.1.2 Model definition

The bridge pier system is described through the modeling of the variables presented in Table 2 as nodes of the network (Figure 4), and through the expert elicitation of conditional probabilities (arcs) in the Conditional Probability Tables (CPTs), which define the relationships between variables. For what concerns the variables, their states were discretized based on values deemed significant for the purposes of the model. For example, continuous variables like inclinometer readings were discretized into three ranges to simplify probabilistic modeling while capturing key behavioral thresholds that correlate with varying damage severities. Regarding the CPTs, an example is provided in Table 3. The table should be read as follows: in a scenario with "No rocking" and "No bending", the bridge pier is most likely (99%) experiencing a low level of inclination (0-0.005 rad). The probabilistic quantification of the network is achieved by specifying the CPTs for each node. For initial nodes (those without parent nodes), this quantification is assigned as prior probabilities. These priors are derived from statistical considerations when available, or by assigning equal probability to each interval into which the variable's domain is divided in cases of total uncertainty. The network's development

reflected the paradigm that effects are conditioned by their causes. Therefore, the BN was designed with a hierarchical structure where variables that represent potential dynamic evidences $\{\alpha, B_d, g\}$ are conditioned by the variables representing phenomena and behaviors $\{R_v, S_f, D_p, \theta, \Delta d, B_{f1}, B_{f2}\}$ that act differently depending on the context condition variables $\{R_e, FT, Ag, K_y, \phi, H, AR, B_s, B_n\}$. The complete model is illustrated in Figure 4.

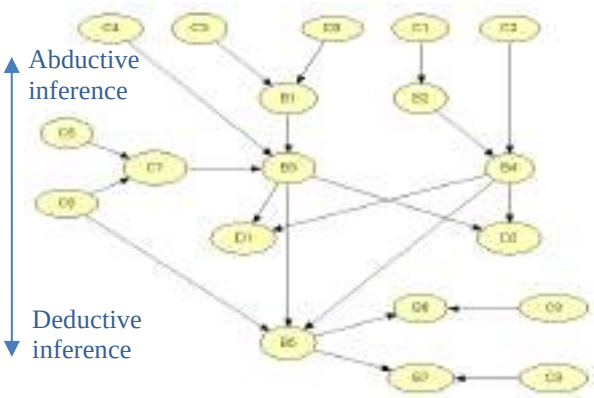


Figure 4. Visualization of nodes (variables) and relationships within the developed BN.

3.1.3 Model validation

Validating an expert-elicited BN involves determining whether the analysis accurately represents the specific concepts it is intended to model for its designated purpose. As a reflection of a mental construct designed to address possible consequences, which may or may not occur, a direct comparison between the BN's model outputs and observations from the real-world system is not feasible. In the context of this study, a comparison with historical data would be inconclusive given the little quantity of available data. However, validation can extend beyond comparisons with observed data by evaluating the internal consistency and logic of the model. Such approaches are widely applied in social science research, system dynamics modeling, and expert-driven BN modeling. A range of specific tests can be performed to assess whether the model meets predefined criteria [5]. Among those, predictive validity tests for BNs can be considered to encompass both the model behavior and the model output. Of particular relevance to establishing confidence in the predictive validity of BNs are Behavior Sensitivity Tests such as the Qualitative Features Tests (QFT). In this test the model response is evaluated in terms of a qualitative understanding on how the system is believed to respond under certain conditions and the sensitivity of the variables. This also helps unearth potential biases introduced by the modeler during the expert knowledge elicitation phase, including those caused by subjective judgment. The QFT is performed by triggering certain states of the BN-variables as inputs, and by inspecting the corresponding model response. The QFT test was performed in this study to qualitatively evaluate the BN and its consistency. Further testing will be performed in future developments, especially to assess the potential bias introduced by the heavy reliance on expert knowledge [5]. In this work, the test was conducted by two domain experts, distinct from those who contributed to the BN modeling phase. The experts have a grounded background in civil engineering with field experience related to the inspection and safety evaluation of existing bridges, obtained by conducting numerous back-office inspections and assessments following the requirements of state-of-the-art Condition Rating Systems [7]. The small validation sample size stems from the specificity of the developed model, which requires civil engineers with proven and established experience in the field of reinforced concrete bridge monitoring and inspection. This may partially limit the comprehensiveness of qualitative validation. However, this limitation does not impact the results of this research that aims at validating a digital twinning methodology. It is part of the future developments of this research to expand the validation sample by including a larger number of domain experts in the validation phase.

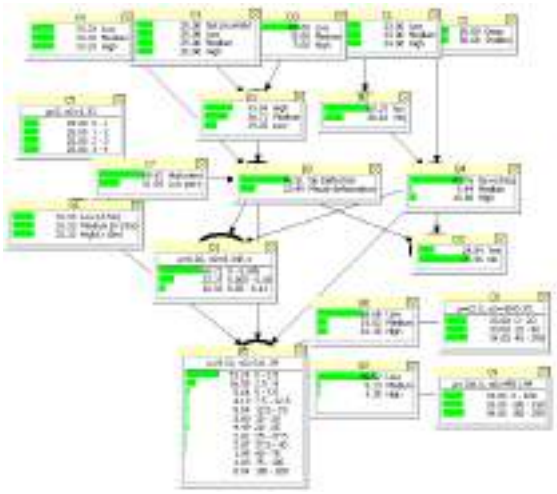
Table 2. Variables used to describe the bridge pier system under consideration.

Variable	Symbol	Node	Evidence from	Discretization
Inclinometer	α	D1	Inclinometer sensor data	{0-0.005, 0.005-0.05, >0.05} [rad]
PiersBaseDamage	B_d	D2	Inspection information	Yes/No
EarthquakeLevel	g	D3	accelerometer data	Low/medium/high
Risk exposure	R_e	C1	Expert assessment	Low/medium/high
FoundationType	FT	C2	Design/survey information	Deep/shallow
DesignGroundAcc	Ag	C3	Design information	Not provided/low/medium/high
RC_Yield_Curvature	K_y	C4	Design information	Low/medium/high
PierDiameter	ϕ	C5	Design/survey information	{0-1, 1-2, 2-3, 3-4} [m]
PierHeight	H	C6	Design/survey information	{2-5, 5-15, 15-40} [m]
AspectRatio	AR	C7	Design/survey information	Slender/Non-slender

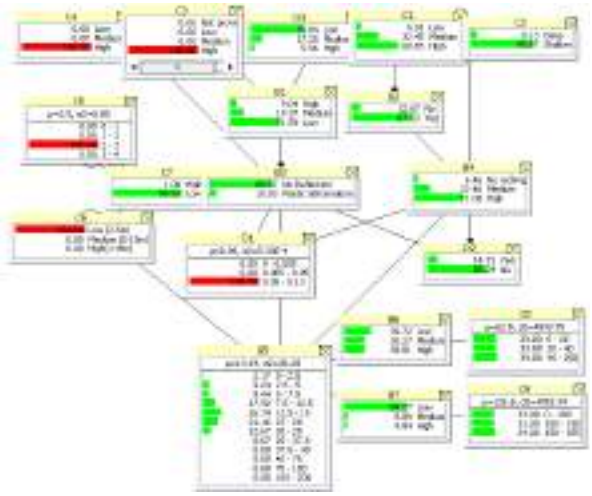
SteelBearing	B_s	C8	Design/survey information	{0-20, 20-40, 40-250} [mm] [8]
NeopreneBearing	B_n	C9	Design/survey information	{0-100, 100-150, 150-255} [mm] [9]
BridgeSeismicHazard	R_v	B1	-	Low/medium/high
Foundation Settlement	S_f	B2	-	Yes/No
Bending	D_p	B3	-	No bending/plastic deformation
Rocking	θ	B4	-	No rocking/medium/high
TopDisplacement	Δd	B5	-	{0-2.5, 2.5-5, 5-7.5, 7.5-12.5, 12.5-15, 15-20, 20-25, 25-37.5, 37.5-40, 40-75, 75-100, 100-200} [mm]
BearingFailure	B_{f1}	B6	-	Minor/moderate/extensive [8]
BearingFailure2	B_{f2}	B7	-	Minor/moderate/extensive [9]

Table 3. Conditional Probability Table (CPT) of the node D1.

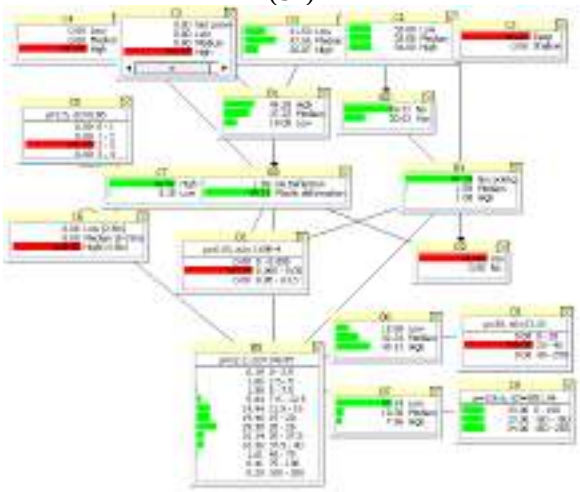
B4	No rocking		Medium		High	
B3	No bending	Plastic deformation	No bending	Plastic deformation	No bending	Plastic deformation
0 - 0.005	0.99	0.1	0.05	0.01	0.01	0.0
0.005 - 0.05	0.01	0.7	0.7	0.65	0.6	0.1
> 0.05	0.0	0.2	0.25	0.34	0.39	0.9



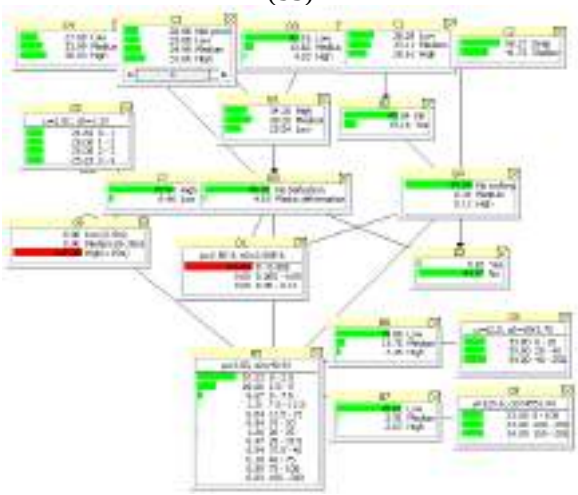
(S1)



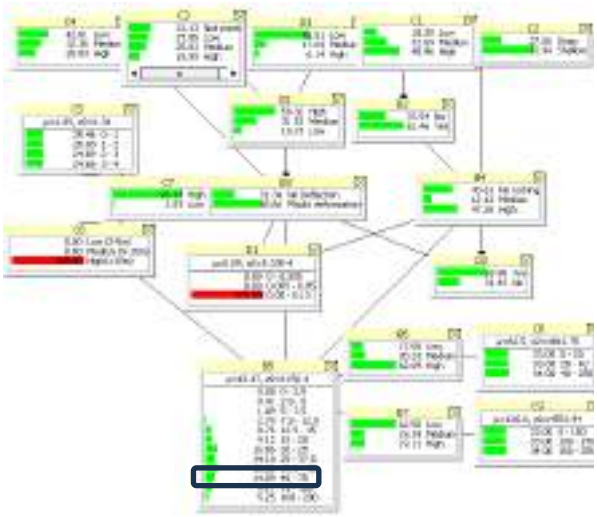
(S3)



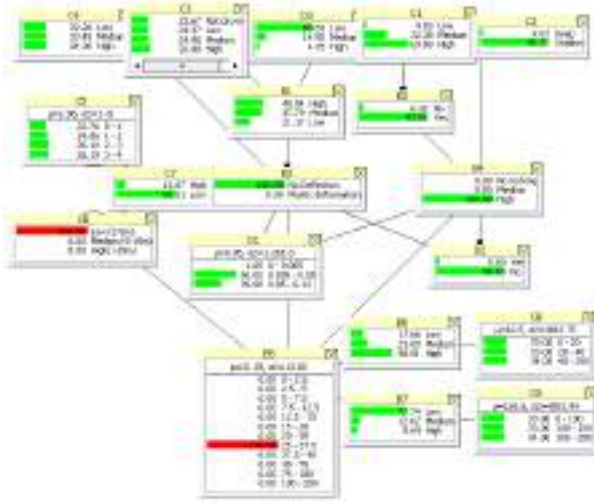
(S2)



(S4)



(S5)



(S6)

Figure 5. Tested scenarios (S1 to S6). Input variables for each scenario are highlighted in red.

Table 4. QFT results.

Test ID	Plausibility assessment
S1	The model schematizes a general conditions in which higher risks are not very likely, which demonstrates a good degree of generalization.
S2	A deep foundation implies a lower chance of rocking. Evidence of bending damage at the base of the pier implies a flexural plastic deformation even though the bridge was designed with appropriate standards. The inclination observed is therefore only given by bending, causing a medium level of displacement at the top of the pier (Δd). A steel bearing would experience extensive damages with high probability. Of interest is also the abductive reasoning process related to the potential causes of such damage: the beliefs of earthquake intensity grew in its *Medium* and *High* states compared to the baseline.
S3	The height of the pier is considered to be *Low (2-5m)*. The measured inclination angle therefore is most probably caused by rocking instead of bending. It implies that the belief of the state *No* of the *PiersBaseDamage* variables is higher than 85%. The significant inclination angle implies a *High* level of rocking. Nevertheless, the displacement on top seems to be not much influenced. This is plausible since the pier is low and displacement Δd grows along with the pier's height H. The believed displacement Δd would be critical in case of steel bearing causing at most moderate damages.
S4	Non-significant tilt angle is observed, bending and rocking effects are strongly believed to be absent. It implies that the displacement on top of the pier is low. Low or no risks for the bearings.
S5	Extremely significant tilt angle is observed, the model is predicting a combined bending-rocking scenario. Bending is more likely than rocking. Indeed the probability is higher, since there is a 50 % chance that rocking is not present. Displacement on top of the pier correctly increases and becomes significant. High risk of failure for the steel bearings, whereas for the neoprene ones additional evidences are required.
S6	Assuming that the displacement has been measured in a range between 20-25 mm, a low pier is 100% experiencing a significant rocking effect. The reason behind that could be the exposure of the pier and its foundation type, which are respectively believed to be most probably *High* and *Shallow*.

3.1.4 Querying the model

The model was integrated with the DT platform as explained in Section 3. The BN was then queried via the REST API service for testing the following processes:

- loading the developed BN model,
- assigning and propagating evidence after new observations,
- retrieving the beliefs of nodes of interest after evidence propagation.

Table 5 shows an example of a test query aimed at retrieving the a posteriori belief of one of the states of Node B5 after evidences in Nodes D1 and C6 are triggered in Scenario n. 5. The Response Body shows a belief of 14.09% for a top displacement in a range from

40 to 75 mm, which corresponds to the one outlined in Figure 5 (S5).

Table 5. Example of a query to the BN model.

Get the a posteriori belief of Node B5 {40-75 [mm]} in Scenario n. 5
GET https://[redacted]/rest/domain/001210-0000-0000-0000-00000/node/00001210-0000-0000-0000-0000000008/getBelief?~9:
{… "Response Headers": { "content-type": "text/plain;charset=utf-8", "content-length": "11", "server": "Jetty(9.2.z-SNAPSHOT)" }, "Response Body": "14.09374811" }

4 Conclusions

This study aims at enhancing the management of complex civil infrastructures by integrating expert-elicited Bayesian Networks (BNs) with Digital Twins (DTs), offering a robust and scalable approach to modeling multi-scale and -domain emergent behaviors.

While the traditional process-based and data-driven models prove useful for data processing, this study offers a solution to the problem of developing high-level knowledge in DTs by endowing them with interpretive value. This is achieved through the development and integration of predictive and explainable BN models. BNs model expert knowledge, enabling the contextualization of heterogeneous multi-domain and multi-scale data (e.g., sensor data streams, observed damages, and design specifications) that are otherwise analytically non-integrable. This contextualization process, capable of assimilating significant new evidence as it arises—even in real time—facilitates the recognition of dynamic system configurations relevant to operational and emergency management needs. In this way, the proposed DT framework functions as a diagnostic and prognostic tool for civil infrastructures, including large-scale systems. It digitizes and accelerates the traditionally slow processes of expert knowledge integration.

The efficacy of this approach was demonstrated through a case study focusing on the dynamic behaviors of reinforced concrete bridge piers. A BN was modeled eliciting the authors' knowledge. The model was first integrated with a web-based DT platform through its REST API interfaces and then validated through Qualitative Features Tests conducted with two external domain experts.

The capacity to bridge human expertise with DT-based diagnostics ensures a cognitive human-machine collaboration that enhances a shared understanding of system behaviors, reflecting the concept of collective intelligence that characterizes complex systems.

Future developments will focus on the extension of the BN model, on its further validation through additional qualitative and quantitative testing and on the integration of the BN web service with the platform backend through a message broker. This would ensure seamless push communication between the service and the backend of the DT platform enabling real-time queries.

References

[1] S. Flumerfelt, K.G. Schwartz, D. Mavris, S. Briceno, Complex Systems Engineering: Theory and Practice, American Institute of Aeronautics and Astronautics, Inc., Reston, VA, 2019. https://doi.org/10.2514/4.105654.

[2] M. Grieves, DIKW as a General and Digital Twin Action Framework: Data, Information, Knowledge, and Wisdom, Knowledge 4 (2024) 120–140. https://doi.org/10.3390/knowledge4020007.

[3] M. Soleimani, F. Campean, D. Neagu, Diagnostics and prognostics for complex systems: A review of methods and challenges, Qual Reliab Eng Int 37 (2021) 3746–3778. https://doi.org/10.1002/qre.2947.

[4] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988.

[5] J. Pitchforth, K. Mengersen, A proposed validation framework for expert elicited Bayesian Networks, Expert Syst Appl 40 (2013) 162–167. https://doi.org/10.1016/j.eswa.2012.07.026.

[6] A.L. Madsen, F. Jensen, U.B. Kjaerulff, M. Lang, The Hugin Tool for Probabilistic Graphical Models, International Journal on Artificial Intelligence Tools 14 (2005) 507–543. https://doi.org/10.1142/S0218213005002235.

[7] J.C. Matos, V. Nicoletti, J. Kralovanec, H.S. Sousa, F. Gara, M. Moravcik, M.J. Morais, Comparison of Condition Rating Systems for Bridges in Three European Countries, Applied Sciences 13 (2023) 12343. https://doi.org/10.3390/app132212343.

[8] J.B. Mander, D.-K. Kim, S.S. Chen, G.J. Premus, Response of Steel Bridge Bearings to Reversed Cyclic Loading, 1996.

[9] R.A. Cook, D.T., Allen, M.H. Ansley, Stiffness Evaluation of Neoprene Bearing Pads Under Long-Term Loads, 2009.

On-site Semantic Mapping and Waypoint Planning for Autonomous Aerial Bridge Monitoring

Yohan Kim[1], Sunwoong Paik[1], and Hyoungkwan Kim[1]

[1]School of Civil and Environmental Engineering, Yonsei University, Republic of Korea.
E-mail: homez815@yonsei.ac.kr, tjsdnd4066@yonsei.ac.kr, and hyoungkwan@yonsei.ac.kr

Abstract –

Effective monitoring of aging bridges is critical for ensuring their safety and maintenance. This study introduces a framework for on-site autonomous aerial bridge monitoring using sensor fusion and SLAM (Simultaneous Localization and Mapping). The proposed method utilizes a lightweight LiDAR sensor and a mini PC onboard a drone to generate real-time 3D semantic maps and flight waypoints. YOLOv8-based image segmentation is employed to identify bridge components, achieving a mean Average Precision (mAP50-95) of 86.6% across test data. Segmentation requires less than 10 milliseconds per frame, while processing LiDAR point clouds takes less than 1 second per frame. Waypoint generation based on the semantic map is completed in under 3 seconds. These results demonstrate the framework's capability to deliver precise and reliable on-site monitoring. This system provides a significant advancement in autonomous aerial bridge inspection by enabling efficient and real-time operation.

Keywords –

Autonomous Drone; Bridge Monitoring; Semantic SLAM; Sensor Fusion

1 Introduction

The need for rapid and precise monitoring of aging bridges has become increasingly critical, and drones have been demonstrated as effective tools for this purpose in various studies [1]. Particularly, autonomous drone navigation minimizes errors caused by subjective human judgment and ensures consistent data quality, making it a promising technology. However, traditional drone-based inspections often rely on predefined flight paths and manually generated waypoints, which can lead to incomplete coverage and inefficiencies in dynamic environments. Additionally, many existing approaches depend on pre-built 3D models, which are often unavailable for aging infrastructure or may become outdated due to structural modifications.

Simultaneous Localization and Mapping (SLAM) is a core technology in autonomous navigation, enabling robots or drones to perceive their surroundings and estimate their position while simultaneously creating a map. SLAM can be categorized into Visual SLAM and LiDAR SLAM. Visual SLAM uses cameras to extract features and reconstruct 3D environments, but its performance degrades in environments with poor lighting or limited texture. On the other hand, LiDAR SLAM generates dense 3D point clouds, offering precise distance measurements and real-time mapping capabilities [2], making it particularly advantageous for large-scale infrastructure inspection.

For autonomous bridge monitoring, maps should include semantic information in addition to geometric data. To achieve this, 3D semantic segmentation techniques can be applied, using either traditional or learning-based approaches. Traditional methods are computationally efficient but lack generalization and accuracy, while learning-based methods provide high accuracy but require significant computational resources.

In this study, a drone equipped with a lightweight LiDAR sensor and a mini PC is utilized to generate high-precision maps in real time, eliminating the need for pre-existing models and enabling adaptive waypoint generation based on continuously updated spatial information. By integrating semantic information with LiDAR SLAM, the proposed approach enhances scene understanding, allowing drones to differentiate bridge components and optimize flight paths accordingly. Camera-LiDAR sensor fusion is employed to achieve fast and accurate 3D bridge segmentation, enabling autonomous navigation with improved structural awareness.

As shown in Figure 1, the proposed method comprises three stages. First, bridges are identified using image segmentation based on deep learning. Second, semantic SLAM is performed by assigning classes to bridge point clouds through camera-LiDAR sensor fusion. Finally, waypoints are generated to provide autonomous flight paths. This approach enables effective aerial monitoring even for aging bridges without pre-

existing 3D models.

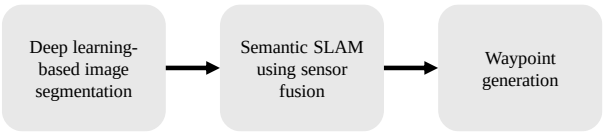


Figure 1. Overview of proposed method

2 Related Works

Most existing studies on waypoint generation and path planning for autonomous drone navigation rely on pre-existing 3D models, such as point cloud, CAD, or BrIM/BIM models. Bolourian and Hammad [3] utilize BrIM/BIM models to define key observation points and apply ray-tracing-based visibility analysis to optimize drone flight paths. Zhao et al. [4] employ a two-stage optimization approach, first determining waypoint locations using genetic algorithms and then refining camera poses through a greedy algorithm to maximize coverage efficiency. Phung et al. [5] model drone path planning as a traveling salesman problem (TSP) and solve it using discrete particle swarm optimization (DPSO), optimizing waypoint sequences to minimize flight distance while ensuring complete coverage. Lin et al. [6] integrate 2D satellite maps for flight planning, enabling flexible data collection missions.

While these methods effectively plan drone flight paths, many aging bridges lack accurate 3D models, and structural changes due to maintenance or disasters may render existing models outdated. To overcome these limitations, some studies have explored on-site mapping approaches that do not rely on pre-existing models. For instance, Narazaki et al. [7] focuses on the rapid post-earthquake inspection of reinforced concrete railway viaducts, specifically targeting damage assessment of columns. To achieve this, their approach generates waypoints based on a progressively built sparse point cloud, enabling drones to autonomously inspect structural elements. While various mapping methods, including SLAM, were considered, their experiments primarily relied on Structure from Motion (SfM) for 3D reconstruction. SfM requires acquiring a sufficient number of images before generating a usable model, which introduces a delay and limits real-time waypoint generation.

This study introduces a real-time LiDAR SLAM-based approach that generates high-precision 3D maps and enables drone-based waypoint generation without relying on pre-existing models. By leveraging LiDAR-generated point clouds with semantic information, our method ensures accurate localization and detailed structural representation, enabling drones to operate based on continuously updated spatial information. This approach facilitates efficient aerial monitoring even for aging bridges without pre-existing 3D models, providing a reliable framework for autonomous navigation in environments where structural changes from maintenance or disasters may render existing models inaccurate or obsolete.

3 Methodology

3.1 Deep Learning-based Image Segmentation

Accurate identification of bridges is a prerequisite for monitoring using autonomous drone navigation. However, processing 3D point clouds in real-time during drone flight is challenging due to computational resource limitations. To address this, the proposed method utilizes deep learning-based image segmentation and sensor fusion to recognize bridges. A real-time deep learning model segments image data captured by the drone's camera into superstructure and substructure. The

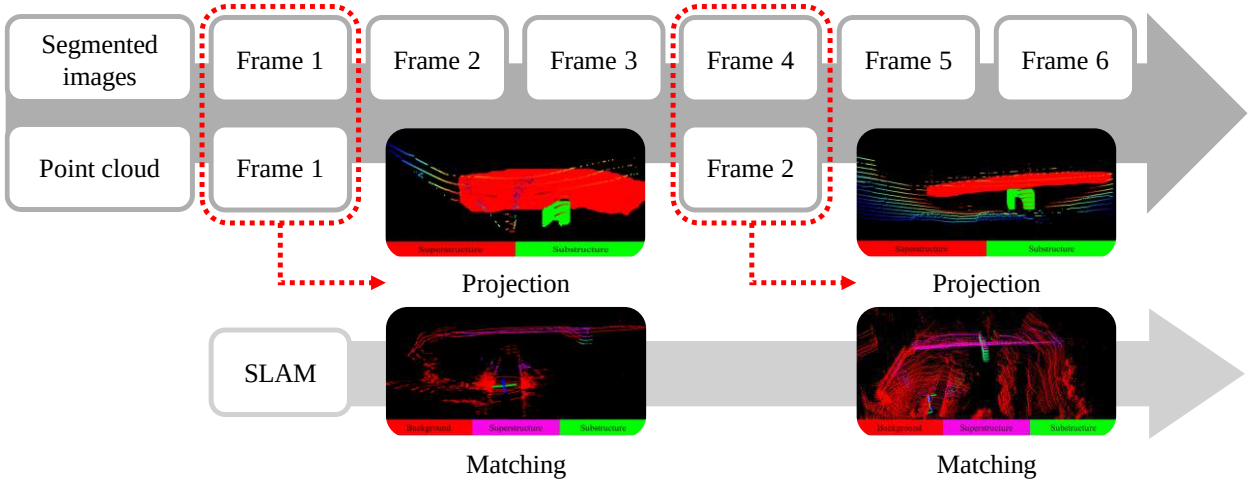


Figure 2. Process of semantic mapping using sensor fusion and SLAM

segmentation results are then used to assign classes to point clouds within the camera view, enabling the system to differentiate between bridge components effectively. This classification is crucial for optimizing waypoint generation and ensuring accurate inspection of critical structural elements.

3.2 Semantic Mapping Using Sensor Fusion and SLAM

To project 3D point clouds acquired by LiDAR onto images, precise calibration between the LiDAR and the camera is required. Calibration aligns the camera's 2D images and the LiDAR's 3D data within the same coordinate system. This involves estimating the camera's intrinsic parameters to correct distortions and matching features between LiDAR point clouds and camera images to calculate the transformation matrix (extrinsic parameters). In this study, a target-based approach using a checkerboard was employed to perform feature matching between the point clouds and images. The checkerboard provides reliable and consistent feature points, ensuring high calibration accuracy.

Once alignment is completed, the segmented images are projected onto the point clouds. The timestamps of the images and point clouds are compared to find the closest frames for matching. Since the number of images is typically greater than that of point clouds within the same duration, the nearest image to each point cloud is selected. Points within the camera's field of view are projected onto the image, and the class of the corresponding pixel is assigned to the points. This ensures that each point in the cloud is semantically labeled with a category, such as superstructure or substructure, and is subsequently utilized in the SLAM and waypoint generation processes. Figure 2 shows the process of semantic mapping in this study.

The LiDAR-Inertial SLAM algorithm integrates data from LiDAR and the Inertial Measurement Unit (IMU) to achieve high-precision real-time localization and map generation. The IMU compensates for motion between LiDAR frames, while LiDAR matching corrects the IMU's drift. This fusion ensures stable performance even during fast movements, rotations, or vibrations of the drone. In this study, calibration between the LiDAR and the drone's IMU was conducted to obtain the transformation matrix, aligning the classified point clouds with the IMU data to generate a real-time semantic SLAM map.

3.3 Waypoint Generation

The generated semantic SLAM map is processed to extract bridge point clouds, which are used as the basis for waypoint generation. First, the point cloud is downsampled to reduce computational load, and noise is removed to enhance the quality of the extracted data. Then, the RANSAC algorithm is applied to extract vertical planes of the bridge, which represent key structural elements such as piers or deck. These vertical planes are critical for determining the spatial boundaries of the bridge and for planning inspection routes.

Second, XY histograms are generated for both superstructure and substructure points, and empty bins around the structure are identified. The center of these bins is used to define the X and Y coordinates, while the Z coordinate is added considering the structure's height to define 3D waypoints. The use of histograms ensures that the generated waypoints are distributed evenly around the structure, optimizing the drone's flight path. This process enables the drone to plan flight paths for efficiently inspecting the superstructure and substructure while maintaining consistent coverage of critical areas.

4 Experiment

4.1 Implementation Details

As shown in Figure 3, the data acquisition platform was built by equipping a drone (DJI Matrice 300 RTK) with a mini PC (Intel NUC) and a LiDAR sensor (Velodyne VLP-16). The camera, LiDAR, and the drone's built-in IMU data were transmitted in ROS (Robot Operating System) message format and stored as bag files. Each message included a timestamp, which was utilized for matching point clouds with images. Data collection was conducted at two bridges in South Korea. Image segmentation was performed in real-time using YOLOv8 [8].

YOLOv8 offers models of various sizes: nano, small, medium, large, and x-large. In this study, YOLOv8n (nano model), which is suitable for real-time processing on CPUs, was selected. The model was pretrained on the COCO dataset and fine-tuned using images captured by the drone camera. Of the total 1,032 images, 722 were used as training data, and 310 as testing data. Figure 4 illustrates an example of annotations.

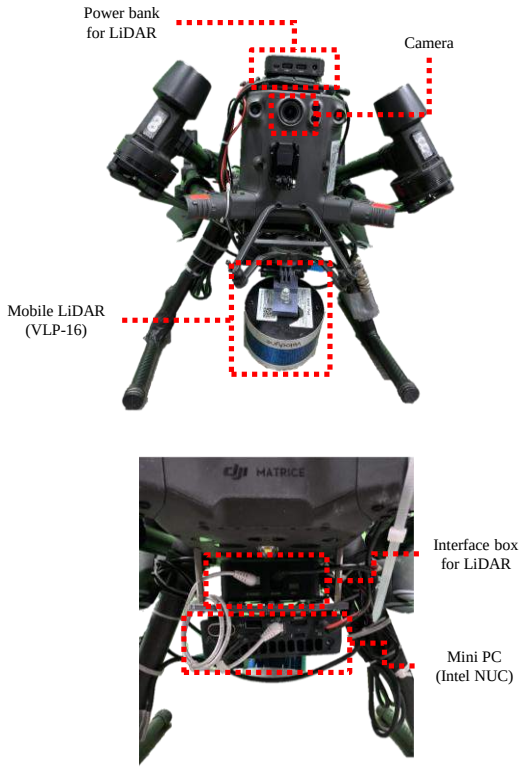


Figure 3. Hardware configuration

The LIO-SAM [9] algorithm was employed for SLAM, organizing sensor data into graph structures and optimizing them to achieve precise localization and mapping. The IMU data was utilized to compensate for the drone's motion, ensuring accurate alignment and consistency within the generated point cloud map. For waypoint generation, it was assumed that the bridge points were sufficiently accumulated on the map, with the initial flight performed using GPS-based or manual control. The map was used for waypoint generation once the bridge points exceeded 100,000 in number.

Semantic mapping, waypoint generation, and the training and testing of the YOLOv8 model were performed on a computer running Ubuntu 18.04 with an Intel Core i7-10700KF CPU and an Nvidia GeForce RTX 3090 GPU.

4.2 Experimental Results

Table 1 and Figure 5 present the quantitative results of image segmentation and examples of segmented images, respectively.

Table 1. Quantitative result of image segmentation

Class	mAP(50-95)
Superstructure	88.1%
Substructure	85.1%
All	86.6%



Figure 5. Visualization of segmented image

The segmentation results achieved an overall mAP(50-95) of 86.6%, where mAP(50-95) refers to the mean Average Precision calculated over IoU (Intersection over Union) thresholds ranging from 0.50 to 0.95 at 0.05 intervals. This metric provides a comprehensive evaluation of the model's detection performance, considering both loose and strict matching

criteria. The slightly higher accuracy for superstructure (88.1%) compared to substructure (85.1%) suggests that superstructure features may be more distinctive or easier to identify in the captured images. This highlights the robustness of the YOLOv8 model in recognizing bridge components under varying conditions, though substructure classification still has room for improvement.

Figure 6 shows the projection of point clouds onto images using the transformation matrix obtained through camera-LiDAR calibration, as well as the generated semantic SLAM map. This figure confirms that the LiDAR data has been accurately projected onto the image and that class information is being appropriately incorporated during the mapping process. The process of waypoint generation are visualized in Figure 7.

In waypoint generation, the bin size of the XY histogram was set to 2 meters, considering the size of the drone and its distance from the bridge. Similarly, the Z-axis interval was also set to 2 meters. Waypoints were initially generated at the midpoint height of the structure, and additional waypoints were generated above and below, considering the structure's height. The generated waypoints were verified against the structural geometry to ensure complete coverage of critical areas. Figure 8 shows the generated waypoints on the SLAM map.

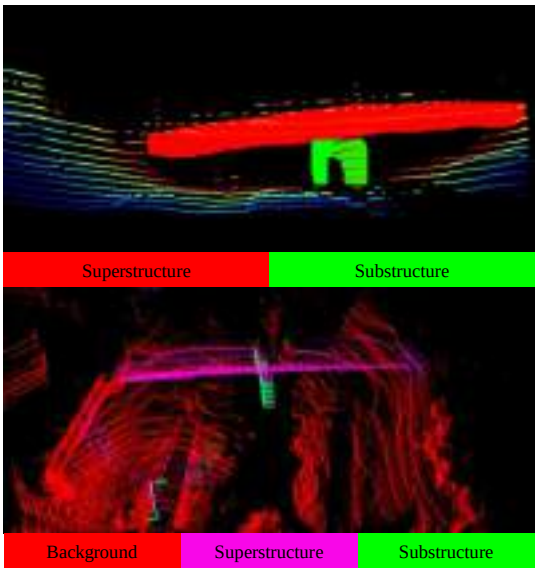


Figure 6. Visualization of projection and mapping

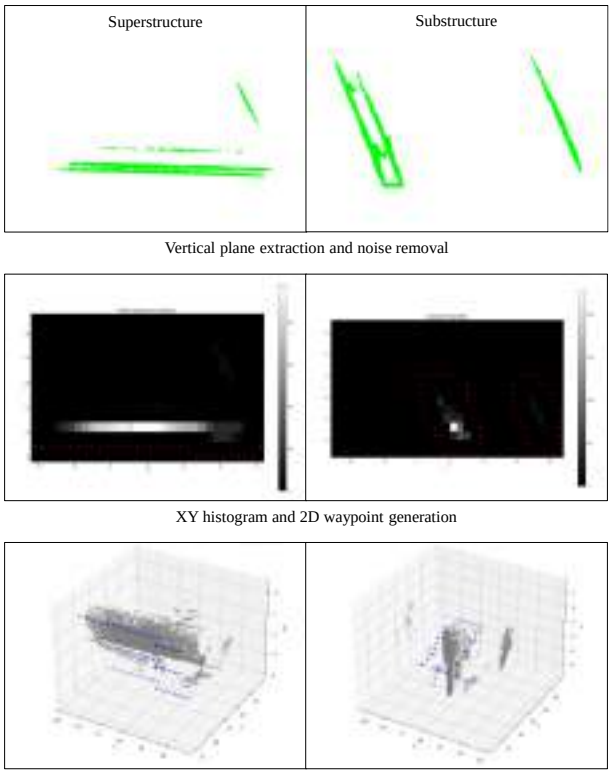


Figure 7. Visualization of waypoint generation result

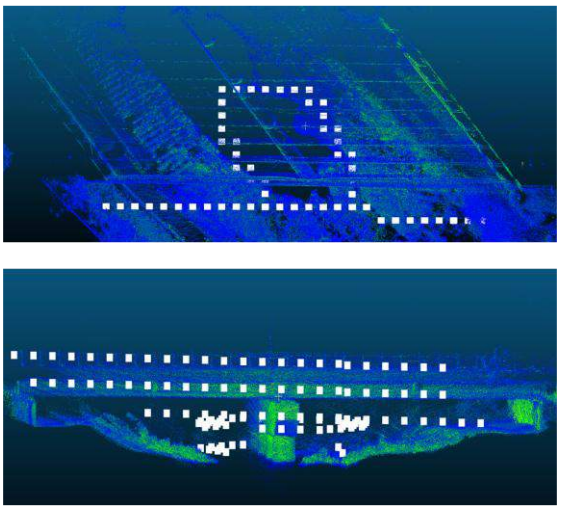


Figure 8. Visualization of generated waypoints on the SLAM map

To validate the effectiveness of the generated waypoints, a coverage evaluation was conducted using the whole bridge point cloud acquired through manual flight. The number of bridge points covered from each waypoint was measured, considering the LiDAR sensor's field of view (FOV). The visibility assessment incorporated a vertical FOV of 30° (-15° to 15°) and a horizontal FOV of 360°. Additionally, the detection

range was limited to a maximum of 10 meters to exclude excessively distant points.

In this experiment, coverage was calculated based on a voxel-based approach. The space was divided into voxels of 0.5m size, and only those containing a certain minimum number of points were considered occupied. Using this information, a ray-casting technique was applied at each waypoint to evaluate visibility. During the ray-casting process, if the line of sight from a waypoint was blocked by bridge components (e.g., girders, decks, or columns) or other obstacles acting as occluding elements, the obstructed area was considered not visible and thus not covered.

As summarized in Table 2, 144 waypoints were generated for Bridge 1, covering 119,956 bridge points, which accounts for approximately 41.21% of the total 291,064 points. Similarly, 113 waypoints were generated for Bridge 2, covering 133,837 bridge points, corresponding to 30.19% of the total 443,280 points.

Table 2. Coverage of bridge point cloud from waypoints

Bridge	Way points	Covered points	Total points	Coverage ratio (%)
1	144	119,956	291,064	41.21
2	113	133,837	443,280	30.19

These results demonstrate that the generated waypoints can effectively serve as an initial exploration path, even with data acquired only from one side of the bridge. The coverage ratios of Bridges 1 and 2 indicate that despite being derived from a limited initial map, the waypoints provided visibility to a significant portion of the bridge structure. As the drone follows these waypoints, it progressively captures more of the bridge, accumulating additional information about the structure. This allows for the continuous refinement of waypoints, enabling a gradual expansion of coverage until the entire bridge is mapped.

The execution time analysis in Table 3 demonstrates the feasibility of on-site processing. Image segmentation requires only ~10 milliseconds per frame, making it highly suitable for on-the-fly data processing during drone flights. Point cloud processing, which includes projection and semantic labeling, takes ~1 second per frame, reflecting efficient handling of LiDAR data. Waypoint generation is completed in ~3 seconds, ensuring that the overall process from image capture to waypoint planning is achieved within ~4 seconds. This performance supports the system’s applicability for on-site operations, allowing for timely and autonomous bridge inspection.

Table 3. Execution time for each process

Process	Time
Image segmentation	~10 milliseconds
Point cloud processing	~1 second
Waypoint generation	~3 seconds
Total	~4 seconds

5 Discussion

The experimental results demonstrate that unmanned and autonomous monitoring of bridges is feasible even in the absence of prior information. The proposed method achieved real-time and highly accurate bridge detection through deep learning-based image segmentation, while sensor fusion and SLAM facilitated the rapid creation of a 3D map containing the necessary bridge information for waypoint generation. The overall processing time was sufficiently short, confirming the method's applicability for on-site operations. However, the proposed method has certain limitations that require further improvement in future research.

First, the process of incorporating image segmentation results into the point cloud needs refinement. Currently, the segmentation result of the single image closest to the point cloud's timestamp is used. If the selected image's segmentation is inaccurate, the error could propagate to waypoint generation. To address this, multiple images could be utilized simultaneously, leveraging only regions identified consistently over a certain number of images, or by evaluating the reliability of segmentation results and selectively incorporating them.

Second, the waypoint generation process requires additional research. This study generated waypoints based on an initial map containing over 100,000 accumulated bridge points, but this map can be continuously updated during the drone's flight. Since the initial map was created using data collected from only one side of the bridge, the opposite side and substructure may not have been sufficiently represented. As the drone collects new data while following the initial waypoints, a dynamic approach could be employed to update the map and generate additional waypoints. For instance, after the drone passes a certain number of waypoints, the map could be updated to reflect new areas, and waypoints could be dynamically generated for further exploration.

Finally, it is essential to ensure the generalizability of the proposed method. While this study demonstrated meaningful results by applying the approach to two bridges, further validation is needed across a wider range of bridge types and environmental conditions.

Addressing these improvements would enable the proposed method to evolve into a more robust and accurate drone-based bridge monitoring system, adaptable to various environments.

6 Conclusion

This study proposed a drone-based monitoring system to address the absence of 3D models and structural changes in aging bridges. By integrating YOLOv8-based image segmentation and camera-LiDAR sensor fusion, bridge point clouds were segmented in real time, while LiDAR-Inertial SLAM enabled stable localization and continuous 3D map generation. The generated semantic SLAM map facilitated waypoint planning, allowing the drone to autonomously inspect bridges. The experimental results demonstrated that the proposed method enables adaptive bridge monitoring and autonomous flight even with limited prior information.

This study focused on adapting and integrating existing techniques into a unified system specifically designed for drone-based bridge inspection. By combining LiDAR SLAM, semantic segmentation, and waypoint optimization, the proposed framework overcomes challenges such as occlusions, real-time adaptability, and the lack of pre-existing 3D models. This integration ensures flexible and efficient coverage of bridge structures, highlighting the practical applicability of this approach in real-world scenarios.

Future research will focus on enhancing waypoint generation by incorporating multi-frame segmentation results and refining real-time map creation and path planning. Additionally, validating the system across a diverse range of bridge types will help assess its generalizability. Further optimization efforts will aim to ensure stable operation in resource-constrained environments, enabling more efficient deployment in large-scale infrastructure monitoring. This study contributes to the advancement of autonomous drone technologies for ensuring the safety and maintenance of aging bridges, demonstrating that existing algorithms can be effectively tailored to meet the specific demands of structural inspection.

Acknowledgements

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Ministry of Education (RS-2018-NR031077) and the Ministry of Science and ICT (RS-2024-00457308).

References

[1] P. Huthwohl, R. Lu, and I. Brilakis. “Challenges of bridge maintenance inspection.” In Proceedings of the International Conference on Computing in Civil and Building Engineering, pages 51–58, 2016.

[2] A. Yarovoi and Y. K. Cho. "Review of simultaneous localization and mapping (SLAM) for construction robotics applications." Automation in Construction, 162:105344, 2024.

[3] N. Bolourian and A. Hammad. "LiDAR-equipped UAV path planning considering potential locations of defects for bridge inspection." Automation in Construction, 117:103250, 2020.

[4] Y. Zhao, B. Lu, and M. Alipour. "Optimized structural inspection path planning for automated unmanned aerial systems." Automation in Construction, 168:105764, 2024.

[5] M. D. Phung, C. H. Quach, T. H. Dinh, and Q. Ha. "Enhanced discrete particle swarm optimization path planning for UAV vision-based surface inspection." Automation in Construction, 81:25–33, 2017.

[6] J. J. Lin, A. Ibrahim, S. Sarwade, and M. Golparvar-Fard. “Bridge inspection with aerial robots: Automating the entire pipeline of visual data capture, 3D mapping, defect detection, analysis, and reporting.” Journal of Computing in Civil Engineering, 35(2):04020064, 2021.

[7] Y. Narazaki, V. Hoskere, G. Chowdhary, and B. F. Spencer Jr. “Vision-based navigation planning for autonomous post-earthquake inspection of reinforced concrete railway viaducts using unmanned aerial vehicles.” Automation in Construction, 137:104214, 2022.

[8] YOLOv8. Online: https://yolov8.org/, Accessed: 24/12/2024.

[9] T. Shan, B. Englot, D. Meyers, W. Wang, C. Ratti, and D. Rus. "Lio-sam: Tightly-coupled lidar inertial odometry via smoothing and mapping." In Proc., 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 5135–5142, 2020.

Development of a Hand-Arm Vibration Syndrome (HAVS) Detection Tool for Construction Workers

S M Jamil Uddin[1], Abdur R. Shahid[2], Mohd. Farhan Israk Soumik[2], Ziyu Jin[1]

[1]Department of Construction Management, Colorado State University, USA
[2]School of Computing, Southern Illinois University, USA
smj.uddin@colostate.edu, shahid@cs.siu.edu, mohdfarhanisrak.soumik@siu.edu, ziyu.jin@colostate.edu

Abstract –

Hand-Arm Vibration Syndrome (HAVS) results from prolonged exposure to hand-held vibrating tools, such as jackhammers, drills, and grinders. Construction workers frequently use these tools for extended periods and are therefore prone to developing HAVS over time. HAVS can affect the vascular, neurological, and musculoskeletal systems of workers, leading to both short-term and long-term disability. Despite its severity, HAVS is often under-reported due to a lack of awareness and delayed detection. Existing methods for detecting HAVS primarily focus on assessing the vibration levels of hand-held equipment rather than the condition of the exposed workers. This study proposes a machine learning tool to assess the hand and arm conditions of construction workers exposed to vibrating equipment. The tool is capable of analyzing hand drawings to detect hand tremors, an early symptom of HAVS. This innovative approach offers a potential solution for the regular monitoring of construction workers' health, enabling early detection and management of HAVS symptoms.

Keywords – HAVS; Machine Learning, Construction Workers Health; Hand Tremor

1 Introduction

The construction industry is one of the most hazardous industries around the world in terms of both fatal and non-fatal injuries. Global reports estimate approximately 60,000 fatal injuries take place within the construction industry around the world [1]. In the United States alone, approximately 200,000 non-fatal injuries and illnesses take place every year and the number go significantly higher when it comes to global injuries in the construction industry [2].

Since construction is a labor-intensive industry, the health and safety of construction workers become significantly important. A major factor affecting the health and safety of these workers is the physical demanding and hazardous nature of their work environment, particularly their reliance on heavy, hand-held machineries such as jackhammers, drills, and grinders to complete tasks [3]. These tools are heavily used in the industry for works such as demolition, drilling, cutting, etc. These tools generate intense vibrations due to their high speed and powerful mechanical operations which are then transmitted to the hands and arms of the workers.

Depending on the equipment type and specifications, the level of vibration can range from moderate to extreme. Studies show that approximately 1.2 million workers in the United States are potentially exposed to hand-arm vibration syndrome risks and among them 500,000 are construction workers [4]. To make things worse, often these tools generate more vibrations than the manufacturers declare. The Hand-Arm Vibration Database [5] on the Physical Agents Portal website collected a total of 4,913 field vibration measurements from 2,897 pieces of equipment and compared their field values with the manufacturer-declared values. This database shows that in the majority of cases, the field values are significantly higher than the declared ones [5]. Exposure to these tools with higher vibration rates and longer duration can cause significant damage to the construction workers' health. Over time, these vibrations can damage soft tissues, nerves, and blood vessels, particularly when workers are exposed for extended periods without adequate breaks or protective measures [6-8]. Additionally, weather conditions, particularly cold and wet environments, significantly worsen the risk of HAVS for construction workers [9]. Cold weather constricts blood vessels, increasing susceptibility to Vibration White Finger (VWF), while wet conditions reduce grip efficiency, amplifying strain on hands and arms [10]. These factors, combined with prolonged outdoor exposure, heighten workers' vulnerability to developing HAVS over time.

Despite these eminent dangers, HAVS is often under-reported or undetected in the construction industry due to the intermittent nature of symptoms. Studies suggest that workers wait more than 3 years after symptom onset before seeking any treatment related to HAVS and

approximately 9 years before receiving a proper assessment [11]. Workers also tend not to report the early symptoms due to lack of awareness. Hence, it is necessary to assess HAVS among the workers regularly, especially those involved in hand-held heavy machinery usage.

This paper proposes a machine learning tool that can assess the workers' hand tremor condition by analyzing their hand drawing. The tool is trained on Parkinson's disease patients' dataset and can analyze the condition of the construction workers' tremor condition and characterize them as HAVS.

2 Hand-Arm Vibration Syndrome (HAVS)

Hand-Arm Vibration Syndrome (HAVS) is a progressive condition caused by continuous exposure to vibration from tools and machinery [12, 13]. It is characterized by symptoms such as tremor, tingling, numbness, reduced grip strength, and loss of fine motor skills, and can affect the vascular, neurological, and musculoskeletal systems [6, 14]. At advanced stages, HAVS can lead to chronic pain and permanent loss of hand function, severely impairing a worker's ability to perform tasks and maintain their livelihood [15, 16]. Vibration exposure can also exacerbate other musculoskeletal disorders, creating compound health risks. Beyond its physical toll, HAVS has economic consequences for both workers and employers, including increased healthcare costs, compensation claims, and decreased productivity [17, 18].

HAVS has been reported across different industries such as manufacturing, mining, forestry, automotive etc. where workers operate vibration-intensive equipment. The prevalence of HAVS among workers regularly exposed to vibration averages 50% and increases over time if corrective measures are not implemented [19]. Studies have shown a high prevalence of HAVS in specific occupational groups, such as foundry workers, where up to 47% of exposed workers may have advanced stages of the syndrome [12].

While HAVS is a concern in different industries, it is particularly prevalent and severe in the construction industry [20]. Workers in the construction industry are regularly exposed to vibrating tools such jackhammers, concrete saws, angle grinders, pneumatic drills etc. The construction environment exacerbates HAVS risks due to prolonged tool usage, lack of awareness, and inconsistent implementation of safety measures [21, 22]. Studies suggest that up to 63% of the construction workers may be affected by HAVS to some extent [23]. In Europe, it is estimated that one in four workers are exposed to either hand-arm vibration or whole-body vibration [24]. In the United States, an estimated 1.45 million workers use powered tools that generate hand-arm vibration, with the prevalence of HAVS ranging from 6% to 100%, with an average of about 50% [25].

3 Point of Departure and Research Objectives

In order to prevent the severe health consequences of HAVS, early detection and intervention is crucial. HAVS detection involves evaluating workers' physical symptoms, functional impairments, and their exposure history [26-29]. There are clinical and technological approaches to detect HAVS among workers. For example, the Vibrotactile test evaluates the level of sensitivity of the mechanoreceptors in the hand [30], while the Thermal Aesthesiometry test assesses the sensitivity of thermal receptors in the hand to changes in temperature [31]. There are other tests such as the Purdue Pegboard test [32], Grip Force Measurement test [33], Cold Provocation test [34] etc. that assess the patients' level of sensitivity and fine motor skills in hands. Although these assessment methods are clinically effective, these strategies can be applied only when the workers self-report some of the symptoms of HAVS. However, in many cases it has been observed that workers often overlook the early symptoms and do not self-report the symptoms [7, 11, 35, 36]. By the time they report their symptoms, it is often too late.

In recent times, researchers have come up with some strategies to detect potential HAVS among construction workers by measuring the level of vibration in hand-held equipment. For example, studies demonstrated the technical feasibility of using the integrated microphone and accelerometer from commercial smartwatches in identifying the level of vibration the hand-held tools generate [25, 37, 38]. However, these methods only assess the hand-held tools' level of vibration rather than the condition of the workers' hands and arms due to the exposure to these vibrations. Additionally, these tools heavily rely on data collected through sensors in smart watches. It is often observed that these sensors collect a lot of noise from surroundings due to the dynamic nature of construction activities.

To address these issues and detect HAVS symptoms among construction workers, this research proposes a machine learning tool that can analyze the hand-drawing of construction workers and assess the condition of their HAVS in real time and continuously over a period. The tool is easy to use and does not require any wearable sensors. The tool can analyze the hand-drawing of workers, especially those exposed to vibrating tools for a longer duration and detect any vibrating syndrome such as tremor in their hands.

4 Methodology

4.1 Development of the HAVS Assessment Tool

Our methodology for developing the HAVS assessment tool is rooted in the motivation of detecting various diseases, such as Parkinson's disease, from hand drawings. Despite originating from distinct causes and affecting different parts of the nervous system, HAVS and Parkinson's disease share similarities, particularly in their effects on motor function and the nervous system. One of the common symptoms for both diseases is hand tremors. Although the causes for hand tremors differ, their effects are indistinguishable. Consequently, given the apparent scarcity of datasets dedicated to HAVS detection, our proposed machine learning model is trained using datasets primarily created for detecting Parkinson's disease [39]. The dataset primarily consists of drawings of wave and spiral shapes, both from healthy individuals and patients affected with Parkinson's. Spiral drawings, in particular, are highly effective in recognizing hand tremors of both resting and active types associated with Parkinson's. This is because tremors cause drawn lines to become relatively uneven and wavy compared to smooth spirals drawn by healthy individuals. As HAVS primarily involves hand arm vibration, it can also be associated with tremoring hands, making this approach particularly useful.

The Hand Arm Vibration Syndrome (HAVS) detection framework proposed in this study can be divided into three distinct modules which will be described in the following subsections sequentially. Figure 1 describes the structure of the proposed system.

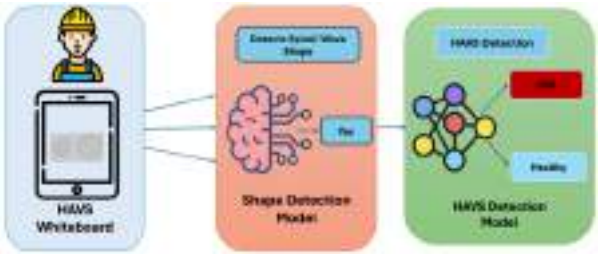


Figure 1. Proposed HAVS System Architecture with three modules.

4.2 HAVS-Whiteboard

To facilitate the collection of drawings from construction workers, a standardized online virtual drawing board is essential. While numerous virtual drawing tools are available, a dedicated tool is necessary to create a compact system that is entirely independent of any third-party application. This system must ensure the privacy and security of construction workers' sensitive information from various malicious inference attacks, such as Health Attribute Inference Attack and Personality Inference Attack, as it is shown that drawing-based inference attacks allow unauthorized intruders to accurately predict the personality and health condition of the concerned individual [40]. We envision to develop the HAVS-Whiteboard as a mobile application.

The data collection process in the HAVS-Whiteboard system begins with initialization of an empty data structure to store all the drawing-related information. Each stroke, defined as the continuous movement of the pen from contact to lift-off, is assigned a unique identifier to ensure proper segmentation of the data. This enables the system to differentiate between individual drawing actions. A sample HAVS-Whiteboard interface is demonstrated in Figure 2.

When a user starts drawing (pen-down event), the system captures the starting position of the pen on the canvas (x, y coordinates), the pressure applied, and the precise timestamp of the interaction. This starting point is the beginning of a new stroke, and a unique stroke ID is generated. As the user continues to draw (pen-move event), the system records data points at regular intervals. Each data point includes the pen's current position, the pressure applied, the angle of the pen, and the timestamp. Additionally, the speed of the pen's movement is calculated in real time by dividing the distance between two consecutive points by the time elapsed between them.

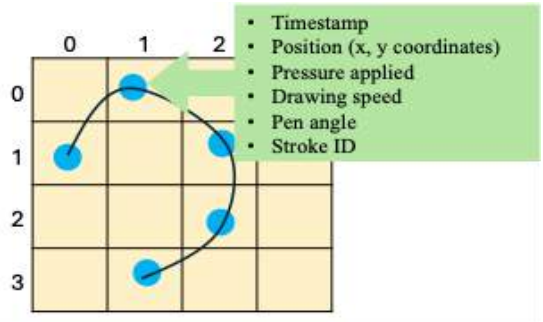


Figure 2. Data collection on HAVS-Whiteboard interface.

When the user lifts the pen (pen-up event), the stroke is finalized, and the system compiles all the captured data for that particular stroke. This process repeats for every new stroke. The collected data is securely stored and includes attributes such as timestamp, position, pressure, speed, pen angle, and the corresponding stroke ID.

In our current implementation, the HAVS-Whiteboard, a Python-based standalone application, captures the (x,y) coordinates and timestamp information. We plan to continue enhancing the application by implementing additional features, which we intend to utilize for data collection from construction workers. Our discussion of the subsequent modules of the architecture is focused on the current state of the implementation.

4.3 Disease Detection Model

Our detection model is built upon the concept of training a deep convolutional neural network model using Spiral and Wave images of the aforementioned dataset. Convolutional Neural Networks (CNNs) unlike the traditional machine learning models like SVM, Random Forest etc does not require manual feature extraction and remarkably effective for identifying hierarchical features like textures, edges etc from images while at the same time showing increased resiliency towards noise and distortions. But the problem arises from the fact that, training a deep CNN model is a tough nut to crack due to the requirement of high quantity of task related quality which is pretty difficult to ensure for each task. This seemingly unsolvable dilemma can be solved by a process which is commonly known as Transfer Learning. Transfer Learning involves finetuning a pretrained model, which has been trained on a large dataset, for a particular task. In this work, we used a pretrained deep CNN model called MobileNetV3Large which has been trained on ImageNet dataset and finetuned for our HAVS detection task. Figure 3 below illustrates the architecture of the HAVS detection model.

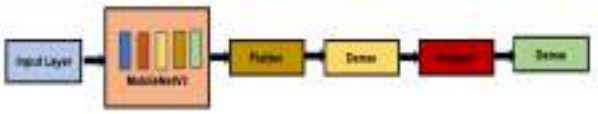


Fig 3. Block Diagram of Detection Model

4.4 Shape Recognition Model

As our proposed detection model is solely trained on Spiral and Wave images, while deploying this very HAVS detection system in real-life scenarios, there remains always a distinct possibility of users generating drawing other than Spirals or Waves. In these types of cases, how robust our detection system may be, it is bound to face failure. So, for uninterrupted operation of the detection model, it must be ensured that only related Spiral and Wave images are forwarded to that system discarding any other shapes. That's where the concept of using Shape Recognition Model comes from. It will primarily work as a filter whose very purpose will be to protect the detection model from unwanted shape images. In this work, we used the same DCNN model called MobileNetLargeV3 that we used for detection task using Transfer Learning. The difference lies in the fact that we finetuned this MobileNetLargeV3 with a dataset to detect whether the input image shape is Spiral or Wave.

MobileNetV3 is an open-source lightweight deep CNN model which has been primarily designed for producing efficient inference results in resource constrained devices such as Smartphones, Embedded Systems and IoT devices. It uses depth wise convolution which separates simpler convolutions into two distinct categories - Depthwise Convolution and Pointwise convolution which reduces the computational cost while still maintaining significant accuracy. Our particular version of MobileNet which is MobileNetV3 uses a squeeze-and-excitation module which helps module to focus on most informative feature channels while ignoring the less relevant ones thus improving its performance over the baseline one. Figure 4 presents the block diagram of the MobileNetV3 architecture, highlighting its key components such as depthwise and pointwise convolutions, bottleneck blocks, and the squeeze-and-excitation module, which collectively enable efficient and accurate feature extraction. As we want to focus on deploying this model in future on edge devices for real time use by the users, this model with its lightweight architecture, less susceptibility to noise and requirement of less computational power becomes a natural choice for the proposed HAVS detection system.



Figure 4: Block Diagram of MobileNet Architecture

5 Experimental Setup and Result Analysis

The entire experiment was implemented in Python using Pytorch framework for training and testing the model and NVIDIA's A100 GPU. Different performance metrices starting from Accuracy, Confusion Matrix, Precision, Recall and ROC-AUC curve have been used for evaluating the performances proposed Shape Recognition and HAVS detection model.

Confusion matrix is a tabular summary of model's predictions compared to actual outcomes which is widely used for evaluating model's performance. Table 1 describes the confusion matrix for both HAVS detection and Shape Recognition Model.

From Table 1, it is clearly visible that our modified finetuned MobileNetV3Large based Shape Detection model performs exceptionally well for both Healthy and Parkinson/HAVS categories where it is nearly perfect while correctly predicting Spiral and Wave Shapes. This Shape detection model works as a gateway to the detection mechanisms, it's highly accurate performance suggests its reliability as a filtering tool. In the case of HAVS model, it is evident that it's ability to differentiate between Healthy and Parkinson/HAVS drawing is significantly better for Spirals compared to Wave drawings where it produces some False Positives.

Table 1. Category Specific Confusion Metrices

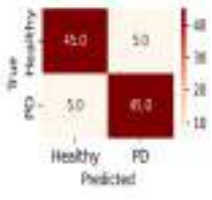
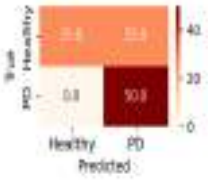
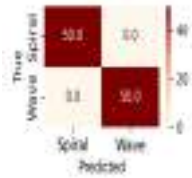
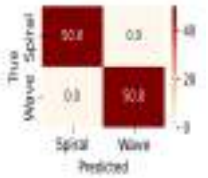
HAVS Detection Model		
	Spiral	Wave
Shape Detection Model		
	Healthy	Parkinson/HAVS

Table 2 demonstrates the category specific results of other metrices such as Accuracy, Precision, Recall, F-1 Score which are usually generated from confusion matrix for both models.

Table 2. Category Specific Results of Performance Metrices.

		Accuracy	Precision	Recall	F-1
HAVS Detection Model	**S**	0.83	0.82	0.87	0.84
	W	0.93	0.93	0.93	0.93
Shape Detection Model	**H**	1.00	1.00	1.00	1.00
	P	1.00	1.00	1.00	1.00

Here S=Spiral, W=Wave, H=Healthy and P=Parkinson/HAVS.

From the above table, it can be observed that the shape recognition model works perfectly for both categories, HAVS model performs well especially for Wave drawings but there are areas of improvement to be done.

6 Discussion

The development of a HAVS assessment tool leveraging machine learning and existing Parkinson's datasets represents an innovative approach to address the challenge of limited HAVS-specific data. By using Spiral and Wave drawings as a means to detect hand tremors, the study successfully demonstrates the potential of integrating this tool as a method of continuous monitoring of hand-tremor development among construction workers. The integration of a custom-built virtual drawing tool i.e., HAVS-Whiteboard, enhances data security and privacy, addressing critical concerns in real-world applications.

Another finding of this research is the effectiveness of transfer learning with MobileNetV3Large in achieving robust performance on both Shape Recognition and HAVS Detection tasks. The Shape Recognition model displayed near-perfect accuracy, acting as a reliable filter to ensure only relevant input reaches the detection model. However, while the HAVS Detection model performed exceptionally well on Wave drawings, its slightly reduced accuracy for Spiral drawings reveals areas requiring further refinement. This discrepancy highlights the importance of feature representation in the input data, as Spiral patterns may involve subtler tremor-induced variations compared to Wave patterns.

Additionally, the study underscores the importance of implementing a lightweight architecture for edge deployment. MobileNetV3's efficiency, low computational overhead, and resilience to noise make it a natural choice for resource-constrained environments, such as construction sites.

The inclusion of multiple performance metrics ensures a comprehensive evaluation of the proposed system. However, analyzing false positives and negatives reveals opportunities for improving the model's ability to differentiate between Healthy and HAVS-related cases, particularly in complex scenarios. Furthermore, the reliance on Parkinson's datasets introduces potential biases, necessitating the collection of HAVS-specific data to enhance generalizability and reduce overfitting to Parkinson's-like features.

7 Conclusion

The study presents the development of a machine-learning-based tool designed to detect Hand-Arm Vibration Syndrome (HAVS) among construction workers. This innovative tool utilizes drawing patterns, captured through timestamps and (x, y) coordinates of drawing, to identify potential symptoms of HAVS.

While the current proof of concept has been validated using simple spiral and wave-like drawings, future iterations of the tool will incorporate additional features to enhance accuracy. In the next phase of development, we will expand the dataset by collecting real-world drawing samples from construction workers who are exposed to hand-held heavy machinery and integrating richer data, such as pressure, speed, and pen tilt. This will enable us to fine-tune the model and improve its reliability in identifying HAVS among construction workers.

The findings highlight the potential of machine learning to revolutionize occupational health monitoring by providing a scalable, efficient, and accurate tool for

early detection of HAVS symptom. The integration of such tools can significantly improve the well-being of workers, reduce the burden of health-related risks, and enhance safety outcomes in high-risk industries. With further refinements and real-world validation, this framework can serve as a foundation for broader applications in health diagnostics and safety monitoring.

References

[1] Lingard, H., *Occupational health and safety in the construction industry*. Construction Management and Economics, 2013. **31**(6): p. 505-514.

[2] Bls, *Injuries, Illnesses, and Fatalities*, in *U.S. Bureau of labor Statistics*. 2021.

[3] Deshmukh, S.V. and S.G. Patil, *A Review Of Influence Of Hand Transmitted Vibration On Health: Due To Hand Held Power Tools.* International journal of engineering research and technology, 2012. **1**.

[4] Wasserman, D.E., et al., *Industrial vibration—An overview.* ASSE Journal, 1974: p. 38-43.

[5] Portal, P.A. *Hand-Arm Vibration Database*. 2008 [cited 2024 12/17/2024]; Available from: https://www.portaleagentifisici.it/index.php?lg=EN.

[6] Nilsson, T., *'The hand–arm vibration syndrome'—a prevention challenge or a price to pay?* Occupational Medicine, 2003/08/01. **53**(5).

[7] B, B., et al., *The NIOSH review of hand-arm vibration syndrome: vigilance is crucial. National Institute of Occupational Safety and Health - PubMed.* Journal of occupational and environmental medicine, 1998 Sep. **40**(9).

[8] Heaver, C., et al., *Hand–arm vibration syndrome: a common occupational hazard in industrialized countries.* Journal of Hand Surgery (European Volume), 2011-02-10. **36**(5).

[9] OSHA. *SawMills eTools*. [cited 2025 1/15/2025]; Available from: https://www.osha.gov/etools/sawmills/plant-wide-hazards/health-hazards/vibration.

[10] L, B., et al., *White fingers, cold environment, and vibration--exposure among Swedish construction workers - PubMed.* Scandinavian journal of work, environment & health, 2010 Nov. **36**(6).

[11] T, B., et al., *Health-care barriers for workers with HAVS in Ontario, Canada - PubMed.* Occupational medicine (Oxford, England), 2015 Mar. **65**(2).

[12] NIOSH. *Vibration Syndrome*. 1983 [cited 2024 12/17/2024]; Available from: https://www.cdc.gov/niosh/docs/83-110/default.html.

[13] CREOD. *Hand-Arm Vibration Syndrome*. A lay-language research synthesis from the Centre for Research Expertise in Occupational Disease (CREOD) 2010; Available from: https://creod.on.ca/wp-content/uploads/2023/03/CREOD-HAVS-summary.pdf.

[14] Mason, H.J., K. Poole, and J. Elms, *Upper limb disability in HAVS cases—how does it relate to the neurosensory or vascular elements of HAVS?* Occupational Medicine, 2005/08/01. **55**(5).

[15] Cederlund, R., Å. Isacsson, and G. Lundborg, *Hand function in workers with hand-arm vibration syndrome*. Journal of Hand Therapy, 1999. **12**(1).

[16] House, R., et al., *Upper extremity disability in workers with hand–arm vibration syndrome.* Occupational Medicine, 2009. **59**(3).

[17] Thompson, A.M.S., et al., *Compensation of hand-arm vibration syndrome in Canada.* Canadian Acoustics, 2011. **39**: p. 112-113.

[18] S, Y., *The compensation experience of hand-arm vibration syndrome in British Columbia - PubMed.* Occupational medicine (Oxford, England), 2012 Sep. **62**(6).

[19] Maruti, S., *Hand-arm vibration syndrome in the workplace.* BC Medical Journal, 2023. **65**(3): p. 97-99.

[20] Fisk, K., C. Nordander, and Å. Ek, *Hand–arm vibration: Swedish carpenters' perceptions of health and safety management.* Occupational Medicine (Oxford, England), 2023. **73**(2).

[21] TA, S., et al., *Hand-arm vibration syndrome among a group of construction workers in Malaysia - PubMed.* Occupational and environmental medicine, 2011 Jan. **68**(1).

[22] Edwards, D.J. and G.D. Holt, *Hand‐arm vibration exposure from construction tools: results of a field study.* Construction Management and Economics, 2006-2-1. **24**(2).

[23] Akinlolu, M. and T.C. Haupt. *Prevalence of the Hand-Arm Vibration Syndrome (HAVS) Among Construction Workers in South Africa*. 2023. Cham: Springer International Publishing.

[24] Coggins, M.A., et al., *Evaluation of Hand-Arm and Whole-Body Vibrations in Construction and Property Management.* The Annals of Occupational Hygiene, 2010. **54**(8).

[25] Rashid, K.M., V. Kumar, and S.Y. Gupta, *Automated Hand-Arm Vibration Monitoring of Construction Worker Using Smartwatch and Machine Learning*. 2020.

[26] Haines, T. and J.P. Chong, *Peripheral neurological assessment methods for workers exposed to hand-arm vibration. An appraisal.* Scandinavian journal of work, environment & health, 1987. **13 4**: p. 370-4.

[27] Yoon, J.K., et al., *Early Objectified Detection Method of Sensorineural Component in Hand Arm Vibration Syndrome.* Annals of occupational and environmental medicine, 2009. **21**: p. 143-153.
[28] Matoba, T. and T. Sakurai, *Physiological methods used in Japan for the diagnosis of suspected hand-arm vibration syndrome.* Scandinavian journal of work, environment & health, 1987. **13 4**: p. 334-6.
[29] Aw, T.C., et al., *Diagnostics of hand-arm system disorders in workers who use vibrating tools.* Occupational and Environmental Medicine, 1997. **54**: p. 90 - 95.
[30] Ye, Y., et al., *Assessment of thermotactile and vibrotactile thresholds for detecting sensorineural components of the hand–arm vibration syndrome (HAVS).* International Archives of Occupational and Environmental Health 2017 91:1, 2017-09-16. **91**(1).
[31] Poole, C.J.M., H. Mason, and A.-H. Harding, *The relationship between clinical and standardized tests for hand–arm vibration syndrome.* Occupational Medicine, 2016/06/01. **66**(4).
[32] J, D., et al., *The Purdue Pegboard Test: normative data for people aged 60 and over - PubMed.* Disability and rehabilitation, 1995 Jul. **17**(5).
[33] Rashid, Z., et al., *Hand Arm Vibration, Grip Strength Assessment and the Prevalence of Hea.* Advances in Intelligent Systems and Computing, 2018.
[34] Proud, G., et al., *Cold provocation testing and hand–arm vibration syndrome—an audit of the results of the Department of Trade and Industry scheme for the evaluation of miners.* British Journal of Surgery, 2003/08/20. **90**(9).
[35] Hill, C., et al., *Assessment of hand-arm vibration syndrome in a northern Ontario base metal mine.* Chronic Dis Can, 2001. **22**(3-4): p. 88-92.
[36] RF, M., et al., *An epidemiologic study of carpal tunnel syndrome and hand-arm vibration syndrome in relation to vibration exposure - PubMed.* The Journal of hand surgery, 1994 Jan. **19**(1).
[37] Matthies, D.J., G. Bieber, and U. Kaulbars. *AGIS: automated tool detection & hand-arm vibration estimation using an unmodified smartwatch*. in *Proceedings of the 3rd International Workshop on Sensor-based Activity Recognition and Interaction*. 2016.
[38] Altheimer, J. and J. Schneider, *Smart-watch-based construction worker activity recognition with hand-held power tools.* Automation in Construction, 2024/11/01. **167**.
[39] Zham, P., et al., *Frontiers | Distinguishing Different Stages of Parkinson's Disease Using Composite Index of Speed and Pen-Pressure of Sketching a Spiral.* Frontiers in Neurology, 2017/09/06. **8**.
[40] Shahid, A.R. and A. Imteaj, *Securing User Privacy in Cloud-Based Whiteboard Services Against Health Attribute Inference Attacks | IEEE Journals & Magazine | IEEE Xplore.* IEEE Transactions on Artificial Intelligence, 2024. **5**(8).

Deep Learning for Automated Crack Recognition: Insights, Challenges, and Future Directions

Ghodsiyeh Rostami[1], Po-Han Chen[1], Mahdi S. Hosseini[2]

[1]Department of Building, Civil, and Environmental Engineering, Concordia University, Canada
[2]Department of Computer Science and Software Engineering, Concordia University, Canada
rose.rostami@concordia.ca, pohan.chen@concordia.ca, mahdi.hosseini@concordia.ca

Abstract – Crack detection in concrete structures is a critical component of infrastructure health monitoring, as cracks pose significant risks to structural integrity. Traditional manual inspection methods, while effective, are resource-intensive, time-consuming, and prone to human error. Automated crack detection using vision-based techniques has become a focal point of research, aiming to enhance accuracy, efficiency, and consistency in inspections. Deep learning (DL) methods have revolutionized crack detection by enabling end-to-end systems that automatically learn from data and perform pixel-level segmentation. This study provides an overview of recent advances in deep learning-based crack recognition, highlighting the mainstream detection methods, and revealing ongoing challenges including the need for large and diverse datasets, the development of lightweight models for real-time performance, and model generalization across diverse conditions.

Keywords –
Crack Detection; Deep Learning; Computer Vision; Infrastructure Health Monitoring

1 Introduction

Civil infrastructure, such as buildings, bridges, and dams, develop defects over time due to material weathering, cyclic loads, environmental factors, and natural hazards. In concrete structures, cracks are the most common defect which can degrade structural integrity. Therefore, early detection of cracks is crucial to prevent catastrophic damages, making regular visual inspections critical for safety and timely maintenance planning. However, traditional manual crack inspections rely on specialized equipment, and skilled professionals, and often disrupt regular operations, making the process resource-intensive, costly, time-consuming, and prone to human error. To address these challenges, automated vision-based crack recognition methods have become a focal point of research, aiming to provide efficient, cost-effective, and consistent inspection methods.

Early efforts in vision-based crack recognition relied on digital image processing techniques, however, these techniques demonstrated limited efficacy in handling the complex and irregular geometries of cracks, as well as the varying illumination and background noise in real-world images. The introduction of machine learning (ML) provided a more adaptive approach, allowing algorithms to recognize patterns from image data and improve detection accuracy. Nevertheless, ML algorithms rely on manual feature engineering, making the process time-consuming and difficult to scale to more complex environments, which limits their broader application.

The advent of deep learning (DL), particularly Convolutional Neural Networks (CNNs), has revolutionized crack detection, enabling models to automatically learn features from data without manual intervention. CNN-based methods have shown superior performance in crack segmentation tasks, to identify cracks at the pixel level, overcoming the limitations of traditional image processing and ML algorithms. In addition, the introduction of Vision Transformers (ViTs) has enabled deep models to capture global and long-range dependencies, further enhancing the ability to detect cracks in complex backgrounds and under varying conditions. This has led to a shift in focus toward deep models capable of segmenting cracks with higher accuracy, robustness, and speed.

This study aims to provide an overview of the recent advancements in automated crack recognition, with a focus on deep learning-based methods that have revolutionized the field through their accuracy, scalability, and ability to handle complex patterns in diverse scenarios. Challenges introducing potential future directions for research and development are also explored and discussed.

2 Background

Over the past two decades, image-based crack detection has been a critical and active research area for achieving efficient and reliable automated inspections

[1,2]. Crack recognition, which involves identifying cracks in digital images or videos and quantifying their dimensions (i.e., length and width), is being extensively studied within the fields of image processing and computer vision. By definition, images are composed of pixels, which can be represented as numerical matrices, where the dimensions of the matrix are determined by the resolution of the image, and the values within the matrix represent the color or intensity of each pixel. This matrix representation enables the application of various mathematical and computational operations for tasks such as object detection and segmentation [3]. Crack detection involves highlighting pixels representing the crack, which can be analyzed at three levels: image, region, and pixel [4]. At the image level, images are classified based on whether they contain cracks (image classification) [5]. At the region level, cracks are identified and localized with bounding boxes (object detection). At the pixel level, segmentation identifies the exact boundary of cracks by highlighting pixels that represent the crack area (Figure 1) [1,6]. The goal of crack recognition is not only to detect cracks but also to measure their dimensions, where researchers explore various approaches to improve accuracy and robustness in crack segmentation. However, crack detection is challenging due to the irregular geometry and non-rigid topology of cracks, which vary in shape and size [7]. Additionally, images captured by inspection devices, such as UAVs, present diverse conditions including variations in illumination, noise, motion blur, and background complexity. For years, digital image processing techniques were explored for crack detection, however, due to their limitations remained impractical.

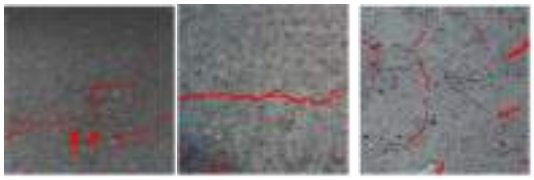


Figure 1. Examples of crack segmentation

The advent of learning-based and data-driven approaches has transformed the field, specifically Deep Learning (DL) become the mainstream approach for handling complex and high-dimensional data. The key innovation is the model's ability to automatically learn features from raw data and complete tasks without the need for manual feature engineering (Figure 2) [8]. In the field of computer vision, DNNs, especially Convolutional Neural Networks (CNNs) and Vision Transformers (ViT), have consistently outperformed traditional image processing techniques. This advancement has led to a paradigm shift in crack recognition research. The focus of many studies has been on deep models that offer faster and more generalized crack detection, where the model learns to extract features from inspection images autonomously and can identify cracks at the pixel level, a process known as semantic segmentation. Given the multi-scale nature of cracks, which can appear in different sizes within an image, and their non-rigid changing forms, several network architectures have been explored to address these challenges.

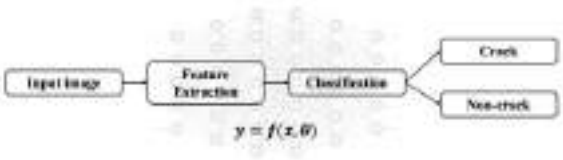


Figure 2. Representation learning by deep neural networks

3 Methodology

This review is based on a systematic content analysis of the existing literature on crack recognition. In the preliminary phase, relevant research papers were collected from databases, including Google Scholar, Scopus, ScienceDirect, and Web of Science, and using backward and forward snowballing techniques.

The keywords "concrete" AND "crack" AND ("segmentation" OR "detection"), "concrete" AND "crack" AND ("segmentation" OR "detection") AND ("measurement" OR "quantification"), and "deep" AND "crack" AND "detection" AND "quantif*" were used to search for the publications. The asterisk (*) wildcard was applied to include variations such as "quantify" and "quantification" in the search results.

During this phase, the collected papers were screened based on their abstracts, focusing on studies employing deep learning methodologies for crack detection. Only the most relevant studies were selected for an in-depth review. The analytical findings from this review were synthesized into a conceptual framework, as illustrated in Figure 3. Overview of the sections and subsections of this study . The remainder of this paper is structured according to the sections outlined in Figure 3. Overview of the sections and subsections of this study .

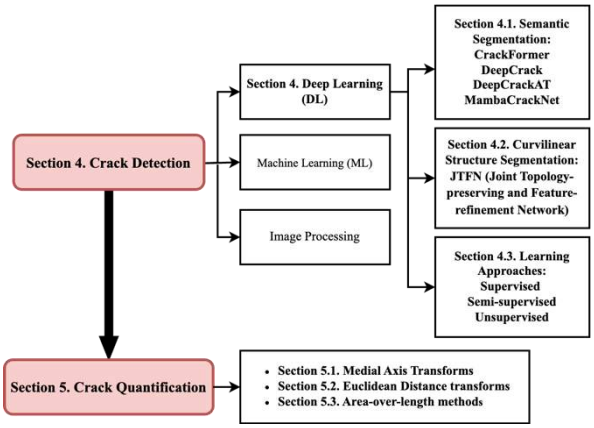


Figure 3. Overview of the sections and subsections of this study

4 Deep Learning-based Crack Detection

This section discusses various crack detection methods based on deep neural networks, with a summary of the selected studies provided in Table 1.

4.1 Semantic Segmentation

Convolutional Neural Networks (CNNs). The work by [9] introduced an early attempt at end-to-end crack segmentation using deep neural networks (DNNs), proposing a Fully Convolutional Network (FCN)-based semantic segmentation model trained on 500 concrete crack images of size 227×227 pixels. While the model outperformed Otsu's method with a 90% F1-score and precision in extracting crack paths, it failed to quantify crack dimensions due to noisy crack-like artifacts in the segmented images, a limitation stemming from the hierarchical processing of FCNs. In FCN, detailed information often from early layers is lost, which is crucial for crack detection as it requires both low-level and high-level features for fine details and semantic context, respectively. Additionally, the inductive bias of CNNs toward local feature extraction limits their ability to capture global contextual information, reducing their effectiveness in distinguishing cracks from complex backgrounds, such as material imprints.

To overcome these limitations, researchers have explored various mechanisms in designing crack segmentation models. These include dilated convolutions for enlarging the receptive field without losing resolution [10], attention mechanisms for focusing on relevant features [11−13], and Vision Transformers for capturing long-range dependencies in the data [14–19]. For instance, [20] presented a deep multi-scale crack feature learning model, named DeepCrackAT, based on an encoder-decoder architecture employing dilated convolution and attention mechanism. The model was trained on two datasets including 1440 images of masonry cracks and 2736 images of concrete and wall cracks, all resized to 224×224 pixels in size. The final model achieved 76.67% F1-score and 64.44% mIoU, nevertheless, it had limitations in segmenting cracks with significant variations in width.

Vision Transformers (ViT), introduced by [21] in 2020, employ self-attention mechanisms to capture global patterns and long-range dependencies. Demonstrating strong generalization capabilities, researchers have leveraged ViTs to enhance model robustness in new scenes and to improve crack detection in complex backgrounds [14]. For instance, [22] addressed the challenge of detecting fine-grained cracks by proposing a transformer-based segmentation network, CrackFormer. The model adapts SegNet architecture, introducing novel self-attention modules with 1×1 convolutional kernels to capture efficient contextual information across features channels, and efficient positional embeddings. New scaling-attention modules are also proposed to suppress non-semantic features and sharpen semantics when combining encoder-decoder outputs. Their comparative results demonstrated CrackFormer's superiority in multi-scale analysis of pavement crack images with higher performance and fewer floating-point operations (FLOPs).

Vision transformers, however, are dependent on large data and are computationally expensive. Therefore, researchers have proposed more efficient hybrid architectures combining convolutions and transformer operations. For instance, [18] proposed a dual-encoder network fusing Transformers and CNNs to capture local and global context information in crack inspection images with complex backgrounds. The model was trained on the Crack3238 dataset and achieved 76.44% F1-score and 78 FPS, which was higher than that of other networks in the comparative analysis.

Vision Mamba is a state-of-the-art high-performance computational framework that incorporates bidirectional state-space models for data-dependent global visual context modeling and position embeddings for location-aware visual recognition [23]. It is designed for efficient and precise data processing, offering scalability and flexibility for diverse applications. In a recent study, [24] proposed a hybrid architecture, MambaCrackNet, combining Mamba and CNN approaches to improve fine crack segmentation accuracy while reducing misdetections of background stains. Comparative experiments on two crack datasets demonstrated MambaCrackNet's superior performance over conventional segmentation models. Nevertheless, the model has a large number of learnable parameters due to the addition of two feature processing paths, leading to increased model complexity and slower inference speed.

Additionally, the Mamba module imposes restrictions related to input image size and the handling of extra spatial dimensions, which can limit the model's flexibility and adaptability in crack detection applications.

4.2 Curvilinear Structure Segmentation

Another approach in crack detection is curvilinear structure segmentation. These methods focus on identifying elongated, narrow structures such as cracks, blood vessels, roads, or rivers in images [25]. Preserving continuity, connectivity, and shape while minimizing false detections are the emphasis of these methods. Early approaches relied on edge detection and hand-crafted features, while modern techniques leverage deep learning models including CNNs and U-Nets, incorporating multi-scale feature fusion and global context modeling. Despite advances, challenges such as fragmentation, noise sensitivity, and scalability remain [26,27].

4.3 Learning Approaches

The studies on crack detection can also be categorized based on their learning approaches, including supervised, semi-supervised, and unsupervised learning methods. Supervised learning approaches have been widely studied in the literature. However, one of the challenges in crack detection is the lack of large-scale annotated datasets, hindering the full employment of deep learning models to achieve robust and generalized models. The most large-scale crack detection study is [28], which recently introduced OMNICRACK30K, a comprehensive collection of decentralized crack segmentation datasets, aimed at establishing a sustainable benchmark. Therefore, to address the challenge of large-scale labeled datasets, some studies investigated semi-supervised learning methods to leverage both small-scaled annotated data and largely available unlabelled datasets. [29] introduced a semi-supervised framework for crack segmentation on concrete and asphalt surfaces. The framework involves three stages: 1) supervised learning with different fractions of labeled data (2%, 10%, 14%, 20%, and 25%) using transfer learning, 2) consistency regularization for stability between original and perturbed images, and 3) self-training with pseudo-labeling and retraining the model from scratch on the expanded dataset for two times. According to the results, despite using only 2% of the data as labeled data, the model's performance (mIoU) was only 2.6% to 4.7% lower than using 100% of the labeled data, showing the effectiveness of the semi-supervised approach with limited annotations.

Furthermore, a novel representation learning framework was introduced by [30]. Addressing the ambiguity of marginal non-crack regions in crack segmentation, this study proposed a clustering-inspired representation learning framework (CIRL) consisting of two phases, where in phase 1, marginal non-crack regions are localized by introducing an Ambiguity-aware segmentation loss (Aseg Loss) to learn prediction variances. Next, to prevent the disturbance of marginal non-crack during model training, in the second phase, a Clustering-inspired loss (CI Loss) was proposed for ambiguous pixels based on the idea that features that are located close/farther in feature space should have consistent/inconsistent prediction. Their comparative analysis of their CIRL with other state-of-the-art crack segmentation models and loss functions showed the superiority of CIRL on four different datasets.

Table 1 Summary of the selected studies on crack detection

Category	Ref.	Model	Approach	Dataset	Performance	Advantages	Limitations
CNN-Based Approaches	[9]	FCN-based semantic segmentation	Supervised	500 concrete crack images (227×227)	90% F1-score	Early work on end-to-end crack segmentation, Outperforms Otsu's method	Fails to quantify crack dimensions, loss of crack details
	[20]	DeepCrackAT (Encoder-Decoder + Dilated Conv + Attention)	Supervised	1440 masonry, 2736 concrete/wall cracks (224×224)	76.67% F1-score, 64.44% mIoU	Captures multi-scale features	Struggles with variable crack widths

Vision Transformer-Based	[22]	CrackFormer (Transformer + SegNet)	Supervised	Two pavement crack datasets (CrackTree260, CrackLS315), one stone crack (Stone331)	Higher accuracy with lower FLOPs (88% average precision)	Multi-scale analysis, effective contextual learning	Computationally expensive
Hybrid Approaches	[18]	DTrC-Net with a Dual-encoder (Transformer + CNN)	Supervised	Crack3238	76.44% F1-score, 78 FPS	Balances local and global feature extraction	High computational cost
	[24]	MambaCrack Net (Mamba + CNN)	Supervised	CCSD (458 high-resolution concrete crack), CD (776 pavements/building walls cracks)	Higher performance than conventional models	Improved fine crack detection	High parameter count, slower inference
Curvilinear Structure Segmentation	[26]	JTFN (Joint Topology-preserving and Feature-refinement Network)	Supervised	concrete/pavement cracks (Crack500, CrackTree260,), 40 retina images (DRIVE), 1160 Road cracks	84% F1-score	Preserves connectivity and shape	Fragmentation, noise sensitivity
Learning Approaches	[29]	U-Net	Semi-supervised framework (transfer learning + pseudo-labeling)	Concrete & asphalt datasets (2%-25% labeled)	mIoU only 2.6%-4.7% lower than fully supervised	Effective with limited labels	Dependence on pseudo-labeling quality
	[30]	JTFN + CIRL	CIRL (Clustering-inspired representation learning)	Concrete/pavement cracks (CrackTree260, Crack500, CrackSeg5K), retina images (DRIVE)	Outperforms state-of-the-art methods on ambiguous crack pixels	Reduces marginal non-crack disturbances	Computationally intensive due to the clustering algorithm

5 Crack Analysis

As the literature review revealed, most studies have focused on enhancing crack detection models, while only a few have addressed methods for crack analysis (i.e., width measurement). This aspect is challenging, as high-quality crack segmentation is crucial for reliable and accurate measurement, especially considering variations in crack inspection images, such as scene complexity, background noise, and texture. Various methods have been proposed for crack quantification, including skeleton extraction techniques such as thinning algorithms, medial axis transforms, and distance transforms [31,32], as well as area-over-length methods [33] and edge detection approaches [34]. An overview of these methods is provided in Table 2. Relevant studies with respect to each approach are further discussed in the following sections.

Table 2 Overview of the crack quantification methods

Measurement Method	Relevant Studies	Advantages	Limitations
Medial Axis Transform (MAT)	[35]	- Provides a skeletonized representation of cracks, preserving	- Sensitive to noise, leading to fragmented or broken skeletons.

		topology. - Applicable to thin, continuous cracks.	- Over-simplifies complex crack structures, losing fine details. - Computationally intensive, especially when combined with deep learning.
Euclidean Distance	[36,37]	- Directly measures crack width by computing the shortest distance between two edges. - Works with both 2D images and 3D point clouds.	- Highly dependent on segmentation quality. Errors in crack edges lead to inaccurate width estimation. - Fails for irregular or branching cracks, as it assumes relatively parallel edges. - May overestimate or underestimate width if cracks have varying thicknesses.
Area-over-Length	[33,38]	- Robust against noise since it does not rely on edge detection. - Provides a simple, efficient way to estimate crack width. - Applicable to thin and elongated cracks.	- Inaccurate for irregular crack shapes, as it assumes uniform width. - Affected by small gaps or interruptions in cracks, leading to overestimation. - Struggles with very thick or non-linear cracks, where area may not properly reflect actual width.

5.1 Medial Axis Transform

To enhance microcrack detection in blurry and low-resolution images taken by UAVs, [35] proposed a three-stage crack recognition method by Super-Resolution Reconstruction (SRR), semantic segmentation, and an improved Medial Axis Transform (MAT) for crack width and length quantification. Their study showed the effectiveness of super-resolution reconstruction where reconstructed images from degraded images lead to similar segmentation output as with original high-resolution images. Nevertheless, the method is a two-stage process which cannot fulfill the real-time demand in practice.

5.2 Euclidean Distance

[36] proposed detection and quantification of global concrete cracks in panorama images using a three-stage process, including image classification of cropped panorama images, crack segmentation of positive outputs from the classification stage and crack quantization. For crack measurement, they used the coordinates of pixels to measure the Euclidean distance between two edges of cracks at different heights (lengths). To overcome the challenge of crack-like artifacts in the segmented images, they quantified cracks with a height of 50 pixels and above only. They calculated the measurements in real-world sizes by multiplying actual pixel sizes by the measurements in pixel value. The relative error of crack width measurements was 3.87%. Furthermore, [37] proposed a point cloud-based approach for analyzing cracks and spalls using semantic segmentation of raw point cloud data. The study developed a Normal Vector Enhanced Dynamic Graph CNN (NVE-DGCNN) to enable the direct analysis of the raw point cloud dataset without further conversions. For each identified defect, a minimum bounding box was generated, allowing for precise calculation of defect dimensions, including length, width, and depth. These measurements were determined through Euclidean distance calculations between the bounding box corners. However, achieving real-time crack recognition using point cloud data remains an open research challenge.

5.3 Area-over-Length

[38] proposed a two-stage crack detection method using a CNN classifier (pre-trained SqueezeNet), Deeplabv3+ to first classify crack and non-crack images, then segment cracks in positive images, and crack quantification using the regionprops function in MATLAB. The function calculates the width of each ellipse in the image by dividing its area over the length. Their results showed a lower average error of regionprops function in measuring crack width and length (7 and 2%, respectively) compared to a skeleton-based method such as medial-axis transform (11 and 17%). To leverage a trade-off between Transformers accuracy and computational cost, [33] developed a lightweight crack segmentation model using CNNs and Transformers named CrackTrNet to enable intelligent concrete dam crack patron inspection. Crack's length was

obtained using a thinning algorithm and width was measured by dividing crack's area over the length. According to the results, the model achieved similar performance compared to mainstream segmentation models with lower FLOPs and fewer parameters. 89.10% mIoU was achieved with 6.11 million parameters, which shows the feasibility of high segmentation performance with lightweight models. Yet, the study has limitations due to the non-diverse dataset, which consists of close-up images with relatively simple backgrounds and noise. Addressing these issues could enhance the model's generalization and robustness.

6 Discussion

The findings indicate that for crack detection, while ViT-based models achieve superior accuracy in controlled settings, they require large-scale datasets and substantial computational resources. For real-time applications, hybrid architectures (CNN-Transformers) offer a more effective balance between accuracy and efficiency, making them promising candidates for deployment. Additionally, optimization techniques such as pruning can enhance memory efficiency and reduce computational costs, further improving model feasibility in practical scenarios.

Moreover, accurate crack width measurement has not been achieved. This is due to limitations of the existing measurement algorithms in this task, as they are not specifically designed for crack measurement. For instance, techniques such as the medial axis transform, as discussed by [39], often generate redundant branches due to the irregular and non-rigid shape of cracks, which in turn distorts measurement accuracy. Furthermore, current crack analysis methods rely on perfectly accurate binary segmentation masks, which have not yet been achieved with the current crack detectors proposed in the literature. Imperfect masks, which may include noise, discontinuities in crack structures, or impurities that resemble cracks, affect the precision of the measurement process, making this an area in need of significant improvement.

7 Challenges, Opportunities, and Future Directions

Fully autonomous crack recognition remains an unsolved issue, as accurate width measurement is dependent on high-quality segmentation, specifically in complex and noisy environments. Therefore, the development of robust crack detection models is essential. One of the main challenges is the lack of large-scale and diverse datasets that can represent the wide range of real-world conditions in which cracks appear [40]. Additionally, the variability in crack structures makes it challenging to design models that can effectively capture the multi-scale and non-rigid structure of cracks. Future research can focus on developing advanced segmentation models and effective operations to enhance detection accuracy and improve generalization across diverse conditions.

Furthermore, in real-world applications, processing speed and computational efficiency are of paramount importance for real-time crack detection during infrastructure inspections or on-site assessments. One challenge is optimizing the trade-off between inference speed, accuracy, and computational cost of the detection models. Moreover, despite advancements in semantic segmentation, multi-stage measurement processes often impact real-time applicability [41]. Developing more robust, noise-resistant quantification methods while improving real-time efficiency can support large-scale, automated infrastructure assessment.

On the other hand, the literature has mainly focused on single-defect detection through binary segmentation whereas, in real-world scenarios, multiple defects (e.g., rust, cracks, and demolition) often coexist, requiring detailed assessments (i.e., area of the damaged surface structure). While some methods have been proposed [42,43], mainly for rough multi-defect localization using object detection and bounding boxes [44,45], future research can shift towards multi-defect segmentation to enable more precise and comprehensive assessment. A major challenge in this direction is the lack of multi-class annotated datasets for multi-defect segmentation. Addressing this limitation through the development of diverse, high-quality datasets and the adoption of multi-task learning frameworks could enhance the versatility of automated inspection systems.

Finally, from a technical standpoint, researchers utilize programming languages such as Python and MATLAB to address these challenges, along with deep learning frameworks including PyTorch or TensorFlow for model development [1]. Additionally, computer vision libraries such as OpenCV and scikit-image facilitate image processing, and automated annotation tools such as LabelMe or CVAT assist in preparing labeled datasets. For practical deployment, optimized frameworks such as TensorFlow Lite and ONNX Runtime enable efficient inference on edge devices. The advancement of these tools allows for the development of more accurate and efficient crack detection solutions.

8 Conclusion

The advancement of deep learning and computer vision technologies has significantly transformed automated crack detection, enabling more precise, and efficient visual crack assessment. This review provided an analysis of the state-of-the-art methodologies in crack

detection and quantification. Key developments, existing challenges, and potential future research directions were provided.

Through a systematic literature review, studies have been analyzed based on their computer vision task, learning approach, and dataset scale, and advantages and limitations were discussed. Furthermore, various crack measurement methods proposed in the literature were reviewed highlighting their effectiveness and shortcomings.

The review revealed that despite the accuracy of ViT-based models in controlled environments, their reliance on large-scale datasets and high computational power limits their practical deployment. Hybrid architectures, such as CNN-Transformers, offer an enhanced balance between accuracy and efficiency, especially when combined with optimization techniques such as pruning to reduce computational overhead. However, accurate crack width measurement remains an unresolved issue due to the limitations of existing algorithms and the reliance on imperfect segmentation masks. Addressing these challenges is crucial for advancing real-world applications.

References

[1] Ai D, Jiang G, Lam S-K, He P, Li C. Computer vision framework for crack detection of civil infrastructure—A review. Engineering Applications of Artificial Intelligence 2023;117:105478. https://doi.org/10.1016/j.engappai.2022.105478.

[2] Halder S, Afsari K. Robots in Inspection and Monitoring of Buildings and Infrastructure: A Systematic Review. Applied Sciences 2023;13:2304. https://doi.org/10.3390/app13042304.

[3] Gonzalez R, Woods R. Digital Image Processing 2017. https://elibrary.pearson.de/book/99.150005/9781292223070 (accessed October 10, 2024).

[4] Szeliski R. Computer Vision: Algorithms and Applications. Springer Nature; 2022.

[5] Prince SJD. Computer Vision: Models, Learning, and Inference. Cambridge University Press; 2012.

[6] Forsyth DA, Ponce J. Computer Vision: A Modern Approach. Prentice Hall Professional Technical Reference; 2002.

[7] Munawar HS, Hammad AWA, Haddad A, Soares CAP, Waller ST. Image-Based Crack Detection Methods: A Review. Infrastructures 2021;6:115. https://doi.org/10.3390/infrastructures6080115.

[8] Ian Goodfellow, Yoshua Bengio, Aaron Courville. Deep Learning. MIT Press; 2016.

[9] Dung CV, Anh LD. Autonomous concrete crack detection using deep fully convolutional neural network. Automation in Construction 2019;99:52–8. https://doi.org/10.1016/j.autcon.2018.11.028.

[10] Ren Y, Huang J, Hong Z, Lu W, Yin J, Zou L, et al. Image-based concrete crack detection in tunnels using deep fully convolutional networks. Construction and Building Materials 2020;234:117367. https://doi.org/10.1016/j.conbuildmat.2019.117367.

[11] Deng L, Yuan H, Long L, Chun P, Chen W, Chu H. Cascade refinement extraction network with active boundary loss for segmentation of concrete cracks from high-resolution images. Automation in Construction 2024;162:105410. https://doi.org/10.1016/j.autcon.2024.105410.

[12] Liang J, Gu X, Jiang D, Zhang Q. CNN-based network with multi-scale context feature and attention mechanism for automatic pavement crack segmentation. Automation in Construction 2024;164:105482. https://doi.org/10.1016/j.autcon.2024.105482.

[13] Pan Y, Zhang G, Zhang L. A spatial-channel hierarchical deep learning network for pixel-level automated crack detection. Automation in Construction 2020;119:103357. https://doi.org/10.1016/j.autcon.2020.103357.

[14] Asadi Shamsabadi E, Xu C, Rao AS, Nguyen T, Ngo T, Dias-da-Costa D. Vision transformer-based autonomous crack detection on asphalt and concrete surfaces. Automation in Construction 2022;140:104316. https://doi.org/10.1016/j.autcon.2022.104316.

[15] Chen Z, Asadi Shamsabadi E, Jiang S, Shen L, Dias-da-Costa D. An average pooling designed Transformer for robust crack segmentation. Automation in Construction 2024;162:105367. https://doi.org/10.1016/j.autcon.2024.105367.

[16] Wang W, Su C. Automatic concrete crack segmentation model based on transformer. Automation in Construction 2022;139:104275. https://doi.org/10.1016/j.autcon.2022.104275.

[17] Wu Y, Li S, Zhang J, Li Y, Li Y, Zhang Y. Dual attention transformer network for pixel-level concrete crack segmentation considering camera placement. Automation in Construction 2024;157:105166. https://doi.org/10.1016/j.autcon.2023.105166.

[18] Xiang C, Guo J, Cao R, Deng L. A crack-segmentation algorithm fusing transformers and convolutional neural networks for complex detection scenarios. Automation in Construction 2023;152:104894. https://doi.org/10.1016/j.autcon.2023.104894.

[19] Zhang E, Shao L, Wang Y. Unifying transformer and convolution for dam crack detection. Automation in Construction 2023;147:104712. https://doi.org/10.1016/j.autcon.2022.104712.

[20] Lin Q, Li W, Zheng X, Fan H, Li Z. DeepCrackAT: An effective crack segmentation framework based on learning multi-scale crack features. Engineering Applications of Artificial Intelligence 2023;126:106876. https://doi.org/10.1016/j.engappai.2023.106876.

[21] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale 2021. https://doi.org/10.48550/arXiv.2010.11929.

[22] Liu H, Miao X, Mertz C, Xu C, Kong H. CrackFormer: Transformer Network for Fine-Grained Crack Detection. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021, p. 3763–72. https://doi.org/10.1109/ICCV48922.2021.00376.

[23] Zhu L, Liao B, Zhang Q, Wang X, Liu W, Wang X. Vision Mamba: Efficient Visual Representation Learning with Bidirectional State Space Model 2024. https://doi.org/10.48550/arXiv.2401.09417.

[24] Han C, Yang H, Yang Y. Enhancing pixel-level crack segmentation with visual mamba and convolutional networks. Automation in Construction 2024;168:105770. https://doi.org/10.1016/j.autcon.2024.105770.

[25] Hu X, Li F, Samaras D, Chen C. Topology-Preserving Deep Image Segmentation. Advances in Neural Information Processing Systems, vol. 32, Curran Associates, Inc.; 2019.

[26] Cheng M, Zhao K, Guo X, Xu Y, Guo J. Joint Topology-preserving and Feature-refinement Network for Curvilinear Structure Segmentation. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada: IEEE; 2021, p. 7127–36. https://doi.org/10.1109/ICCV48922.2021.00706.

[27] Mosinska A, Marquez-Neila P, Kozinski M, Fua P. Beyond the Pixel-Wise Loss for Topology-Aware Delineation. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, p. 3136–45. https://doi.org/10.1109/CVPR.2018.00331.

[28] Benz C, Rodehorst V. Omni-Crack30k: A Benchmark for Crack Segmentation and the Reasonable Effectiveness of Transfer Learning. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA: IEEE; 2024, p. 3876–86. https://doi.org/10.1109/CVPRW63382.2024.00392.

[29] Asadi Shamsabadi E, Erfani SMH, Xu C, Dias-da-Costa D. Efficient semi-supervised surface crack segmentation with small datasets based on consistency regularisation and pseudo-labelling. Automation in Construction 2024;158:105181. https://doi.org/10.1016/j.autcon.2023.105181.

[30] Chen Z, Lai Z, Chen J, Li J. Mind marginal non-crack regions: Clustering-inspired representation learning for crack segmentation. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA: IEEE; 2024, p. 12698–708. https://doi.org/10.1109/CVPR52733.2024.01207.

[31] Bae H, An Y-K. Computer vision-based statistical crack quantification for concrete structures. Measurement 2023;211:112632. https://doi.org/10.1016/j.measurement.2023.112632.

[32] Lee JS, Hwang SH, Choi IY, Choi Y. Estimation of crack width based on shape-sensitive kernels and semantic segmentation. Structural Control and Health Monitoring 2020;27:e2504. https://doi.org/10.1002/stc.2504.

[33] Li M, Yuan J, Ren Q, Luo Q, Fu J, Li Z. CNN-Transformer hybrid network for concrete dam crack patrol inspection. Automation in Construction 2024;163:105440. https://doi.org/10.1016/j.autcon.2024.105440.

[34] Ding W, Yang H, Yu K, Shu J. Crack detection and quantification for concrete structures using UAV and transformer. Automation in Construction 2023;152:104929. https://doi.org/10.1016/j.autcon.2023.104929.

[35] Xiang C, Wang W, Deng L, Shi P, Kong X. Crack detection algorithm for concrete structures based on super-resolution reconstruction and segmentation network. Automation in Construction 2022;140:104346. https://doi.org/10.1016/j.autcon.2022.104346.

[36] Song L, Sun H, Liu J, Yu Z, Cui C. Automatic segmentation and quantification of global cracks in concrete structures based on deep learning. Measurement 2022;199:111550. https://doi.org/10.1016/j.measurement.2022.111550.

[37] Bahreini F, Hammad A. Dynamic graph CNN based semantic segmentation of concrete defects and as-inspected modeling. Automation in Construction 2024;159:105282. https://doi.org/10.1016/j.autcon.2024.105282.

[38] Teng S, Chen G. Deep Convolution Neural Network-Based Crack Feature Extraction, Detection and Quantification. J Fail Anal and Preven 2022;22:1308–21. https://doi.org/10.1007/s11668-022-01430-9.

[39] Yu Z, Shen Y, Sun Z, Chen J, Gang W. Cracklab: A high-precision and efficient concrete crack segmentation and quantification network. Developments in the Built Environment 2022;12:100088. https://doi.org/10.1016/j.dibe.2022.100088.

[40] Deng W, Mou Y, Kashiwa T, Escalera S, Nagai K, Nakayama K, et al. Vision based pixel-level bridge structural damage detection using a link ASPP network. Automation in Construction 2020;110:102973. https://doi.org/10.1016/j.autcon.2019.102973.

[41] Miao P, Srimahachota T. Cost-effective system for detection and quantification of concrete surface cracks by combination of convolutional neural network and image processing techniques. Construction and Building Materials 2021;293:123549. https://doi.org/10.1016/j.conbuildmat.2021.123549.

[42] Al Deen Taher SS, Dang J. Autonomous Multiple Damage Detection and Segmentation in Structures Using Mask R-CNN. In: Wu Z, Nagayama T, Dang J, Astroza R, editors. Experimental Vibration Analysis for Civil Engineering Structures, Cham: Springer International Publishing; 2023, p. 545–56. https://doi.org/10.1007/978-3-030-93236-7_45.

[43] Hoskere V, Narazaki Y, Hoang TA, Spencer Jr. BF. MaDnet: multi-task semantic segmentation of multiple types of structural materials and damage in images of civil infrastructure. J Civil Struct Health Monit 2020;10:757–73. https://doi.org/10.1007/s13349-020-00409-0.

[44] Cha Y-J, Choi W, Suh G, Mahmoudkhani S, Büyüköztürk O. Autonomous Structural Visual Inspection Using Region-Based Deep Learning for Detecting Multiple Damage Types. Computer-Aided Civil and Infrastructure Engineering 2018;33:731–47. https://doi.org/10.1111/mice.12334.

[45] Jiang Y, Pang D, Li C. A deep learning approach for fast detection and classification of concrete damage. Automation in Construction 2021;128:103785. https://doi.org/10.1016/j.autcon.2021.103785.

Unified Network-Based Representation of BIM Models for Embedding Semantic, Spatial, and Topological Data

Jin Han[1], Xin-Zheng Lu[1], and Jia-Rui Lin[1,*]

[1]Department of Civil Engineering, Tsinghua University, China
han-j23@mails.tsinghua.edu.cn, luxz@tsinghua.edu.cn, lin611@tsinghua.edu.cn

Abstract –

Building Information Modeling (BIM) has revolutionized the construction industry by providing a comprehensive digital representation of building structures throughout their lifecycle. However, existing research lacks effective methods for capturing the complex spatial and topological relationships between components in BIM models, which are essential for understanding design patterns and enhancing decision-making. This study proposes a unified network-based representation method that integrates the "semantic-spatial-topological" multi-dimensional design features of BIM models. By extending the IFC (Industry Foundation Classes) standard, we introduce local spatial relationships and topological connections between components to enrich the network structure. This representation method enables a more detailed understanding of component interactions, dependencies, and implicit design patterns, effectively capturing the semantic, topological, and spatial relationships in BIM, and holds significant potential for the representation and learning of design patterns.

Keywords –

BIM representation; Network; Multi-dimensional; Design features

1 Introduction

Building Information Modeling (BIM) is a digital methodology employed in construction to design, manage, and improve detailed 3D models of buildings across their entire lifecycle [1]. By consolidating detailed information about a structure's physical and functional properties into one unified model, BIM has greatly enhanced the efficiency and effectiveness of engineering applications [2]. Consequently, BIM has been widely studied in various fields, including visualization [3], human-machine collaborative design [4] [5], and life cycle assessment [6] [7], and its adoption is steadily increasing. However, the inherent complexity of BIM models, which contain vast amounts of semantic, spatial, and topological information, makes it challenging to leverage their full potential for automated analysis and decision-making [8]. Traditional methods of working with BIM models often struggle to capture the nuanced relationships between components, such as spatial proximities, functional dependencies, and topological connections, hindering advanced tasks like design validation, conflict detection, and knowledge extraction. Therefore, how to extract and represent relevant design features and patterns from BIM data to enhance the design, construction, and maintenance of buildings has become a key focus of scholarly research.

To address the demand for a standardized and interoperable framework in construction, the Industry Foundation Classes (IFC) were introduced as an open data format to facilitate the exchange of BIM information across various tools [9] [10]. However, IFC's complex structure and irrelevant data for many tasks have led researchers to represent BIM models as vectors or networks [11]. Some studies have focused on extracting key component attributes, converting them into tensors for machine learning applications, such as clash detection and classification [12]-[14]. While these approaches address component attributes, they overlook important spatial and topological relationships, which are crucial for understanding component interactions. To capture these relationships, other researchers have modeled BIM components as nodes and edges in a network, representing attributes as node or edge features [15][17]. While these studies address some limitations of the first method, the spatial and topological relationships they capture are often simplistic and insufficient. Furthermore, the graph construction process is typically tailored to specific tasks, lacking flexibility and general applicability.

Therefore, this study proposes a unified network-based representation method for the BIM model's "semantic-spatial-topological" multi-dimensional design features. Building upon IFC and previous research, it further incorporates local spatial relationships and topological relationships between components to enrich the local information within the network. This approach not only facilitates the integration of different data types

but also enables the identification of implicit design patterns, thereby enhancing the understanding of component interactions and dependencies.

2 Background

To address the growing demand for a standardized and interoperable framework in the construction industry, IFC were introduced [9]. IFC was developed as an open, neutral data format to ensure the exchange of information across various BIM tools without data loss or misinterpretation [10]. However, due to the non-intuitive nature of the data in IFC and the presence of irrelevant information for most tasks, many researchers have attempted to further represent BIM models as vectors, or networks based on IFC and BIM software.

The first type primarily involves extracting key attributes of components and converting them into independent tensors for training and prediction. Lin & Huang (2019) extracted data from clash detection reports, encoded it into tensors, and used machine learning to detect unrelated conflicts [12]. Wang & Leite (2013) selected three types of attributes—geometric properties (e.g., "Volume"), usage and materials (e.g., "Material Flexibility"), and clash information (e.g., "Clashing Volume"), which were converted into tensors and used to train machine learning algorithms for predicting clash responses [13]. Liu et al. (2024) used component images and basic attributes like thickness and concrete grade to train a multi-model deep learning system for classification [14]. Although the above studies considered key component attributes, their use of tensors overlooked important spatial and topological relationships, such as relative positions, connections, and collaborative functions (e.g., "force transfer mechanisms"). These relationships are not only key to understanding component interactions and functional coordination but also crucial in capturing design patterns. For instance, the spatial relationship between beams and columns reflects how they transfer forces together. Without this information, models cannot fully capture the implicit interactions between components.

Therefore, scholars have attempted to represent BIM models as networks, where components and their relationships are depicted as nodes and edges, and attributes are represented as node or edge features. For example, Wang et al. (2022) treat point-based instances (e.g., “Pipe Fitting”) and curve-based instances (e.g., “Pipe Curve”) as nodes and edges in a network, using graph matching techniques to detect connection errors between components [15]. Hu et al. (2020; 2023) represent spatial and topological relationships, such as clashes and connections, as edges for collision detection and optimization [5] [16]. In contrast to the fine-grained representation, Li et al. (2024) treat rooms as nodes, with edges representing their hierarchical relationships, enabling automatic modular building design [17]. While these methods effectively capture certain spatial and topological properties of BIM models, they show considerable variation in network construction depending on the task, lacking a standardized approach to BIM model representation. Moreover, the absence of spatial relationships between components leads to incomplete information in the network.

3 Unified Network-based Representation Method for BIM

While the IFC data format can describe the geometry, attributes, and relationships of building components in detail, its storage format involves complex hierarchical structures and a vast amount of semantic information. However, deep learning networks, which are commonly used for data analysis, require numerical, flattened formats (such as images, matrices, or vectors) as input. Therefore, in order to leverage the data within existing BIM models for analysis, learning, and reuse, it is necessary to extract and preprocess the data from IFC to simplify its structure and extract key features.

In this study, we propose a unified network-based representation method for BIM models that retains critical information while structuring the data to accommodate various BIM models and research questions. Specifically, using IFC files exported from BIM software, this study extracts and calculates the properties of various components and their spatial and topological relationships, as illustrated in Table 1. Subsequently, all components are then represented as different types of nodes based on their classification in the IFC, with the extracted semantic properties represented as node features. Meanwhile, the spatial and topological relationships between components are represented as edges connecting the corresponding nodes. This network-based representation method lays the foundation for subsequent learning and transfer of design patterns within BIM models.

3.1 Semantic Data Representation of Components

Specifically, in representing the semantic properties of components, this study selects only the attributes related to each component itself from the IFC files exported by the BIM model, including basic information, geometric shapes, material properties, and other relevant attributes. The basic information includes attributes such as component ID, name, and FamilyName, which are used to uniquely identify the component. Geometric shapes are described by the attribute "Representation," including BoundingBox, Brep, SectionedSpine, and

others. Material properties are stored in entities such as "IfcMaterial" and "IfcMaterialLayerSetUsage," describing the composition and characteristics of the component. Other relevant attributes are defined through "IfcPropertySet", including durability, thermal conductivity, compressive strength, and more.

3.2 Topological Data Representation of Components

For topological properties, we selected the host and connection relationships contained in the IFC. In IFC, entities such as doors (IfcDoor) and windows (IfcWindow) are typically associated with wall entities (IfcWallStandardCase) through an opening entity (IfcOpeningElement). However, in this study, the opening entity is ignored, and the host relationship is directly associated with the wall and window or door entities. Besides, connection relationships are obtained through "IFCRELCONNECTSPATHELEMENTS," which describe the connections between components.

In addition, this study introduces a touch floor relationship to describe the vertical connection between components and floors, providing a more intuitive representation of the functional layout and component distribution across different building levels. For example, this includes the connection between walls and floors, or the installation positions of equipment within a floor relative to the floor slab. This topological relationship is extracted by calculating the positional relationship between the bounding boxes of components and floors. The three types of topological relationships extracted and calculated are shown in Figure 1.

Table 1 Unified attributes extraction and calculation method for semantic, topological, and spatial properties of components

Properties	Attributes	Description
Semantic	Basic information	Including attributes such as component ID, name, and FamilyName, which are used to identify the component uniquely.
	Geometric shapes	Defined by the attribute "Representation" from IFC, including BoundingBox, Brep, and others, this is used to represent the geometric shape of components.
	Material properties	Defined by attributes such as "IfcMaterial" and "IfcMaterialLayerSetUsage" in IFC, describing the composition and material of the component.
	Other relevant attributes	Defined by "IfcPropertySet", including durability, thermal conductivity, compressive strength, and more.
Spatial	Location	Coordinates of component positioning points or positioning line endpoints. It is defined by "ObjectPlacement," including the local coordinate system and reference coordinate system.
	Spatial relationships	The positional relationship between two components in space, including five categories: 1) different surface, 2) interface non-parallel, 3) interface parallel, 4) point-to-line, and 5) point-to-point.
	Angle	The intersection angle between two components' positioning lines; if one component uses point positioning, the angle is 0.
	Shortest distance vector coordinates	Endpoint coordinates of the shortest distance vector between the positioning points or lines of two components.
	Shortest distance	Actual spatial separation between two components. A negative value indicates an overlap or collision between the components, while a positive value denotes a gap.
	Angle with the horizontal plane	The angle between the plane or line formed by the two components and the horizontal plane. For spatial relationships in categories 2-5, the angle with the horizontal plane represents the angle between the plane or line formed by their positioning lines or points and the horizontal plane. In category 1, the angle is 0.
Topological	Host relationship	Represents the inclusion relationship, where one component is fully contained within another, such as doors/windows within walls.
	Connection relationship	Defined by "IFCRELCONNECTSPATHELEMENTS," which describe the connections between components.
	Touch floor	Describe the vertical connection between components and floors.

relationship

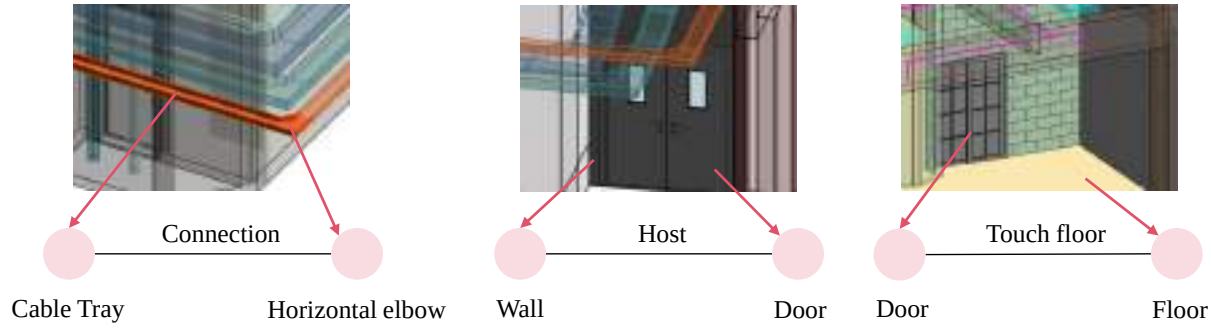


Figure 1. Three types of topological relationships

3.3 Spatial Data Representation of Components

This study first extracts the attributes representing the location of components from the IFC files, which are described by "ObjectPlacement," including the local coordinate system and reference coordinate system.

However, simple coordinate information alone is insufficient to intuitively reflect the spatial relationships between components. To address this limitation, this study further incorporates local spatial relationships of nearby components to provide additional geometric information and contextual understanding. Specifically, a local spatial relationship identifies all components within a specified distance (0.5m in this study) from a given component and calculates their relative positional relationships, including: 1) spatial relationship, 2) angle, 3) shortest distance vector coordinates, 4) shortest distance, and 5) angle with the horizontal plane, as illustrated in Figure 2.

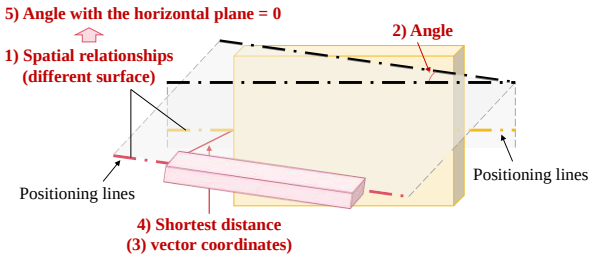


Figure 2. Example of local spatial relationship

The "spatial relationship" refers to the positional relationship between two components in space, which is determined by their positioning points or lines, and it can be classified into five distinct types: 1) different surfaces, 2) non-parallel interfaces, 3) parallel interfaces, 4) point-to-line, and 5) point-to-point. The first three types apply when both components are positioned using lines, the fourth type applies when one component is positioned by points and the other by lines, and the fifth applies when both components are positioned by points. The "angle" refers to the intersection angle between the positioning lines of the two components. If one component uses point positioning, the angle is considered 0°. The "shortest distance vector coordinates" correspond to the endpoints of the shortest distance between the positioning points or lines of the two components. The "distance" value indicates the spatial separation between the components. A negative distance suggests overlap or collision, while a positive value indicates a gap, showing that the components are separated by a specific distance. The "angle with the horizontal plane" refers to the angle between the plane or line formed by the two components and the horizontal plane. For spatial relationships falling under categories 2 to 5, the "angle with the horizontal plane" is calculated as the angle between the plane or line formed by the positioning points/lines and the horizontal plane. For category 1, the angle is 0°. Based on geometric principles, it is clear that these five calculated spatial relationships can comprehensively describe the relative positional relationship between two components.

3.4 Network-based Representations of BIM Models

This section further converts the components, their attributes, and relationships extracted from Sections 2.1, 2.2, and 2.3 into network nodes, node features, edges, and edge features, to achieve the network representation of the BIM model. The construction process is outlined in Table 2.

Specifically, all components in the BIM model are treated as nodes in the network. Based on the classification of component categories in IFC, we treat these nodes into different types, such as walls, beams, pipes, and so on. At the same time, all semantic properties of the components extracted in Section 2.1 are converted

into the corresponding node features. The topological relationships extracted in Section 2.2 are represented as edges connecting the corresponding component nodes. These edges are further classified into three categories based on the three different types of topological relationships. Additionally, the spatial relationships between a component and all components within a local 0.5m range, as described in Section 2.3, are represented as the fourth type of edge in the network. The calculated relative positional relationships between two components are used as the features of this type of edge.

It should be noted that, when two component nodes are connected by a topological relationship, they are inherently adjacent, with a relative distance of zero. This proximity implies an inherent spatial relationship between them, which could lead to redundancy if both spatial and topological relationships are represented as separate edges. To prevent this redundancy, we adopted a strategy where the spatial relationship is disregarded when a topological relationship is present, ensuring that the two component nodes are connected by only one edge. This methodology effectively reduces redundancy while maintaining the integrity of both the spatial and topological relationships within the network.

Table 2. Process of constructing the BIM-specific network

Algorithm

Input: Components $\boldsymbol{\mathcal{C}}$ from BIM model $\boldsymbol{\mathcal{M}}$
Output: Network $\mathcal{N}\{\boldsymbol{\mathcal{V}}, \boldsymbol{\mathcal{E}}\}$; node features $\{\mathbf{n}_v, \forall v \in \boldsymbol{\mathcal{V}}\}$; edge features $\{\mathbf{n}_e, \forall e \in \boldsymbol{\mathcal{E}}\}$

For each component type $\boldsymbol{\mathcal{T}}$ in $\boldsymbol{\mathcal{C}}$ do
  For each component i in $\boldsymbol{\mathcal{T}}$ do
    $v_i^{\mathcal{T}}$ = create_node(i)
    $\mathbf{n}_{v_i^{\mathcal{T}}}$ = extract_semantic_ properties(i)
  end for
end for
For each node $v_i^{\mathcal{T}}$ in $\boldsymbol{\mathcal{V}}$ do
  Nodes $\boldsymbol{v}_{\boldsymbol{all_j}}^{\boldsymbol{\mathcal{T}}}$ = get_host_relationship(i)
  For $v_j^{\mathcal{T}}$ in $\boldsymbol{v}_{\boldsymbol{all_j}}^{\boldsymbol{\mathcal{T}}}$:
    $\left(v_i^{\mathcal{T}}, v_j^{\mathcal{T}}\right)_h$ = create_host_edge$\left(v_i^{\mathcal{T}}, v_j^{\mathcal{T}}\right)$
  end for
  Nodes $\boldsymbol{v}_{\boldsymbol{all_k}}^{\boldsymbol{\mathcal{T}}}$ = get_connection_relationship(i)
  For $v_k^{\mathcal{T}}$ in $\boldsymbol{v}_{\boldsymbol{all_k}}^{\boldsymbol{\mathcal{T}}}$:
    $\left(v_i^{\mathcal{T}}, v_k^{\mathcal{T}}\right)_c$ = create_connection_edge$\left(v_i^{\mathcal{T}}, v_k^{\mathcal{T}}\right)$
  end for
  Nodes $\boldsymbol{v}_{\boldsymbol{all_l}}^{\boldsymbol{\mathcal{T}}}$ = get_touch_floor_relationship(i)
  For $v_l^{\mathcal{T}}$ in $\boldsymbol{v}_{\boldsymbol{all_l}}^{\boldsymbol{\mathcal{T}}}$:
    $\left(v_i^{\mathcal{T}}, v_l^{\mathcal{T}}\right)_T$ = create_touch_floor_edge$\left(v_i^{\mathcal{T}}, v_l^{\mathcal{T}}\right)$
  end for
  For each node $v_m^{\mathcal{T}}$ in $\boldsymbol{\mathcal{V}}$ and $m \neq i$:
    if (distance between m and i < spatial relationship threshold) and (not $\left(v_i^{\mathcal{T}}, v_m^{\mathcal{T}}\right)_h$&$\left(v_i^{\mathcal{T}}, v_m^{\mathcal{T}}\right)_c$&$\left(v_i^{\mathcal{T}}, v_m^{\mathcal{T}}\right)_T$):
      $\left(v_i^{\mathcal{T}}, v_m^{\mathcal{T}}\right)_S$ = create_spatial_edge$\left(v_i^{\mathcal{T}}, v_m^{\mathcal{T}}\right)$
      $\mathbf{n}_{\left(v_i^{\mathcal{T}}, v_m^{\mathcal{T}}\right)_S}$ = calculate_relative_position(i, m)
    end for
  end for

4 Case Study

Our network-based representation method achieves seamless integration with mainstream BIM software by leveraging the standardized IFC format as the universal data interface. Since IFC serves as the open BIM interoperability standard supported by all major platforms (Revit, ArchiCAD, etc.), our method is theoretically applicable to nearly all BIM software. Users simply need to export the model as an IFC file and use the proposed program for conversion. This section uses a residential floor as a case study to present how the method proposed in this research can be applied to convert a BIM model into a network. The resulting overall network and the zoomed-in local network are shown in Figure 3. Notably, node positions in the figure are determined by the coordinates of their corresponding components' central points.

As shown in Figure 3, representing a BIM model as a network effectively simplifies the complex building information and clearly expresses the multi-dimensional features of components and their relationships, such as semantic, spatial, and topological relationships. Additionally, the introduction of the touch floor relationship and the spatial relationships within a 0.5m local range further enhances the network by adding more local relational information. Overall, this method effectively compresses redundant geometric data while retaining as much of the hierarchical component attributes and spatial-temporal dependencies as possible, resulting in a structured yet information-rich network representation.

With the rapid development of machine learning, particularly in the context of nonlinear and fuzzy learning, there is an increasing need for structured tabular data for training. Our network-based BIM model addresses this requirement by providing a clear and structured format that can be used for training machine learning models, including graph neural networks (GNNs). Given that our model incorporates numerous spatial and topological features, it offers a powerful foundation for learning and reusing design features from historical models. For instance, GNNs can be employed to mask and reconstruct

node features in an unsupervised manner, enabling the model to deeply capture implicit design patterns within the BIM data. This allows the model to understand complex relationships and interactions between components that are not explicitly labeled but are critical for tasks like design optimization, automated checks, and conflict detection. Additionally, our method is directly applicable to downstream tasks. For example, in automated checks and conflict detection, we can convert these tasks into edge relationship prediction tasks, where the goal is to predict the relationships between two nodes (components) in the network. Using GNNs, we can predict whether any potential conflicts or design issues exist based on the spatial and topological relationships encoded in the network. This approach not only enables accurate detection but also facilitates the optimization and validation of designs in a more automated and efficient manner.

In summary, the network-based representation we propose supports machine learning algorithms, especially GNNs, by providing a structured format that captures both explicit and implicit design features. This facilitates tasks such as conflict detection, design optimization, and more, by making the data more accessible and interpretable for downstream analysis.

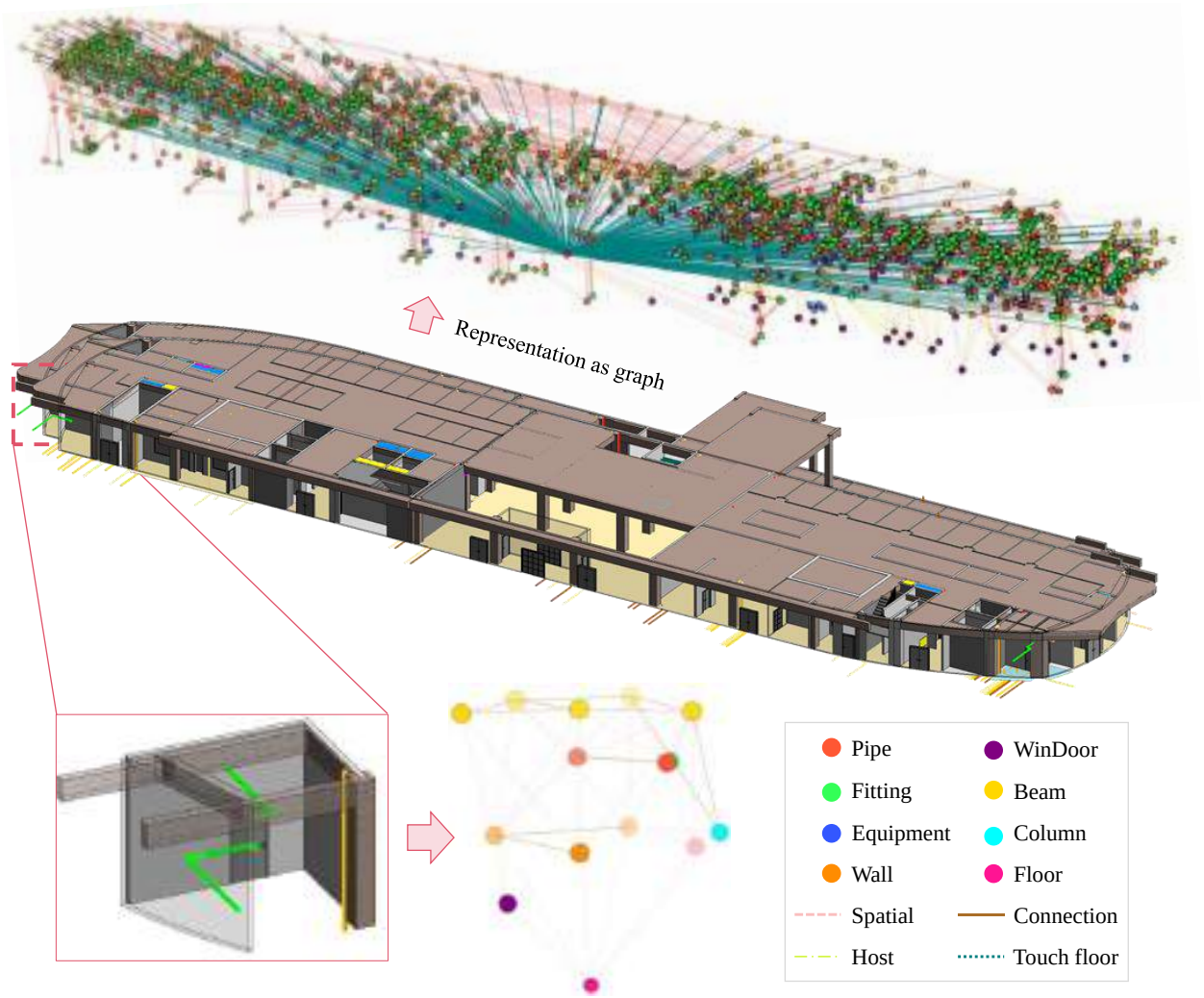


Figure 3. Example of local spatial relationship

5 Conclusion

This study proposes a unified network-based representation method for BIM models, effectively capturing the semantic, spatial, and topological features of building components. By transforming IFC data into a structured network format, our approach not only simplifies the representation but also enhances its utility for machine learning and other advanced analytical tasks. The proposed method enables more efficient design feature extraction, conflict detection, and optimization,

offering a promising framework for leveraging BIM data in various downstream applications.

Beyond the scope of the current research, future improvements can be pursued through the following approaches:

(a) Additional component features, such as time or design phase, can be integrated into the node and edge attributes to account for the impacts of construction processes and design stages, thus broadening the scope of downstream transfer learning tasks. For instance, in spatiotemporal clash detection, properties like construction priority and design phase can be incorporated. Similarly, for operation and maintenance applications, attributes such as access space and frequency of access can be included.

(b) By applying advanced machine learning techniques such as GNNs to the building data converted into a network, we can learn and capture implicit design patterns within the BIM data, enabling the transfer and reuse of historical data.

Acknowledgments

The authors are grateful for the financial support received from the National Key R&D Program of China (No. 2023YFC3804600) and the National Natural Science Foundation of China (No. 52378306).

References

[1] Smith, P. (2014). BIM implementation–global strategies. *Procedia engineering*, *85*, 482-492.

[2] Antwi-Afari, M. F., Li, H., Pärn, E. A., & Edwards, D. J. (2018). Critical success factors for implementing building information modelling (BIM): A longitudinal review. *Automation in construction*, *91*, 100-110.

[3] Leng, S., Lin, J. R., Li, S. W., & Hu, Z. Z. (2021). A data integration and simplification framework for improving site planning and building design. *IEEE Access*, *9*, 148845-148861.

[4] Lin, J.R.*, Chen, K.Y., Pang, P., Wu, D.P., Zhang, J.P. (2024). ChatBIM: Interactive Data Retrieval and Visualization from BIM. *the 10th International Conference on Innovative Production and Construction (IPC 2024)*, 265-266. Perth, Australia.

[5] Hu, Y., Xia, C., Chen, J., & Gao, X. (2023). Clash context representation and change component prediction based on graph convolutional network in MEP disciplines. *Advanced Engineering Informatics*, *55*, 101896.

[6] Zheng, Y., Zhang, Y., & Lin, J. (2023). BIM–based time-varying system reliability analysis for buildings and infrastructures. *Journal of Building Engineering*, *76*, 106958.

[7] Rad, M. A. H., Jalaei, F., Golpour, A., Varzande, S. S. H., & Guest, G. (2021). BIM-based approach to conduct Life Cycle Cost Analysis of resilient buildings at the conceptual stage. *Automation in Construction*, *123*, 103480.

[8] He, Z., Wang, Y. H., & Zhang, J. (2025). Generative AIBIM: An automatic and intelligent structural design pipeline integrating BIM and generative AI. *Information Fusion*, *114*, 102654.

[9] IFCwiki, Open source projects supporting IFC, http://www.ifcwiki.org/index.php/Open_Source (accessed: June 16, 2024).

[10] Zhou, Y. C., Zheng, Z., Lin, J. R., & Lu, X. Z. (2022). Integrating NLP and context-free grammar for complex rule interpretation towards automated compliance checking. *Computers in Industry*, *142*, 103746.

[11] Zhou, Y. W., Hu, Z. Z., Lin, J. R., & Zhang, J. P. (2020). A review on 3D spatial data analytics for building information models. *Archives of Computational Methods in Engineering*, *27*(5), 1449-1463.

[12] Lin, W. Y., & Huang, Y. H. (2019). Filtering of irrelevant clashes detected by BIM software using a hybrid method of rule-based reasoning and supervised machine learning. *Applied Sciences*, *9*(24), 5324.

[13] Wang, L., & Leite, F. (2013). Knowledge discovery of spatial conflict resolution philosophies in BIM-enabled MEP design coordination using data mining techniques: a proof-of-concept. In *Computing in Civil Engineering (2013)* (pp. 419-426).

[14] Liu, H., Gan, V. J., Cheng, J. C., & Zhou, S. A. (2024). Automatic Fine-Grained BIM element classification using Multi-Modal deep learning (MMDL). *Advanced Engineering Informatics*, *61*, 102458.

[15] Wang, Y., Zhang, L., Yu, H., & Tiong, R. L. (2022). Detecting logical relationships in mechanical, electrical, and plumbing (MEP) systems with BIM using graph matching. *Advanced Engineering Informatics*, *54*, 101770.

[16] Hu, Y., Castro-Lacouture, D., Eastman, C. M., & Navathe, S. B. (2020). Automatic clash correction sequence optimization using a clash dependency network. *Automation in Construction*, *115*, 103205.

[17] Li, K., Gan, V. J., Li, M., Gao, M. Y., Tiong, R. L., & Yang, Y. (2024). Automated generative design and prefabrication of precast buildings using integrated BIM and graph convolutional neural network. *Developments in the Built Environment*, *18*, 100418.

Vision Transformer-based Local Physical Fatigue Assessment using Electromyography (EMG) Signals for Construction Worker Health Monitoring

Yogesh Gautam,[1] Yuming Zhang, [2] Amit Ojha[3] and Houtan Jebelli, Ph.D.[4,*]

[1,2,3] Ph.D. Student, Department of Civil and Environmental Engineering, University of Illinois Urbana-Champaign, Champaign, IL.
[4,*]Assistant Professor, Department of Civil and Environmental Engineering, University of Illinois Urbana-Champaign, Champaign, IL. (corresponding author)
ygautam2@illinois.edu ,yumingz4@illinois.edu, amito2@illinois.edu, hjebelli@illinois.edu

Abstract

Construction workers frequently endure high levels of physical strain, which can lead to fatigue and long-term musculoskeletal disorders. Traditional methods of fatigue assessment often fail to provide timely and accurate diagnoses, particularly in dynamic work environments. In this paper, we present a Vision Transformer-based approach for assessing localized physical fatigue using Electromyography (EMG) signals. This method leverages the high learning capacity of Vision Transformers to capture both spatial and temporal features from EMG data, enabling precise detection and prediction of fatigue. By focusing on specific muscle groups that are prone to fatigue during construction tasks, our model delivers detailed insights into the onset of fatigue at a localized level. Experimental evaluations conducted on roofing tasks targeting the biceps brachii and brachioradialis demonstrate that the proposed method can predict fatigue to a localized level. The results demonstrate the potential of this approach for enhancing on-site health monitoring systems and establishing a foundation for proactive interventions aimed at reducing the risk of fatigue-induced injuries among construction workers.

Keywords –

EMG ; Vision Transformer; CNN; Fatigue Detection

1 Introduction

Physical fatigue is a critical factor influencing worker productivity, health, and safety in various industries such as manufacturing, construction, and logistics [1]. Workers exposed to prolonged or repetitive physical tasks often experience muscle fatigue, which can impair motor control, reaction times, and cognitive function. This leads to a heightened risk of injuries and accidents, particularly in jobs that involve lifting, repetitive movements, or prolonged exertion [2]. Given the high stakes, detecting and managing physical fatigue in real-time is vital for reducing workplace injuries and ensuring worker well-being [3].

Accurately assessing physical fatigue remains a challenging problem due to its subjective nature and the variability in how it manifests across different individuals. Traditional fatigue assessment methods, such as self-reported surveys (e.g., Borg Fatigue Scale) or physical performance tests, are limited by their inherent subjectivity and time-consuming nature [4,5]. Moreover, these methods often fail to capture real-time changes in muscle fatigue, making them impractical for dynamic, high-risk environments where quick decision-making is crucial.

Electromyography (EMG) is an objective method for assessing muscle activity by measuring the electrical signals generated by muscle contractions. As muscles fatigue, changes occur in both the amplitude and frequency of the EMG signals, making it possible to quantify muscle fatigue objectively [6]. However, analyzing EMG data for fatigue detection presents challenges due to the complex, non-linear nature of the signals, the presence of noise, and variability caused by different muscle groups, task conditions, and individual characteristics.

In recent years, machine learning (ML) has emerged as a promising approach for analyzing complex bio-signals, including EMG [7]. By learning patterns from data, ML models can detect subtle changes in EMG signals associated with muscle fatigue. Despite the promise of ML, traditional models such as support vector machines and convolutional neural networks rely heavily on manual feature extraction and are often limited by their inability to capture long-range dependencies in the

data, particularly in time-series signals like EMG.

Transformers, originally developed for natural language processing, have gained attention in the field of computer vision due to their powerful ability to capture global dependencies within input data [8]. The Vision Transformer (ViT), in particular, has shown state-of-the-art performance on various image classification tasks. Unlike traditional CNNs, which use localized filters to extract features, the ViT relies on a self-attention mechanism to capture both local and global features from input images or 2D data. This makes ViTs particularly well-suited for analyzing spectrograms of EMG signals, where capturing both temporal and frequency information is crucial for fatigue detection.

One of the key challenges in applying ML models to EMG data is the variability and noise present in time-domain signals. By transforming EMG signals into spectrograms (time-frequency representations), it becomes possible to extract both temporal and spectral features from a single representation [9]. This approach not only reduces noise but also enhances the model's ability to detect tempo-spectral patterns in the data and provides image representation which can directly processed by the ViT without any manual feature extraction. As ViTs can model both local and global context in EMG spectrograms, it makes them well-suited for distinguishing between different levels of physical fatigue.

In this paper, we propose a Vision Transformer-based approach for assessing local physical fatigue using EMG signals. The method involves converting EMG signals into spectrograms, which are then input into a fine-tuned Vision Transformer model for classification. The objective is to demonstrate that this approach can effectively classify different levels of fatigue (low, medium, high) with higher accuracy and robustness compared to traditional machine learning techniques. By leveraging the power of the ViT model, we aim to provide a scalable, real-time solution for fatigue monitoring, particularly in labor-intensive work environments.

2 Background and Related Work

2.1 EMG and Fatigue Assessment

Electromyography (EMG) has emerged as a crucial technique for evaluating muscle activity by capturing electrical signals generated during muscle contractions. This method offers direct insights into muscle activation patterns and serves as a valuable tool for assessing muscle fatigue [10]. As muscles experience fatigue, EMG signals undergo characteristic changes, including amplitude increases and frequency decreases, reflecting the underlying physiological processes such as metabolic byproduct accumulation and additional motor unit recruitment. The analysis of EMG signals for fatigue detection involves a multi-step process:

1. Signal preprocessing (filtering, normalization)
2. Feature extraction
3. Pattern recognition

Traditional approaches have relied on signal processing methods like Fourier and Wavelet Transforms for feature extraction. However, these techniques often require manual feature selection and struggle to capture the complex, non-linear patterns associated with muscle fatigue [4]. Furthermore, EMG signals are susceptible to various sources of noise and artifacts, including movement and electrode displacement, which complicate the analysis process. Time-frequency domain analysis, particularly through the use of spectrograms, has gained traction as an effective method for representing EMG signals. Spectrograms offer a visual depiction of a signal's frequency content evolution over time, facilitating the identification of fatigue-related patterns. This representation not only enhances the interpretability of EMG data but also paves the way for the application of advanced machine learning techniques in fatigue classification. Recent advancements in machine learning have opened new avenues for EMG signal analysis: Deep Learning Models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have shown promise in automatically extracting relevant features from raw EMG data. Ensemble methods such as combining multiple classifiers has demonstrated improved robustness and accuracy in fatigue detection. On the other hand, transfer learning methods leveraging pre-trained models on large datasets has enabled more efficient training on smaller, task-specific EMG datasets [11–13].

Despite these advancements, challenges remain in developing localized fatigue assessment methods for diverse muscle groups and tasks. The integration of EMG-based fatigue detection systems into practical workplace settings, particularly in dynamic environments like construction sites, presents additional hurdles that require innovative solutions.

2.2 Machine Learning for Fatigue Classification

Machine learning (ML) has revolutionized the analysis of complex bio-signals, particularly EMG, for physical fatigue detection and classification. The field has evolved from traditional ML approaches to sophisticated deep learning methods, each offering unique advantages in processing EMG data [14].

Early efforts in EMG-based fatigue classification relied on conventional ML algorithms such as Support Vector Machines (SVM), Random Forests, and Decision Trees [15]. These methods demonstrated varying degrees of success in handling high-dimensional data and non-

linear classification problems [16].

The advent of deep learning has significantly improved EMG signal analysis. Convolutional Neural Networks (CNNs) have shown exceptional performance in automating feature extraction from raw or minimally processed EMG data. Liu et al. proposed a hybrid CNN-LSTM model that outperformed standalone LSTM, CNN, and RF models approaches in muscle fatigue recognition [17]. Long Short-Term Memory (LSTM) networks have also demonstrated promise in capturing temporal dynamics in EMG signals, with a study by Wang et al. developing an LSTM-based muscle fatigue recognition model that surpassed CNN and SVM in classification performance [18].

Despite these advancements, several challenges persist in EMG-based fatigue classification. These include the generalizability of models across different tasks and individuals, the ability to capture long-range dependencies in EMG signals, and the development of real-time processing capabilities for dynamic work environments. Recent developments aim to address these challenges through hybrid models, time-frequency analysis, and transfer learning approaches [7].

The field is moving towards more sophisticated and integrated approaches, including multi-modal analysis combining EMG with other physiological signals, the development of explainable AI for better interpretation of model decisions, and adaptive learning techniques that can adjust to individual differences and changing conditions in real-time. As research progresses, the integration of advanced ML techniques with domain-specific knowledge in physiology and biomechanics promises to yield more accurate, robust, and practical solutions for EMG-based fatigue detection and classification, particularly in dynamic work environments such as construction sites.

2.3 Vision Transformers and Bio-signals

Vision Transformers (ViTs) represent a significant leap forward in deep learning, particularly in computer vision tasks. Unlike traditional Convolutional Neural Networks, ViTs leverage the Transformer architecture's self-attention mechanism to capture both local and global dependencies within input data. This unique approach allows ViTs to model complex spatial relationships and long-range interactions, making them highly effective for tasks that require a comprehensive understanding of the entire input [19]. In the realm of bio-signal processing, especially for EMG-based fatigue detection, ViTs offer several compelling advantages. By leveraging EMG spectrograms—images representing time-frequency information [20]—ViTs can directly analyze temporal and spectral information in EMG signals. Their ability to process and understand 2D spectrograms makes them well-suited for identifying subtle changes in EMG signals that correlate with different levels of muscle fatigue. Moreover, the global context capture capability of ViTs enhances their robustness against noise and variability in EMG data, potentially leading to more accurate and reliable fatigue classification.

Recent literature has demonstrated several applications of transformer-based architectures in bio signal processing [21][22].

These advancements suggest that transformer-based models, including ViTs, have significant potential in bio-signal processing. Their ability to capture complex spatial and temporal relationships in data makes them particularly suitable for analysing EMG signals for fatigue detection. This work leverages the advancements in ViT and electromyography techniques to develop framework for localized fatigue estimation in construction workers.

3 Methodology: ViT-based Fatigue Assessment

The proposed methodology for EMG-based fatigue assessment consists of three main parts: data acquisition, spectrogram generation, and fatigue classification using a Vision Transformer (ViT) model. In the first stage EMG signals is acquired from the muscle group of interest using surface electrodes, and bio-amplifier. These signals are then processed, which includes labeling

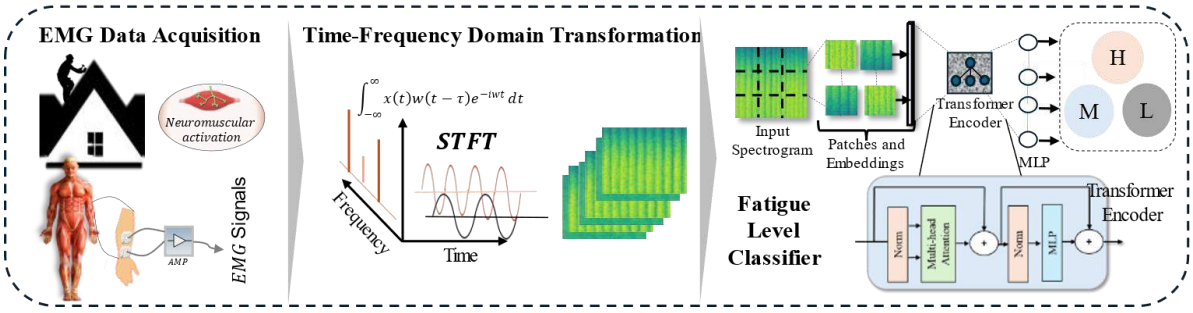


Figure 1: Overview of the methodology. The acquired EMG signals are converted to spectrograms using STFT

which is then used as input to the Vision Transformers model.

and windowing. For labeling, the Borg Fatigue rating scale is employed, providing a standardized measure of perceived exertion. This step is crucial for creating a labeled dataset that can be used to train and evaluate the Vision Transformer model. The acquired signals is used to generate a spectrogram through short-time Fourier transform. The generated spectrogram is then used as input to the Vit model which then leverages transformer architecture to classify between low, medium and high levels of fatigue (Figure 1).

3.1 Spectrogram Generation

The second phase involves transforming the time-domain EMG waveforms into 2D spectrograms using the Short-Time Fourier Transform (STFT). The STFT is defined by the equation 1:

$$STFT\{x\}(\tau,\omega) = X(\tau,\omega) = \int_{-\infty}^{\infty} x(t)w(t-\tau)e^{-iwt}dt \quad (1)$$

Where, x represents the raw EMG signals in the time domain, and w(t) is a non-zero window function. This transformation allows for the visualization of frequency changes over time, providing a richer representation of the EMG data compared to the raw time-domain signal. The use of spectrograms offers several significant advantages in EMG signal analysis. Firstly, it captures both temporal and spectral information in a single representation, allowing for a comprehensive view of the signal's characteristics. Secondly, the spectrogram transformation effectively reduces noise and enhances relevant features in the EMG signal, improving the signal-to-noise ratio and facilitating more accurate analysis. Lastly, by converting the EMG data into a 2D image-like format, spectrograms enable the application of powerful image processing techniques to bio-signal data, opening up new possibilities for advanced analysis and feature extraction.

3.2 Fatigue Classification using Vision Transformer

The final stage involves using the generated spectrograms as input to a Vision Transformer (ViT) model for fatigue level classification. The ViT model, pre-trained on the ImageNet-21K dataset , is fine-tuned for the specific task of EMG-based fatigue detection. b Key aspects of the ViT model include:

Self-Attention Mechanism: The model uses self-attention to process the entire spectrogram as a sequence of tokens, capturing long-range dependencies and global context. This is particularly beneficial for understanding complex spatial relationships in the spectrograms.

Model Architecture: The ViT consists of multiple Transformer blocks, each comprising self-attention and feedforward layers. The self-attention mechanism is mathematically expressed as 2:

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2)$$

where Q, K, and V denote the query, key, and value matrices derived from input embeddings, and d_k represents the dimensionality of the key vectors.

Feedforward Neural Networks: Each Transformer block incorporates feedforward neural networks to further process token embeddings:

Fine-tuning Process: The pre-trained ViT model is adapted to the specific task of EMG-based fatigue classification through fine-tuning on a acquired dataset of EMG spectrograms (roofing tasks).

Loss Function: The model is trained using categorical cross-entropy as the loss function (Equation 3).

$$H\big(g_f(I_{2D})\big) = -\sum_{j=1}^{k} y_j \log\left(g_f(I_{2D})\right) \quad (3)$$

Here g_f is the ViT model where whereas I_{2D} is the input 2D spectrogram, y_j represents true label of the class whereas k represents the number of classes.

By using spectrograms as an intermediate representation, the approach makes the use of power architecture like ViT feasible for time-series-based EMG signals.

4 Experimental Case Study

The experimental study was designed to comprehensively assess muscle fatigue in construction workers through EMG signal analysis. Twelve volunteers were recruited to participate in a carefully controlled experimental protocol that simulated realistic construction tasks, specifically focusing on roofing activities that engage the upper limb muscle groups. The participant selection process made sure that all subjects were free from medical conditions that could potentially interfere with the experimental measurements.

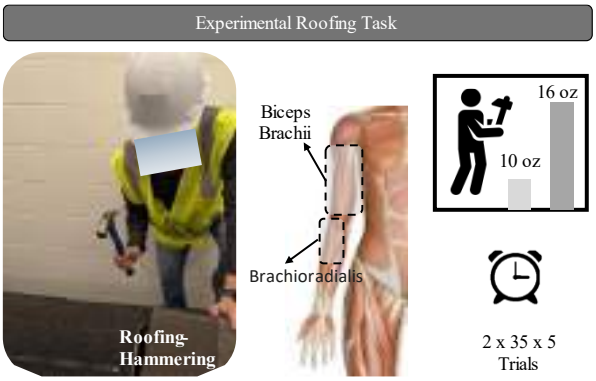


Figure 2: Roofing task was used to evaluate performance of the developed framework. Subjects hammered shingles into the roof structure during the task when EMG signals were acquired.

Electrode placement was a critical aspect of the methodology. Specialized surface EMG electrodes were positioned bilaterally over the biceps and brachioradialis muscle group. A multichannel ADInstruments bio-amplifier was utilized to acquire high-resolution EMG signals, with a sampling frequency of 1 kHz and a window size of 3 seconds. A sliding window of 1 seconds was used for the signal's segmentation. The average time of experimentation was around 12 minutes per subject excluding setup and rest periods. The generated spectrograms were split into 80:20 rationfor training and testing with a stratified split to make it consistent across all the classes.

The electrode configuration allowed for precise tracking of muscle electrical activity during various physical tasks. The Borg Rating of Perceived Exertion (RPE) scale played a crucial role in fatigue classification. This well-established subjective measure allows participants to rate their perceived level of physical exertion on a scale typically ranging from 6 to 20. In this study, the RPE was used to categorize fatigue levels into three distinct classes: low, medium, and high. Participants were trained to provide self-assessments, ensuring the reliability of the subjective fatigue ratings.

The experimental tasks were carefully designed to simulate real-world construction scenarios, with a particular focus on roofing activities. Participants performed controlled roofing tasks that engaged the biceps brachii and brachioradialis muscles, mimicking the physical demands of actual construction work. The tasks were structured to progressively induce muscle fatigue, allowing to capture the nuanced changes in muscle activation patterns over time (using a hammer of increased load). Signal processing played a critical role in transforming raw EMG data into meaningful insights. The acquired signals underwent several preprocessing steps, including filtering to remove noise, normalization to account for variations and transformation into spectrograms using Short-Time Fourier Transform (STFT). This approach converted time-domain signals into two-dimensional representations that could be effectively analysed using the Vision Transformer model. From twelve subjects with

During the experiment, participants were given adequate rest periods between tasks to minimize cumulative fatigue effects and ensure the reliability of the measurements. By integrating objective EMG measurements with subjective fatigue ratings, the study aimed to develop a comprehensive understanding of muscle fatigue in construction workers.

5 Results and Discussion

The developed fatigue estimation model based on a Vision Transformer (ViT) network demonstrated promising performance in classifying fatigue levels using EMG signals from the biceps brachii and brachioradialis muscles. The model's performance varied across different muscle groups, with distinct patterns observed

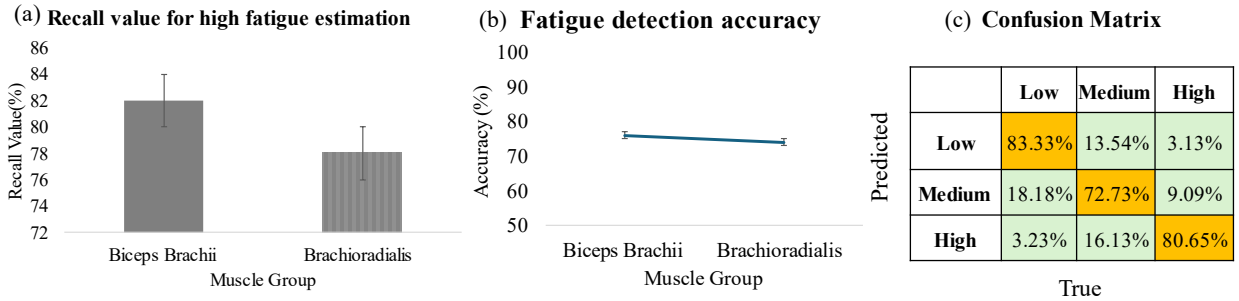


Figure 3: (a) Bar diagram for recall value for the high fatigue estimation. (b) Accuracy for fatigue estimation. (c) Confusion matrix for the fatigue estimation through ViT.

in fatigue detection accuracy.

For high fatigue estimation, the recall values showed notable differences between muscle groups, with biceps brachii achieving approximately 82% recall and brachioradialis showing around 78% recall. The overall fatigue detection accuracy remained relatively consistent across both muscle groups, with values around 75% (biceps brachii- 76% and brachioradialis – 74%).

The model showed the strongest performance in identifying low fatigue states (83.3% accuracy) , followed by high fatigue states (80.65% accuracy), while medium fatigue states presented the most challenging classification task (72.73% accuracy). This pattern suggests that the transition between fatigue states, particularly in the medium range, requires careful consideration in future model refinements. The superior performance of the ViT model can be attributed to its ability to capture global context through the analysis of spectrograms. By transforming EMG signals into time-frequency representations, the model leverages its strength in processing two-dimensional data, which allows it to identify subtle changes in muscle activation associated with varying levels of fatigue. The advantages of using spectrograms as input for the ViT model remain significant, facilitating comprehensive analysis of both temporal and spectral features. This multi-dimensional approach enhances the model's ability to assess fatigue accurately across different muscle groups and fatigue states.

For a more comprehensive analysis, the authors also compared the manual feature extraction-based method using both time domain feature (Signal Mean Absolute Value and Root Mean Square) and frequency domain feature (Median Frequency (MDF) and Mean Frequency (MEF) [5] and trained a SVM, ANN, and random forest models from the extracted features. The SVM model provided an average accuracy of 59.4%, which was the highest among the competing model. Table 2 outlines the average accuracy and its comparison with the ViT model.

Table 1: Comparison with ML models using manually extracted feature from EMG signals

Model	Accuracy	% improvement in VIT
SVM	59.4 %	26 %
ANN	51.2 %	46 %
Random Forest	58.1 %	29 %

While the developed mode is able to predict localized fatigue levels with a decent accuracy, several challenges and limitations remain. Real-time processing capabilities are crucial for practical applications, and while our model shows promise, further optimization may be necessary to ensure efficiency in resource-constrained environments. Additionally, variability across participants can affect performance; individual differences in muscle composition and fatigue response patterns may lead to inconsistent results. Moreover, a larger and more diverse dataset would likely improve the robustness of the model.

Looking ahead, there are several promising avenues for future research. Exploring transfer learning from larger biomedical datasets could enhance the model's performance by providing additional training data that captures a wider range of muscle activity patterns. Moreover, integrating multi-modal data—such as heart rate and skin temperature—could offer a more comprehensive view of fatigue status, potentially improving detection accuracy. Investigating adaptive learning techniques may also allow for personalization of the model for individual users, further enhancing its applicability across diverse populations.

Overall, our study highlights the potential of deep learning integrated EMG-based models for localized fatigue detection in physically demanding environments such as construction sites. By addressing current limitations and exploring future directions, we aim to develop a robust system that can contribute significantly to worker safety and productivity through effective fatigue monitoring.

6 Conclusions

This study presents a novel approach for local physical fatigue assessment using a Vision Transformer-based model applied to EMG spectrograms. The developed system demonstrates significant potential for real-time fatigue monitoring in occupational settings, particularly in dynamic environments such as construction sites. With an achieved accuracy of 75% and and recall of over 80%, our model outperforms traditional machine learning approaches in classifying different levels of muscle fatigue. The success of this ViT-based approach highlights the advantages of leveraging advanced deep learning techniques for bio-signal processing. By capturing both spatial and temporal features from EMG data, the model provides a more nuanced understanding of fatigue progression, crucial for timely interventions in workplace safety. The use of spectrograms as input to the ViT model proved particularly effective, allowing for a comprehensive analysis of muscle activation patterns associated with fatigue. The implications of this research extend beyond mere performance metrics. Accurate and real-time fatigue detection systems have the potential to revolutionize workplace safety practices. By enabling

proactive interventions, such systems can help prevent injuries, optimize task assignments, and improve overall worker health outcomes. The integration of such technology into existing ergonomics programs and fatigue management systems could lead to significant improvements in workplace safety and productivity. However, this study also acknowledges several limitations and areas for future research. Expanding the dataset to include a more diverse range of participants and occupational tasks will be crucial for enhancing the model's generalizability. Future work should also explore the integration of multi-modal data and further optimize real-time processing capabilities. Additionally, addressing ethical considerations and privacy concerns will be essential for the widespread adoption of such monitoring systems in workplace environments. In summary, this research represents a significant step forward in the field of occupational health and safety, offering a promising tool for real-time fatigue assessment. As we continue to refine and expand upon this approach, the potential for creating safer, more productive work environments becomes increasingly tangible, underscoring the importance of continued research and development in this critical area.

Acknowledgments

The work presented in this paper was supported by a National Science Foundation Award (No. 2401745, 'Future of Construction Workplace Health Monitoring'). Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] Namian M, Taherpour F, Ghiasvand E, Turkan Y. Insidious Safety Threat of Fatigue: Investigating Construction Workers' Risk of Accident Due to Fatigue. J Constr Eng Manag 2021;147:04021162. https://doi.org/10.1061/(ASCE)CO.1943-7862.0002180/ASSET/C4FFF408-D81F-4F2D-83B4-5E046312FD01/ASSETS/IMAGES/LARGE/FIGURE9.JPG.

[2] Liu Y, Gautam Y, Ojha A, Shayesteh S, Jebelli H. Studying the Effects of Back-Support Exoskeletons on Workers' Cognitive Load during Material Handling Tasks. Construction Research Congress 2024, CRC 2024 2024;1:659–69. https://doi.org/10.1061/9780784485262.067.

[3] Liu Y, Ojha A, Jebelli H. Vision-based Ergonomic Risk Assessment of Back-Support Exoskeleton for Construction Workers in Material Handling Task. 2023 International Conference on Computing in Civil Engineering (I3CE) 2023.

[4] Aryal A, Ghahramani A, Becerik-Gerber B. Monitoring fatigue in construction workers using physiological measurements. Autom Constr 2017;82:154–65. https://doi.org/10.1016/J.AUTCON.2017.03.003.

[5] Jebelli H, Lee S. Feasibility of Wearable Electromyography (EMG) to Assess Construction Workers' Muscle Fatigue. Advances in Informatics and Computing in Civil and Construction Engineering 2019:181–7. https://doi.org/10.1007/978-3-030-00220-6_22.

[6] Stegeman DF, Blok JH, Hermens HJ, Roeleveld K. Surface EMG models: properties and applications. Journal of Electromyography and Kinesiology 2000;10:313–26. https://doi.org/10.1016/S1050-6411(00)00023-7.

[7] Ojha A, Gautam Y, Jebelli H, Akanmu A. Physiological impact of powered back-support exoskeletons in construction: Analyzing muscle fatigue, metabolic cost, ergonomic risks, and stability. Autom Constr 2024;168:105742. https://doi.org/10.1016/J.AUTCON.2024.105742.

[8] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is All you Need. Adv Neural Inf Process Syst 2017;30.

[9] Gautam Y, Jebelli H. Design of flexible polyimide-based serpentine EMG sensor for AI-enabled fatigue detection in construction. Sens Biosensing Res 2024;46:100713. https://doi.org/10.1016/J.SBSR.2024.100713.

[10] Dimitrova NA, Dimitrov G V. Interpretation of EMG changes with fatigue: facts, pitfalls, and fallacies. Journal of Electromyography and Kinesiology 2003;13:13–36. https://doi.org/10.1016/S1050-6411(02)00083-4.

[11] Shayesteh S, Ojha A, Jebelli H. Workers' Trust in Collaborative Construction Robots: EEG-Based Trust Recognition in an Immersive Environment. Automation and Robotics in the Architecture, Engineering, and Construction Industry 2022:201–15. https://doi.org/10.1007/978-3-030-77163-8_10.

[12] Gautam Y, Ojha A, Liu Y, Jebelli H. Advancing Realtime Physiological Health Monitoring in Construction: A Feasibility Study of Fatigue Estimation of Construction Workers on Smart

Devices. 2024 ASCE International Conference on Computing in Civil Engineering, Pittsburgh: 2024.

[13] Ojha A, Sharifironizi A, Liu Y, Jebelli H. Enhancing Human-Centric Physiological Data-Driven Heat Stress Assessment in Construction through a Transfer Learning-Based Approach. Construction Research Congress 2024, CRC 2024 2024;1:157–67. https://doi.org/10.1061/9780784485262.017.

[14] Jebelli H, Seo JO, Hwang S, Lee SH. Physiology-based dynamic muscle fatigue model for upper limbs during construction tasks. Int J Ind Ergon 2020;78:102984. https://doi.org/10.1016/J.ERGON.2020.102984.

[15] Zhang G, Yang B, Zan P, Zhang D. Multilevel Assessment of Exercise Fatigue Utilizing Multiple Attention and Convolution Network (MACNet) based on Surface Electromyography. IEEE Transactions on Neural Systems and Rehabilitation Engineering 2024. https://doi.org/10.1109/TNSRE.2024.3523332.

[16] Chand S, McDaid A, Lu Y. Dynamic muscle fatigue assessment using s-EMG technology towards human-centric human-robot collaboration. J Manuf Syst 2023;68:508–22. https://doi.org/10.1016/j.jmsy.2023.05.022.

[17] Liu J, Tao Q, Wu B. Dynamic Muscle Fatigue State Recognition Based on Deep Learning Fusion Model. IEEE Access 2023;11:95079–91. https://doi.org/10.1109/ACCESS.2023.3309741.

[18] Wang J, Sun S, Sun Y. A Muscle Fatigue Classification Model Based on LSTM and Improved Wavelet Packet Threshold. Sensors 2021, Vol 21, Page 6369 2021;21:6369. https://doi.org/10.3390/S21196369.

[19] Anwar A, Khalifa Y, Coyle JL, Sejdic E. Transformers in biosignal analysis: A review. Information Fusion 2025;114:102697. https://doi.org/10.1016/J.INFFUS.2024.102697.

[20] Zawawi TNST, Abdullah AR, Shair EF, Halim I, Rawaida O. Electromyography signal analysis using spectrogram. Proceeding - 2013 IEEE Student Conference on Research and Development, SCOReD 2013 2013:319–24. https://doi.org/10.1109/SCORED.2013.7002599.

[21] Anwar A, Khalifa Y, Coyle JL, Sejdic E. Transformers in biosignal analysis: A review. Information Fusion 2025;114:102697. https://doi.org/10.1016/J.INFFUS.2024.102697.

[22] Yan G, Liang S, Zhang Y, Liu F. Fusing Transformer Model with Temporal Features for ECG Heartbeat Classification. Proceedings - 2019 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2019 2019:898–905. https://doi.org/10.1109/BIBM47256.2019.8983326.

Automated Site Extraction for Scan-to-BIM Progress Monitoring in Road Network Construction

Ziang Jiang[1], Xuesong Shen[1], Khalegh Barati[1] and James Linke[2]

[1]School of Civil and Environmental Engineering, The University of New South Wales (UNSW), Australia
[2]GeoAI, Australia
ziang.jiang@unsw.edu.au, x.shen@unsw.edu.au, khalegh.barati@unsw.edu.au, j.linke@llsurveys.com.au

Abstract –

Construction progress monitoring is an essential activity for ensuring a project's quality and timely delivery. Scan-to-BIM demonstrates its potential and superiority in capturing the as-built status and handling design changes to achieve real-time project tracking. Previous studies have focused on its main steps, including segmentation, recognition, and reconstruction, but often overlooked target area extraction due to their emphasis on residential buildings. Manual extraction can be tedious, time-consuming, and labor-intensive, especially for large-scale, horizontal transport infrastructures. Although similar applications can be found in other domains, these methods may not be feasible for extracting road networks construction sites due to differences in data structure, the as-built status of roads, and the requirements for the Level of Detail (LoD) in the output. This paper proposes an automated hybrid road construction site extraction framework as the input for subsequent Scan-to-BIM construction progress monitoring. It leverages both local feature analysis and graph theory. The developed road patch generation and connection processes enable the extraction of complex road networks with high generality and versatility. Two sets of Airborne Laser Scanning (ALS) data with different construction progress were used to validate the proposed framework, achieving site extraction accuracies of 99.59% and 94.50%, respectively. These results indicate that the proposed method can effectively serve as a basis for supporting construction progress monitoring through Scan-to-BIM.

Keywords –

Point Cloud; Airborne Laser Scanning; Road Construction; Sit Extraction; Construction Progress Monitoring; Scan-to-BIM

1 Introduction

Frequent project monitoring is crucial for improving performance in the construction and renovation of transport infrastructure [1, 2]. It was reported that recent studies have focused on achieving project monitoring by creating as-built models for structures under construction [3-5]. Indeed, numerous road network extraction approaches can be found in the domains of remote sensing and computer vision, serving as the modeling foundation for supporting transport network design, analysis, optimization, and the generation of high-definition (HD) maps for autonomous driving [6, 7].

Ground-based Mobile Laser Scanning (MLS) and aerial images have been commonly used as primary data sources for modeling urban streets and highway systems in previous studies. However, these sensors are insufficient for capturing complex, large-scale transport infrastructure construction sites in terms of efficiency. As an alternative, ALS demonstrates superiority in road classification for complex environments compared to orthoimages [8]. Although learning-based methods have achieved advancements in object recognition, the lack of training datasets and the complexity of site environments—such as ruts, temporary roads, hills, and trenches—affect the performance of common classifiers developed for photogrammetry and laser scanning [9-12]. Additionally, these site conditions hinder the application of conventional image processing methods, such as edge detection or line fitting. Similarly, many widely used local feature analysis methods, including plane fitting and gradient mapping, face the same challenges. Due to differences in primary data sources, the targeted stages of a road's life cycle, and the expected output based on application requirements, a research gap exists. Currently, no proper automatic method for effectively extracting road networks at construction sites has been developed to support Scan-to-BIM construction progress monitoring through ALS.

The proposed hybrid methodology aims to achieve automated road construction site extraction, starting from the earthwork stage. This methodology is characterized by two key academic contributions: (1) automated site extraction for road construction with complex layouts; and (2) high generality and adaptability to complex site

and surrounding area condition, as well as compatibility with the early earthwork stage. The framework begins with generating a variant of a gradient map to extract grids containing steep slopes. Nonlinear continuous large-scale slopes, such as hills and stockpiles, are then removed by utilizing the grid's Degrees of Freedom (DoF). The remaining grids are grouped using a region-growing algorithm to form a series of boundaries. To extract the target areas, road patches are generated by pairing boundaries and then calculating a series of centroids between those boundary pairs. Subsequently, patch connections are applied by comparing the cross-sectional properties of nearby patches. Finally, the connection results are refined by analyzing the distribution of grids. Thus, the method can be regarded as a hybrid of local feature analysis and graph theory.

The remainder of this paper is structured as follows: Section 2 provides a comprehensive review of previous work on road network extraction. Section 3 describes the proposed site extraction framework in detail. Section 4 presents a validation of the method using two ALS datasets. Finally, Section 5 discusses the conclusions and future work.

2 Literature Review

Transport network modeling is essential task for the subsequent tasks in many domains [13]. Over two decades ago, researchers began exploring the feasibility of using MLS and integrated sensors to capture 3D road models [14]. As-built models of transport infrastructure can be used for traffic analysis, simulation, and optimization, supporting efficient planning, renovation, and system improvement throughout its service life [15, 16]. The conventional Scan-to-BIM workflow consists of object recognition, segmentation, and reconstruction. These processes are responsible for classifying point clouds with semantic tags, segmenting points based on assigned labels, and reconstructing each recognized object for further processing [17-19]. Thus, accurate and effective pavement extraction is crucial for successful Scan-to-BIM, as it narrows the focus area, reduces complexity, and enhances accuracy. However, it should be noted that almost all existing methods are developed for finished road networks through MLS and their core criteria introduced are not suitable for road construction sites. In general, road extraction methods can be categorized into feature-based, learning-based, and graph-based methods.

2.1 Feature-based Road Extraction

Feature-based methods are widely used in road object recognition and information extraction. These methods can be further divided into region growing, feature extraction, feature fitting, and cross-section analysis. Typically, first three approaches are combined sequentially to extract areas of interest [15]. Zhang et al. [20] used region growing to extract the road surface, followed by intensity variance detection, B-spline and Alpha-shape techniques to create a pavement curve model. Such cooperation can also be found in Yadav et al. [21]. A two-step framework that first extracted local features in each grid and clustered smooth grids using a predefined threshold. Then, a plane fitting step was applied to refine the pavement extraction results. Rodríguez-Cuenca et al. [22] collaborated morphological image processing with grid-based local features analysis to precisely depict the road boundaries. Cross-section analysis methods normally create cross-sections along scanning trajectories or road markings [23, 24]. Pavement extraction is then performed by detecting road curbs or consecutive markings. Yang et al. [23] sliced the datasets along scanning lines and a moving window operator was applied to filter out irrelevant points as well as detecting curbs based on predefined patterns. Similarly, Guan et al. [25] developed algorithms to detect curbs using elevation and slope differences for automated road information extraction. Overall, feature-based methods are effective for identifying pavement boundaries by creating gradient maps, calculating elevation differences or detecting linear patterns, followed by numerical model fitting to generate detailed curve or surface models.

2.2 Learning-based Road Extraction

Like fused approaches in Scan-to-BIM, one or multiple procedures can be replaced by learning-based methods, such as curbs extraction and curve fitting [7, 26, 27]. Yang et al. [28] introduced a binary kernel descriptor (BKD) for encoding as-built point clouds, combined with a random forest classifier to extract curbs and road markings. Similarly, Soilán et al. [29] combined heuristic rules and unsupervised learning models to differentiate between pavement and sidewalks. To acquire better performance, Ma et al. [30] used satellite imagery and a deep learning framework, BoundaryNet, for improving road boundary extraction results. This hybrid method fed curb extraction results into a sequential refining model within BoundaryNet. Wei et al. [31] applied a reinforcement learning (RL) algorithm for curve fitting in complex 3D highway modeling. Road curbs are also commonly used in HD map generation tasks as indicators for pavement extraction [6].

2.3 Graph-based Road Extraction

Graph-based methods are less common as they require supplementary information such as transport network maps. Boyko et al. [32] developed a method using 2D road network maps to divide urban area point cloud into suitable size patches along optimized Cardinal

splines. Then, identifying road networks through local elevation changes in patches. In contrast, Ma et al. [33] applied PointNet++ for initiating a rough road surface extraction and refined results using a two-step graph-based method.

3 Methodology

In this section, the proposed framework for extracting road network construction sites will be elaborated in detail. Compared with the subsequent parametric modeling in Scan-to-BIM for road construction, target area extraction is more challenging due to the complex site conditions and the ALS dataset utilized. This framework consists of four main steps: boundary grids candidate extraction, irrelevant grids removal, road patch generation and connection, and boundary grids refinement (Figure 1). The core idea lies in the road patch generation and connection, which utilizes cross-sectional properties to connect patches and form large network clusters. This step effectively extract the target area with different progress and layouts from its surroundings.

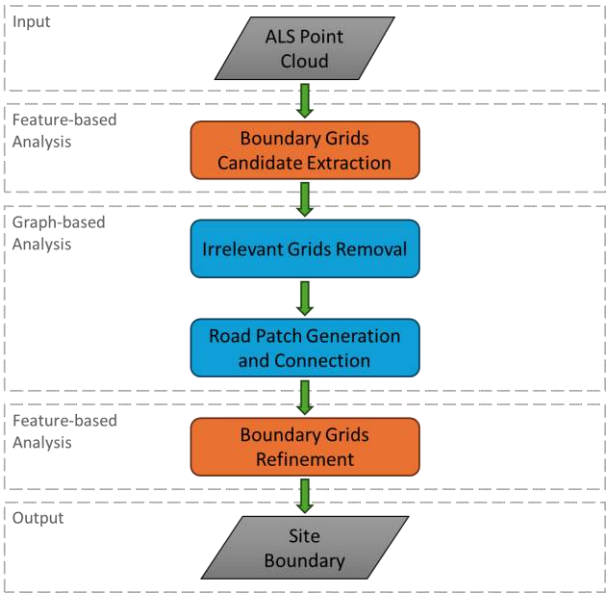


Figure 1. Workflow of the proposed road network construction site extraction framework.

3.1 Boundary Grids Candidate Extraction

Conventional methods for directly extracting ground surface points, such as those based on smoothness or edge detection, are not suitable for road networks under construction captured by ALS. Two main technical challenges identified are monotonous roughness and linear patterns across the entire field. In road construction sites, there is initially no significant roughness difference between the base course and its surrounding areas. Furthermore, unique wave-like patterns in ALS dataset complicates edge detection methods that rely on point distribution analysis. Its uniform distribution causes common local features, such as linearity, planarity, and omni-variance, to yield similar numerical values for areas with and without edges. As a result, instead of directly extracting the ground surface or applying edge detection, we opted to roughly extract boundary candidates using a gradient change ratio map (GCR).

The GCR map is a variant of the gradient map, designed to be more effective for large-scale areas. Initially, the x-y plane is divided into square-shaped grid cells with an empirical edge length of 1.5 m [34, 35]. The GCR is then calculated using the formula shown in Equation (1).

$$GCR = \frac{h_a - h_b}{D_{ab}} \quad (1)$$

where a and b represent the highest and lowest points within the same grid, respectively. Here, h_a and h_b are the elevations of the two points, and D_{ab} is the distance between the two points on the x-y plane. Equation (1) thus computes the gradient change ratio on the plane for each grid. Grids with GCR values exceeding $\mu + \alpha \cdot \sigma$ are defined as boundary candidates and prepared for irrelevant point removal. As an empirical parameter, the factor α is set to 1 for all cases.

3.2 Irrelevant Grids Removal

Irrelevant grid removal aims to eliminate grids belonging to hills and stockpiles that form large-scale nonlinear continuous slopes. This process uses the centroid of each grid as a node and creates a graph by connecting these nodes with edges. For a given group of candidate grids, if their width is more than twice the grid size, the grids in inner layer will have a DoF higher than 6, indicating that at least six grids are adjacent to them. To locate all irrelevant grids, those with a DoF of 7 or 8 are removed first. In the second step, grids with a reduced DoF after the initial removal are also eliminated to completely remove irrelevant areas. Following this, a region growing with an elevation criterion is applied to the remaining grids to cluster adjacent grids together. However, during this removal process, some grids containing boundaries of road construction sites may also be removed, as they are often adjacent to hills or trenches. These grids can still be retrieved through interpolation between the two ends or by fusing the grids with edge detection results. However, this retrieval process falls outside the scope of this paper.

3.3 Road Patch Generation and Connection

After removing irrelevant grids and the subsequent

region-growing process, the adjacent grids are grouped into numerous clusters with either linear or random shapes. To precisely and automatically extract the boundaries of the construction site and establish connectivity between road segments under construction, road patch generation and connection are developed.

Road patch generation consists of two main steps: boundary pairing and centroid computation. In the boundary pairing step, a pair search is conducted within a given radius for each boundary, and all nearby boundaries with a certain proportion of grids within the range are stored. In the centroid computation step, the centroids of paired grids between two paired boundaries are calculated on the x-y plane. The patch is then generated based on the resulting centroids on two ends. The introduction of centroid computation serves two purposes: (1) to provide centroids for locating the two ends of the patches, and (2) to ensure that the patch is formed in a regular shape, even when the two paired boundaries have different lengths. For each grid, it should be noted that centroids are computed between the grid itself and its closest grid, as well as the neighbors of the closest grid in the paired boundary. This approach ensures a relatively smooth transition when curves are represented by staggered grids.

The key idea behind patch connection is similar to the region-growing algorithm. The seed and its neighboring properties are defined as the cross-sections at the ends of each patch. To handle complex road layouts, such as roundabouts, T-junctions, crossroads, and traffic islands, loose criteria are introduced to ensure flexibility. The cross-section is defined by three points: two paired grid centroids and their calculated centroid. The centroids of the two grids use the highest values on the z-axis within their respective grids as their elevation, while the calculated centroid's elevation is derived from the average lowest z-axis values of the two grids. By defining these three points for each cross-section, a triangular regular network (TRN) is established for each patch. This structure allows for easy analysis of the properties at the ends of patches. To evaluate connectivity, two criteria are applied: the aspect ratio (AR) and elevation. The aspect ratio is derived from the difference between the average elevation of two paired grids, $h_{A.max}$, and the elevation of the computed centroid, $h_{A.min}$, divided by the distance between the two centroids of the two paired grids, G_1 and $G_1{}'$ on the x-y plane as shown in Equation (2).

Proximity checking is a prerequisite for connectivity inspection. If the cross-section at one end of a patch is close to that of another, connectivity inspection is applied to determine their relationship. As shown in Equation (3), two road patches are grouped into the same cluster when both the AR and the elevation deviation are smaller than the respective tolerances, δ_{ar} and δ_e. The top ten largest clusters are then output and prepared for refinement.

$$AR = \frac{h_{A.max} - h_{A.min}}{\|G_1 - G_1{}'\|} \quad (2)$$

$$\frac{|AR_i - AR_j|}{min(AR_i, AR_j)} \leq \delta_{ar} \quad (3)$$

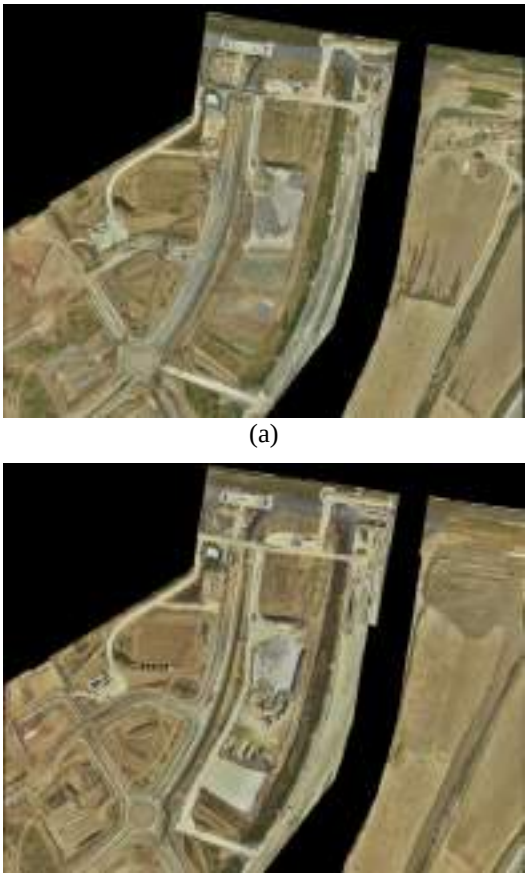
3.4 Boundary Clusters Refinement

Although road patch connections can form a series of large clusters and distinguish them from surrounding neighbors, the main components of some top-ranked clusters may still belong to fragments of irrelevant areas. To address this issue, cluster refinement is introduced to eliminate unwanted large clusters by analyzing the distribution of grid centroids at different resolutions (grid sizes). As the grid size increases, more centroids are included in each grid, resulting in a higher proportion of grids with nonlinear centroids inside. Two indicators, scale-invariant anisotropy and surface area, are used for refinement. Scale-invariant anisotropy ranges from 0 to 1 and increases with the level of randomness in the distribution, while surface area is calculated using a convex hull algorithm, where smaller surface areas indicate more linear distributions. By applying cross-inspection with these two criteria, irrelevant clusters can be effectively identified and removed.

4 Case Study

In this case study, a highway construction site in New South Wales, Australia was selected as the target. The selected area is approximately 1 km by 1 km, and the raw datasets were processed using commercial software to remove non-ground points. The site features complex layouts, including roundabouts, T-junctions, crossroads, and traffic islands. Additionally, due to weather conditions, irrelevant patterns such as truck ruts, mud, and ponding are visible in aerial images. Surrounding the site, temporary roads, hills, and a series of large-scale stockpiles further complicate the extraction of the construction site.

Two datasets were collected on 24 February 2023 and 18 March 2023 to monitor the earthwork at the construction site. As shown in Figure 2b, additional traffic lanes have been excavated on the top-left and bottom-right corners of the roundabout during the past 3.5 weeks. Furthermore, the newly added segments have varying widths, ranging from 8 meters to 15 meters. This information can be used as a reference for determining the search radius in boundary pairing. The ground truth (GT) was generated by manually segmenting the results of irrelevant grid removal, providing a clearer boundary for comparison.



(a) (b)

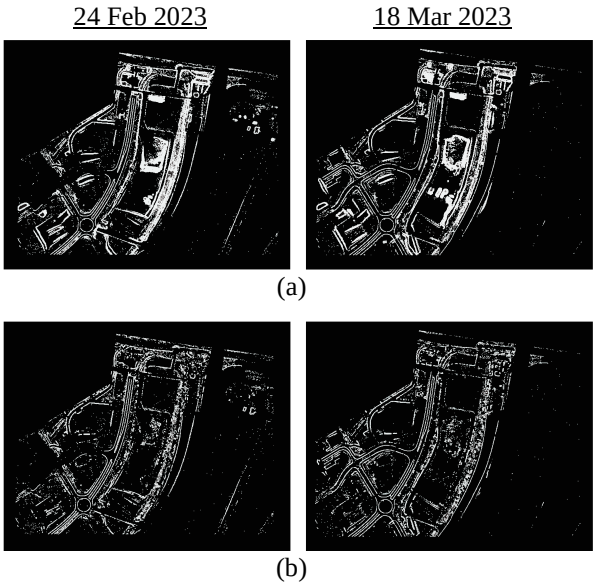
Figure 2. Aerial images of two datasets: (a) captured on 24 February 2023; (b) captured on 18 March 2023.

4.1 Boundary Grids Extraction

To extract boundary candidates, we applied the GCR to each grid to identify candidates with large elevation changes exceeding the sum of the average and standard deviation of GCR. As shown in Figure 3a, the resulting candidates contain hills and stockpiles located around the construction site. After applying irrelevant grid removal, the grids with nonlinear continuous slopes in these areas are removed. Only a few fragments remain in the datasets, as shown in Figure 3b. This process significantly reduces the number of grids introduced into subsequent road patch processing. Moreover, it prevents irrelevant grids from forming patches with other site boundaries, which could otherwise degrade the extraction results.

4.2 Road Patch Processing and Refinement

During the boundary pairing process, the search radius is set to 20 meters, which is slightly larger than the widest part of the entire road network. In this case, if one boundary has at least half of its grids located within this distance from another boundary, they are marked as paired boundaries. Road patch generation is then conducted based on these boundary pairs. Once the generation is completed, proximity and connectivity checks are initiated. The proximity radius threshold R is set to 10 meters, which is half the search radius used in the boundary pairing process. The tolerance for elevation (δ_e) and aspect ratio (δ_{ar}) are set to 10% and 30%, respectively. During the validation process, we observed that δ_{ar} values ranging from 30% to 50% produced similar results. However, due to the presence of T-junctions, where the width of certain areas changes dramatically, δ_{ar} cannot be set below 30%. Lower values lead to failure of grid extractions in those areas. In contrast, high δ_{ar} will include surrounding areas with dissimilar cross-sectional properties, which hinders the subsequent patch processing.



(a) (b)

Figure 3. Results of boundary grids extraction: (a) Candidate grids on 24 February and 18 March 2023; (b) Irrelevant grid removal results on 24 February and 18 March 2023.

As shown in Figure 4a, several irrelevant red areas are extracted due to their large size. However, these clusters can be easily classified by inspecting the trends of local feature changes as the resolution decreases, given that the fragments and short boundaries are distributed randomly. During this process, the upper and lower bounds for the grid size are set at 7.5 meters and 4.5 meters, respectively. This ensures that the new grid contains at least three centroids of the old 1.5 meters grids without including multiple boundaries belonging to the roads. For each cluster, the number of grids with anisotropy and surface area exceeding the average values will increase due to the higher level of randomness introduced during the change of resolution. as the grid size approaches the upper bound,

the proportion of grids with exceeded values in irrelevant clusters will first surpass 50%. By utilizing this principle, nearly all irrelevant red clusters are successfully removed, as shown in Figure 4b.

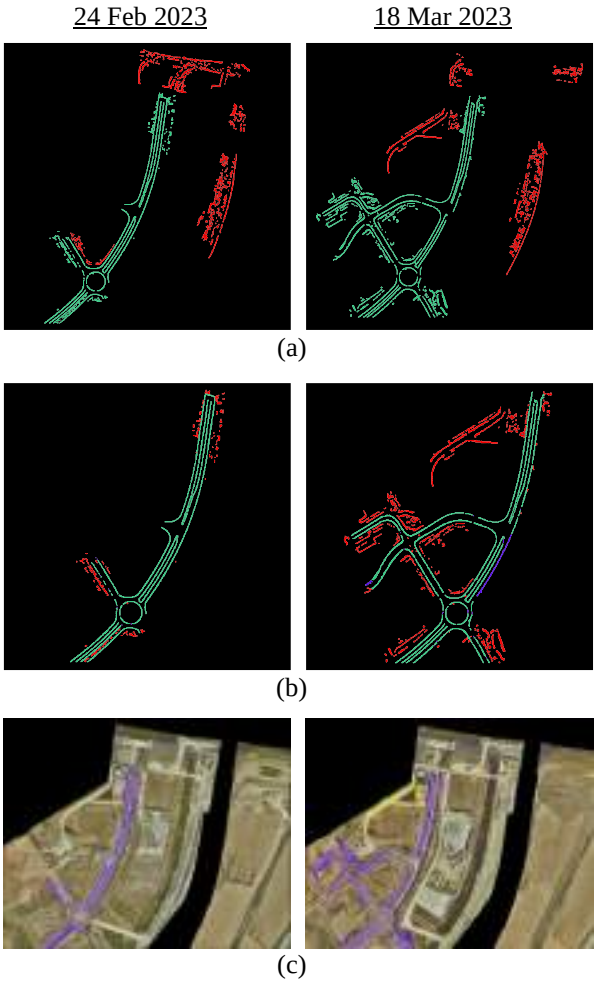


Figure 4. Results of road construction site boundary extraction: (a) Top 10 largest clusters on 24 Feb and 18 Mar 2023; (b) Correct boundaries (green) with missing parts (blue) and excess parts (red) on 24 Feb and 18 Mar 2023; and (c) Overlaid site boundaries on aerial images on 24 Feb and 18 Mar 2023.

The performance evaluation was conducted by calculating the overlap ratio of grids between the extraction results and the GT. As shown in Table 1, the accuracy of the proposed framework for the two datasets is 99.59% and 94.50%, respectively. The deviation is mainly caused by the rearrangement of stockpiles adjacent to the missing boundary, highlighted in blue in Figure 4b on 18 March. This rearrangement results in more fragments of stockpiles remaining near the missing boundary, which affects patch generation and connection. Additionally, the excess ratio for the second dataset is significantly higher than that of the first, at approximately 58.32%. This is clearly visible in Figure 4b on 18 March, where a curved temporary road is highlighted in red. Over the past 3.5 weeks, a series of road barriers have been installed along this temporary road, providing distinctive geometric features that help generate a more complete road boundary. Due to its clear linear pattern, the temporary road was retained during refinement. In Figure 4c, the extracted site boundaries are overlaid on aerial images. The grids align precisely along the boundaries of the road network construction site and can be updated to reflect changes in the construction process. This demonstrates that the proposed method can reliably provide the target area for the subsequent Scan-to-BIM process.

Table 1. Results of road construction site extraction with accuracy (%) and excess ratio (%).

Number of Grids	*24-Feb*	*18-Mar*
Site Boundary	2,466	5,008
Ground Truth	1,944	2,781
Overlapped	1,936	2,628
Accuracy (%)	99.59	94.50
Excess	530	2,380
Excess Ratio (%)	27.26	85.58

5 Conclusions

This paper presented an automated site extraction framework to support Scan-to-BIM progress monitoring in road construction. By leveraging the advantages of local feature analysis and graph theory, this hybrid approach successfully segments the target area from its surroundings. The application of GCR candidate extraction reduces computational requirements and addresses the challenges posed by complex site conditions, which conventional pavement extraction methods struggle to handle. The irrelevant grids removal process effectively fragments grids of nonlinear slopes into small clusters, which can then be easily eliminated. Patch generation and connection provide a feasible solution for segmenting the construction site from the surrounding area with similar cross-sectional features. The criteria utilized are well-suited for handling complex road shapes and layouts. After cluster refinement, the framework achieves a minimum accuracy of 94.50%. The results indicate that the proposed method can be applied to datasets with different road layouts and construction progress, highlighting its generality and potential for datasets from other projects.

Some limitations remain in the proposed framework.

The centroid generation process lacks stability when curves cause multiple grids to select the same closest grid in the paired boundary during centroids computation. Moreover, to ensure a clear end edge when paired boundaries have different lengths, we apply a threshold to handle this case. However, both strategies may result in the omission of centroids in certain specific positions. Thus, for future work, two modifications and a missing parts retrieve step can be incorporated into the framework to further improve its performance. These modifications aim to enhance the accuracy of boundary pairing and centroid computation during road patch generation, as these steps significantly impact overall accuracy. Boundary searching could introduce more multiple paring criteria, rather than relying solely on distance. Centroid calculation could also be refined by incorporating predefined shapes to enhance the smoothness and continuity of centroid transitions. The missing grids retrieve process could be achieved by overlapping boundary curves derived from elevation scalar fields onto the extraction results. Since the target areas have already been identified, the focus would shift to curves that connect extracted grids located far apart. By inspecting grids along these focused curves, missing grids could be effectively retrieved.

Acknowledgement

This research was supported by funding from the Australian Research Council (ARC) under Grant number IH210100048 ARC Research Hub for Resilient and Intelligent Infrastructure Systems (RIIS). We would also like to express our gratitude to GeoAI for their support and collaboration.

References

[1] Y. Chong Heap, R. Lopez, J. Wang, X. Wang, and Z. Zhao, *Comparative Analysis on the Adoption and Use of BIM in Road Infrastructure Projects.* Journal of Management in Engineering, 2016. **32**(6): p. 05016021.

[2] A. Costin, A. Adibfar, H. Hu, and S.S. Chen, *Building Information Modeling (BIM) for transportation infrastructure – Literature review, applications, challenges, and recommendations.* Automation in Construction, 2018. **94**: p. 257-281.

[3] T. Patel, B.H.W. Guo, and Y. Zou, *A scientometric review of construction progress monitoring studies.* Engineering, Construction and Architectural Management, 2021. **ahead-of-print**(ahead-of-print).

[4] Z. Jiang, X. Shen, M. Ibrahimkhil, K. Barati, and J. Linke. *Scan-VS-BIM for real-time progress monitoring of bridge construction project*. in *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 2022.

[5] M.H. Ibrahimkhil, X. Shen, K. Barati, and C.C. Wang *Dynamic Progress Monitoring of Masonry Construction through Mobile SLAM Mapping and As-Built Modeling*. Buildings, 2023. **13**, DOI: 10.3390/buildings13040930.

[6] C. Ye, H. Zhao, L. Ma, H. Jiang, H. Li, R. Wang, M.A. Chapman, J.M. Junior, and J. Li, *Robust Lane Extraction From MLS Point Clouds Towards HD Maps Especially in Curve Road.* IEEE Transactions on Intelligent Transportation Systems, 2022. **23**(2): p. 1505-1518.

[7] B. Suleymanoglu, M. Soycan, and C. Toth *3D Road Boundary Extraction Based on Machine Learning Strategy Using LiDAR and Image-Derived MMS Point Clouds*. Sensors, 2024. **24**, DOI: 10.3390/s24020503.

[8] K. Karila, L. Matikainen, E. Puttonen, and J. Hyyppä, *Feasibility of Multispectral Airborne Laser Scanning Data for Road Mapping.* IEEE Geoscience and Remote Sensing Letters, 2017. **14**(3): p. 294-298.

[9] E. Che, J. Jung, and M.J. Olsen, *Object Recognition, Segmentation, and Classification of Mobile Laser Scanning Point Clouds: A State of the Art Review.* Sensors, 2019. **19**(4).

[10] K. Mirzaei, M. Arashpour, E. Asadi, H. Masoumi, Y. Bai, and A. Behnood, *3D point cloud data processing with machine learning for construction and infrastructure applications: A comprehensive review.* Advanced Engineering Informatics, 2022. **51**: p. 101501.

[11] Y. Xu, X. Shen, S. Lim, and X. Li, *Three-Dimensional Object Detection with Deep Neural Networks for Automatic As-Built Reconstruction.* Journal of Construction Engineering and Management, 2021. **147**(9): p. 04021098.

[12] Y. Guo, X. Shen, J. Linke, Z. Wang, and K. Barati *Quantification of Structural Defects Using Pixel Level Spatial Information from Photogrammetry*. Sensors, 2023. **23**, DOI: 10.3390/s23135878.

[13] R. Wang, J. Peethambaran, and D. Chen, *LiDAR Point Clouds to 3-D Urban Models$:$ A Review.* IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2018. **11**(2): p. 606-627.

[14] I. Kiichiro, T. Jun-ichi, A. Yoshiharu, and H. Takumi. *A Mobile Mapping System for road data capture based on 3D road model*. in *2006 IEEE Conference on Computer Aided Control System Design, 2006 IEEE International Conference on Control Applications, 2006 IEEE International Symposium on Intelligent Control*. 2006.

[15] M. Soilán, A. Sánchez-Rodríguez, P. del Río-Barral,

C. Perez-Collazo, P. Arias, and B. Riveiro *Review of Laser Scanning Technologies and Their Applications for Road and Railway Infrastructure Monitoring*. Infrastructures, 2019. **4**, DOI: 10.3390/infrastructures4040058.

[16] P. Wei, L. Yan, H. Xie, and M. Huang, *Automatic coarse registration of point clouds using plane contour shape descriptor and topological graph voting.* Automation in Construction, 2022. **134**: p. 104055.

[17] N. Abreu, A. Pinto, A. Matos, and M. Pires *Procedural Point Cloud Modelling in Scan-to-BIM and Scan-vs-BIM Applications: A Review*. ISPRS International Journal of Geo-Information, 2023. **12**, DOI: 10.3390/ijgi12070260.

[18] D. Rebolj, Z. Pučko, N.Č. Babič, M. Bizjak, and D. Mongus, *Point cloud quality requirements for Scan-vs-BIM based automated construction progress monitoring.* Automation in Construction, 2017. **84**: p. 323-334.

[19] Q. Wang, J. Guo, and M.-K. Kim *An Application Oriented Scan-to-BIM Framework*. Remote Sensing, 2019. **11**, DOI: 10.3390/rs11030365.

[20] Z. Zhang, J. Li, Y. Guo, C. Yang, and C. Wang, *3D Highway Curve Reconstruction From Mobile Laser Scanning Point Clouds.* IEEE Transactions on Intelligent Transportation Systems, 2020. **21**(11): p. 4762-4772.

[21] M. Yadav and A.K. Singh, *Rural Road Surface Extraction Using Mobile LiDAR Point Cloud Data.* Journal of the Indian Society of Remote Sensing, 2018. **46**(4): p. 531-538.

[22] B. Rodríguez-Cuenca, S. García-Cortés, C. Ordóñez, and M.C. Alonso, *An approach to detect and delineate street curbs from MLS 3D point cloud data.* Automation in Construction, 2015. **51**: p. 103-112.

[23] B. Yang, L. Fang, and J. Li, *Semi-automated extraction and delineation of 3D roads of street scene from mobile laser scanning point clouds.* ISPRS Journal of Photogrammetry and Remote Sensing, 2013. **79**: p. 80-93.

[24] L. Ma, Y. Li, J. Li, C. Wang, R. Wang, and M.A. Chapman *Mobile Laser Scanned Point-Clouds for Road Object Detection and Extraction: A Review*. Remote Sensing, 2018. **10**, DOI: 10.3390/rs10101531.

[25] H. Guan, J. Li, Y. Yu, M. Chapman, and C. Wang, *Automated Road Information Extraction From Mobile Laser Scanning Data.* IEEE Transactions on Intelligent Transportation Systems, 2015. **16**(1): p. 194-205.

[26] N. Wang, Z. Shi, and Z.J.C.J.o.R.S. Zhang, *Road boundary, curb and surface extraction from 3D mobile LiDAR point clouds in urban environment.* 2022. **48**(4): p. 504-519.

[27] J. Balado, J. Martínez-Sánchez, P. Arias, and A. Novo *Road Environment Semantic Segmentation with Deep Learning from MLS Point Cloud Data*. Sensors, 2019. **19**, DOI: 10.3390/s19163466.

[28] B. Yang, Y. Liu, Z. Dong, F. Liang, B. Li, and X. Peng, *3D local feature BKD to extract road information from mobile laser scanning point clouds.* ISPRS Journal of Photogrammetry and Remote Sensing, 2017. **130**: p. 329-343.

[29] M. Soilán, L. Truong-Hong, B. Riveiro, and D. Laefer, *Automatic extraction of road features in urban environments using dense ALS data.* International Journal of Applied Earth Observation and Geoinformation, 2018. **64**: p. 226-236.

[30] L. Ma, Y. Li, J. Li, J.M. Junior, W.N. Gonçalves, and M.A. Chapman, *BoundaryNet: Extraction and Completion of Road Boundaries With Deep Learning Using Mobile Laser Scanning Point Clouds and Satellite Imagery.* IEEE Transactions on Intelligent Transportation Systems, 2022. **23**(6): p. 5638-5654.

[31] Y. Wei, Z. Zhang, X. Huang, Y. Lin, and W. Liu, *3D HIGHWAY CURVE RECONSTRUCTION FROM MOBILE LASER SCANNING POINT CLOUDS THROUGH DEEP REINFORCEMENT LEARNING.* Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., 2023. **XLVIII-1/W2-2023**: p. 55-61.

[32] A. Boyko and T. Funkhouser, *Extracting roads from dense point clouds in large scale urban environment.* ISPRS Journal of Photogrammetry and Remote Sensing, 2011. **66**(6, Supplement): p. S2-S12.

[33] H. Ma, H. Ma, L. Zhang, K. Liu, and W. Luo *Extracting Urban Road Footprints from Airborne LiDAR Point Clouds with PointNet++ and Two-Step Post-Processing*. Remote Sensing, 2022. **14**, DOI: 10.3390/rs14030789.

[34] B. Yang, R. Huang, Z. Dong, Y. Zang, and J. Li, *Two-step adaptive extraction method for ground points and breaklines from lidar point clouds.* ISPRS Journal of Photogrammetry and Remote Sensing, 2016. **119**: p. 373-389.

[35] Y. Zhou, R. Huang, T. Jiang, Z. Dong, and B. Yang, *Highway alignments extraction and 3D modeling from airborne laser scanning point clouds.* International Journal of Applied Earth Observation and Geoinformation, 2021. **102**: p. 102429.

P2P: Point Cloud to Panel Layout Optimization

Nisha Deborah Philips, Yifang Liu, Nolan W. Hayes, Diana Hun, and Bryan P. Maldonado

Buildings and Transportation Science Division, Oak Ridge National Laboratory*, United States of America

philipsn@ornl.gov, liuy5@ornl.gov, hayesnw@ornl.gov, hunde@ornl.gov, maldonadopbp@ornl.gov

Abstract -

Building envelope retrofits, despite their benefits on enhancing energy efficiency, progress slowly due to high operating costs. Overclad panelized systems present an attractive solution to make retrofits affordable and easy to install. However, several stages of the retrofit process remain disconnected, suboptimal, and require significant human intervention. This study aims to bridge the gap between digital twin generation and overclad panel installation by automating the design of an optimal panel layout directly from a building envelope point cloud. In this end-to-end approach, the dimensional twin is generated by segmenting the facade point cloud. The facade then undergoes a three-step process to generate an optimized panel layout for integration with automated placement systems.

Keywords -

Building envelope retrofit, Panel layout design, Constraint satisfaction problem

1 Introduction

Rapid consumption of world energy has raised concerns over the depletion of energy resources and the adverse effects it has on the environment. With the construction industry accounting for 20–40% of the total energy consumption [1], significant effort has been directed to improve the energy efficiency of existing buildings. In response, several government organisations have implemented initiatives to enhance the energy performance of buildings, while extensive research efforts have been made to optimize the energy performance of existing structures. Results have shown that energy consumption can significantly be reduced via retrofitting existing buildings as opposed to constructing new ones [2].

With this surge of interest in retrofitting, leveraging technologies such as Digital Twins, which are digital replicas of physical objects, are imperative to optimize retrofits and improve the energy efficiency of existing buildings [3]. Digital twins have a variety of applications from being integrated with Internet of Things (IoT) for data collection to Artificial Intelligence (AI) for processing the collected data [4]. In this context, digital twins are used to virtually reconstruct a building and generate optimized panel layout designs for each facade through an automated process.

Panel layout generation from architectural drawings falls into two categories: "heavy" panel layouts for facades with openings (windows, doors) and "light" panel layouts for facades without such features [5]. Heavy solutions use prefabricated panels, while light solutions require on-site panel rework. Despite these differences, both types can utilize the same optimization algorithm. Furthermore, automated panel layout tools should be integrated with existing technologies for digital twin generation like the Automatic point Cloud Building Envelope Segmentation (Auto-CuBES) [6] and for automated installation like the Real-Time Evaluator (RTE) [7, 8] to create a unified retrofit framework.

The panel layout designer could either be performed manually by an architect or automated via AI. The manual panel layout design is inherently labor intensive and requires substantial amount of man-hours. Given these challenges, advancements in computational methods have paved the way for automating the panel layout design process using AI. By transforming the problem into a mathematical optimization framework, such as the bin packing problem [9], it becomes possible to model the allocation of panels as a computationally solvable task. The bin packing problem, a well-known combinatorial optimization challenge, involves efficiently packing objects of varying sizes into fixed-capacity bins, minimizing wasted space. For panel layout design, this approach aims to maximize panel coverage while adhering to the architectural constraints.

In this paper, we propose modeling panel layout design as a *meta-rectangle packing problem*, extending traditional rectangle packing [10] to non-uniform bins. This mathematical framing allows us to use AI-driven algorithms, specifically Constraint Satisfaction Problem (CSP), to optimize panel packing. Section 2 describes existing and proposed panel layout design algorithms. Section 3 presents the implementation and analyzes execution time and coverage. Section 4 concludes and outlines future work.

*Notice: This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (https://www.energy.gov/doe-public-access-plan).

2 Optimal panel layout design

2.1 Digital twin generation

Building envelope retrofit extensively uses point cloud data to generate as-built dimensions of the building. However, the generation of digital twins from point cloud data has traditionally been a manual process, making it laborious and time-consuming. To address this limitation, machine learning methods, such as the Automatic point Cloud Building Envelope Segmentation (Auto-CuBES) [6] algorithm, can automate the generation of digital twin from point cloud data.

Auto-CuBES divides the algorithm into sequential tasks that generate a digital twin from building point cloud data of 3mm density. Despite its computational efficiency and accuracy, the input parameters scale linearly with features, requiring considerable time for parameter calibration. To address this bottleneck, the authors developed Auto-CuBES++, a digital tool providing real-time parameter calibration for building facade segmentation, available for academic or commercial use [11]. This tool assists users at every algorithm stage by visualizing results and allowing parameter adjustments in real time without full re-execution. The software also preserves execution states, ensuring input parameters are retained and eliminating the need for re-entry when rerunning the algorithm.

In addition to preserving the state of the run, Auto-CuBES++ also stores the dimensions as a JSON file containing the coordinates of the facade and its associated openings. These JSON files facilitates validating the digital twin's accuracy in virtualizing the building. The coordinates in the JSON file are represented as tuples of real numbers (x, y, z), which are used to map the facade to a Cartesian coordinate system. This representation provides a structured framework for panelizing the facade and seamlessly integrating it with the automated installation systems, such as the RTE.

2.2 Bin packing problems

Panelization, the key theme of this paper, is the process of prefabricating building's structural elements off-site and assembling a panelized system to minimize on-site labor. In a traditional panelized construction of a wall, the wall is divided into a simple layout and the pre-manufactured panels are assembled accordingly. However, traditional techniques of panel layout design are not as straight forward and often require human intervention to optimize the generated layout design [5].

The panel layout design problem can be represented as a *Bin packing problem* [9], and more specifically as a *Rectangle Packing Problem* [12] since we focus on single-wall arrangements. Rectangle Packing is an extension of Bin packing that uses a single bin instead of multiple bins.

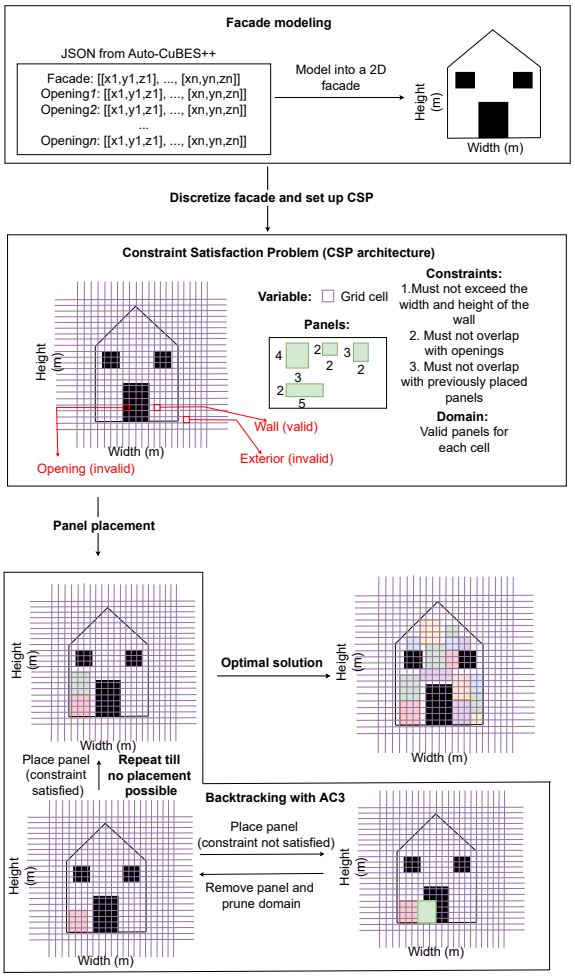


Figure 1. Facade to panel layout

As Bin packing problems are NP-hard, Rectangle Packing Problems inherit this computational complexity and are also NP-complete [10].

These Rectangle Packing Problems can be solved using deterministic algorithms, which use variables and policies to explore the solution space [10], and non-deterministic algorithms, such as Genetic Algorithm (GA) [13] and Simulated Annealing [14]. While deterministic approaches rely on predefined policies and constraints, non-deterministic methods leverage relationships between variables as data structures to search for optimal solutions.

2.3 Proposed panel layout design algorithm

This paper addresses panel packing using a deterministic algorithm, specifically the Constraint Satisfaction Problem (CSP). It combines backtracking [15] and arc consistency (AC-3) [16] to efficiently explore the solution space. Backtracking methodically scans solutions, reverting decisions that violate constraints, while AC-3 prunes infeasible

values to simplify the problem. Figure 1 illustrates this hybrid process, comprising three stages: 1) Facade modeling, 2) CSP formulation, and 3) solving via backtracking.

2.3.1 Facade modeling

Initially, the algorithm processes the JSON files generated by Auto-CuBES++ and extracts the geometric data required for facade modeling. Auto-CuBES++ generates two types of JSON files, one comprising of the facade boundary points, and another set of rectangular prism files for each individual opening in the facade. The algorithm transforms the extracted geometric data from the 3D space to a 2D representation based on the following rules:

- Points (x, z) for a facade parallel to the x and z axes
- Points (y, z) for a facade parallel to the y and z axes

2.3.2 CSP formulation

Following the facade modeling, the CSP problem for meta-rectangle packing is formulated. In a traditional rectangle packing algorithm, a set of rectangular items are packed into a rectangular box. However, due to the algorithm being applied to buildings, the rectangular box is adapted to a polygonal facade with or without openings. Additionally, the rectangular items are customized to panels of varying sizes that are packed on the wall.

Every CSP architecture is uniquely identified by a set of variables, the domains associated with the variables and a set of constraints. In the initial stage, the algorithm assumes a simplified scenario where the grid represents a facade without any openings (light panel layout design). In this case, each cell in the grid is valid, and the algorithm proceeds by assigning panels from the available domain to each of these cells. The process involves evaluating the constraints to ensure that the panels meet the architectural requirement of not aligning with the openings.

Once this basic configuration is established, the algorithm is adapted to handle the more complex scenario involving openings (heavy panel layout design). Openings, such as windows or doors, alter the layout of the facade, and certain cells are no longer suitable for panel placement. Each cell can fall under one of the categories shown in Table 1. Based on this information, the grid is refined to include only the valid cells, more specifically the wall points, so cells falling under any of the other categories are marked invalid and a panel is not placed on those cells.

The algorithm generates solutions that satisfy a set of panel placement constraints, as shown in Algorithm 1. Firstly, IS_WITHIN_BOUNDS() verifies that panels are placed only on valid cells within the grid. Secondly, DOES_NOT_OVERLAP() ensures that newly placed panels do not overlap with existing panels or openings. Finally, for light panel layouts, DOES_NOT_ALIGN() checks that panel edges do not align with opening edges to prevent potential air/heat leakage points.

Table 1. Grid cells vs Validity

Type of grid cell	Validity
Only wall	Valid
Only opening	Invalid
Only exterior	Invalid
Wall and opening	Invalid
Wall and exterior	Invalid
Opening and exterior	Invalid

Algorithm 1 Valid panel placement check

Require: Wall, Cell (x, y), Panel
Ensure: Panel satisfies or violates constraints
1: **Function** CONSTRAINTS(wall, x, y, panel):
2: **for** each cell in panel.area on wall **do**
3: **if** IS_WITHIN_BOUNDS(x, y, PANEL) **AND** DOES_NOT_OVERLAP(x, y, PANEL)) **AND** DOES_NOT_ALIGN_WITH_WINDOW(x, y, PANEL)) **then**
4: **return** True
5: **else**
6: **return** False
7: **end if**
8: **end for**

2.3.3 Backtracking with Arc Consistency

Following the initialization of the grid and the CSP architecture, Backtracking, one of the most commonly used algorithms, was employed to solve the problem. This algorithm exhaustively searches the solution space to find assignments or, in this case, panel placements that satisfy the constraints. However, due to the large search spaces involved in most CSP problems, the algorithm could get computationally expensive. Therefore, to improve the efficiency of the algorithm, backtracking is integrated with AC-3, thereby reducing the complexity of the problem by pruning invalid panel assignments. This hybrid approach is well suited for problems where constraints are heavily co-dependent, as is the case for the panel layout design optimization problem.

As shown in Algorithm 2, Backtracking implements a Depth First Search (DFS) to explore possible solutions. Given grid cells, panel types, and placement rules, the algorithm traverses each cell and attempts to place the largest valid panel from the cell's domain. If the placement satisfies all constraints, the PLACEPANEL() function marks the cells as occupied. If invalid, REMOVEPANEL() reverts the placement and the algorithm backtracks to try the next panel from the domain.

While backtracking ensures completeness, it's limited by exponential growth in the solution space. To improve

Algorithm 2 Backtracking for meta-rectangle packing

Require: Wall dimensions, panels, openings
Ensure: Maximum coverage with minimum panels
Initialize valid wall grid and domains
Apply AC-3 to enforce initial arc consistency
Call BACKTRACK(wall, panels, $x = 0$, $y = 0$, panel_id=1)

Function BACKTRACK(wall, panels, x, y, panel_id):
if $y \geq$ wall.height **then**
  Update best grid and return
end if
if $x \geq$ wall.width **then**
  BACKTRACK(wall, panels, $x = 0$, $y + 1$, panel_id)
  return
end if
if wall.current_grid[y][x] $\neq 0$ **or** $(x, y) \in$ wall.openings **then**
  BACKTRACK(wall, panels, $x + 1$, y, panel_id)
  return
end if
for panel $\in$ SORTBYSIZE(wall.domains[(y, x)]) **do**
  if CONSTRAINTS(wall, x, y, panel) **then**
    PLACEPANEL(wall, x, y, panel, panel_id)
    Apply AC-3 for arc consistency
    BACKTRACK(wall, panels, x+ panel.width, y, panel_id+1)
    REMOVEPANEL(wall, x, y, panel)
  end if
end for
BACKTRACK(wall, panels, $x + 1$, y, panel_id)

Function PLACEPANEL(wall, x, y, panel, panel_id):
for each cell in panel.area **do**
  wall.current_grid[cell] $\leftarrow$ panel_id
end for
wall.panel_count $\leftarrow$ wall.panel_count + 1

Function REMOVEPANEL(wall, x, y, panel):
for each cell in panel.area **do**
  wall.current_grid[cell] $\leftarrow 0$
end for
wall.panel_count $\leftarrow$ wall.panel_count - 1

Algorithm 3 Arc Consistency (AC-3)

Initialize queue with all cells from the grid
while queue is not empty **do**
  $(x, y) \leftarrow$ dequeue(queue)
  if REVISE(wall, x, y) **then**
    for neighbor $\in$ GETNEIGHBORS(x, y) **do**
      Enqueue(neighbor)
    end for
  end if
end while

Function REVISE(wall, x, y):
revised $\leftarrow$ false
for panel $\in$ wall.domains[(y, x)] **do**
  if not CONSTRAINTS(wall, x, y, panel) **then**
    Remove panel from wall.domains[(y, x)]
    revised $\leftarrow$ true
  end if
end for
return revised

efficiency, we implemented AC-3 pre-processing to prune inconsistent panels from a cell's domain after successful placement. As shown in Algorithm 3, AC-3 initializes a queue with all grid cells and iteratively prunes the domains of each cell and its neighbors, removing invalid panel configurations. This maintains local consistency and reduces the search space size.

2.3.4 Complexity Analysis

The backtracking approach remains exponential in complexity as the number of valid grid cells increases. AC-3 improves efficiency by dynamically reducing domain sizes, but its impact depends on constraint density. Two key factors influence performance:

Facade Shape Complexity: Complex geometries increase both the search space and constraint network density, resulting in higher AC-3 processing overhead.

Panel Size Variety: A larger panel dimension set increases domain sizes, creating more potential assignments for backtracking. Though AC-3 prunes infeasible configurations, it cannot eliminate the problem's inherent exponential worst-case growth.

3 Results

The algorithm begins by identifying and importing the JSON files containing the dimensions for a desired facade and its associated openings. These files are then pre-processed to ensure they are aligned either with the (x, z) plane or the (y, z) plane depending on the orientation of the selected facade. For this study, the algorithm is tested on both a facade aligned with the (x, z) plane and another aligned with the (y, z) plane, as depicted in Figure 2.

Following the pre-processing stage, the corresponding 2D points are extracted from the aligned JSON files. These points are used to map the grid onto a cartesian space without loss of any spatial data. This mapping ensures accurate representation of the facade that would serve as an input to the CSP architecture.

Additionally, due to the complexity added by the openings, the algorithm was run independently for two cases: 1) Facades without openings (light solution), and 2) Facades with openings (heavy solution). For both cases, the wall and opening dimensions in the 2D space were scaled to three different units: millimeters (mm), centimeters (cm), and decimeters (dm). This multi-scale analysis highlights the relationship between grid density and computational time. Notably, as the grid density increases tenfold (e.g., transitioning from cm to mm scales), the execution time also increases approximately tenfold, demonstrating the algorithm's scaling behavior.

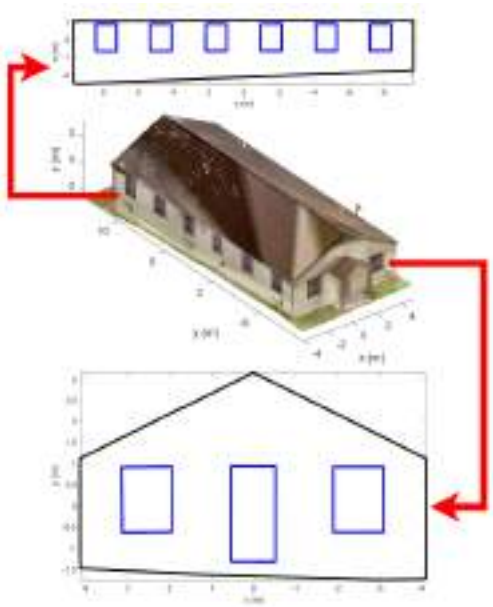


Figure 2. Dimensional facades extracted with Auto-CuBES++ and pre-processed

3.1 Facade without openings

After the initial set up, the inputs for the CSP architecture are set up and initialized. As elaborated in Section 2, the inputs provided to the CSP algorithm are the following:

Grid resolution: determines the scale of each grid cell onto which the facade is mapped. It can take up any value from $\{10, 100, 1000\}$ depending on the dimensions the facade is scaled to: mm, cm, or dm.

Panel sizes: customized values based on the facade under consideration. The height of each panel is fixed to 2.44 m (8 ft), while the width varies in factors of the grid resolution, typically multiples of 1.22 m (4 ft).

Facade grid: is the 2D cartesian representation of the facade generated by Auto-CuBES.

Following the set up of these parameters, the algorithm tackles the problem of mapping the facade to a grid. In order to adapt the rectangle-packing problem to the meta-rectangle packing problem, the area where the panels can be placed has to be clearly defined. This is achieved via the Shapely package in Python [17]. The 2D points extracted from the JSON files generated by Auto-CuBES are mapped onto a rectangular grid mimicking a rectangular bin. Due to the algorithm initially tackling facades without any opening, the valid grid is initialized to the entire facade area.

After grid initialization, each cell receives a list of available panels. The algorithm evaluates three grid resolutions: decimeters (dm), centimeters (cm), and millimeters (mm). Three panel sizes (8×4 ft, 8×8 ft, and 8×12 ft) are scaled to match each resolution: 24×12, 24×24, and 24×36 dm; 240×120, 240×240, and 240×360 cm; and 2400×1200, 2400×2400, and 2400×3600 mm.

In due course, the backtracking algorithm is triggered to explore potential solutions for all the three grid resolutions. This process iteratively constructs panel layouts by placing the panel with the largest area from each cell's domain. while also systematically retracting and reevaluating the assigned panel positions to maximize coverage. As this problem was modeled as an optimization problem, the final result that is visualized aims to maximize the coverage of the valid grid in terms of cells while minimizing the number of panels used.

3.2 Facade with openings

The setup for this case builds on Section 3.1 but accounts for openings like windows and doors. The wall geometry is read from a JSON file containing the facade's polygonal representation, and its integrity is ensured through a validation and repair process. Similarly, all opening geometries are dynamically loaded from matching files, each representing an opening in the wall, and undergo the same validation and repair procedure to ensure they are well-formed polygons.

After preparing the wall and opening polygons, the valid area is computed by subtracting the opening polygons from the wall polygon using a geometric difference operation. This process removes areas covered by openings, ensuring panels are not placed over windows or doors. Additionally, cells along the facade's edges are invalidated to prevent panels from extending beyond the wall's dimensions. The resulting valid area excludes both openings and boundary cells, creating an optimized grid for panel placement.

The algorithm is then run for three facades at different grid resolutions, including one from a two-story experimental building at Oak Ridge National Laboratory (ORNL), known as the Flexible Research Platform 2 (FRP2), as illustrated in Fig.3. The other two facades are from a single-story residential building being retrofitted by the Knoxville's Community Development Corporation (KCDC). The two facades chosen from the single-story building cover the cases of facades being aligned with the $x-z$ plane (KCDC 21) and the $y-z$ plane (KCDC 11), as illustrated in Fig.4 and Fig.5, respectively. In these figures, part (a) shows the panel placement with openings, while part (b) assumes no openings, with panels placed to fully cover any openings. These facades were converted into all three grid resolutions being tested and evaluated based on the time taken for initialization, valid grid computation,

Table 2. Result: Panel placement execution time

Dimension	Initializing (sec)	Valid grid computation (sec)	Backtracking + AC-3 (hrs)	Coverage (%)	Number of panels used
FRP2 (with opening)				Excluding openings	
Decimeter (dm)	1.87	0.44	2.38	60.51	10
Centimeter (cm)	69.89	36.33	17.31	54.27	10
Millimeter (mm)	2805.06	2595.82	>30	~55	10
FRP2 (without opening)					
Decimeter (dm)	2.05	1.03	1.4	76.24	12
Centimeter (cm)	13.96	36.65	10.89	75.22	15
Millimeter (mm)	479.21	3392.06	>30	~ 75	15
KCDC 11 (with opening)				Excluding openings	
Decimeter (dm)	0.20	0.14	0.01	25.2	2
Centimeter (cm)	8.84	9.47	0.48	23.06	2
Millimeter (mm)	822.82	1111.53	28.11	22.87	2
KCDC 11 (without opening)					
Decimeter (dm)	0.55	0.51	0.63	57.44	3
Centimeter (cm)	7.26	11.52	9.33	55.54	3
Millimeter (mm)	533.11	1048.32	>30	~ 55	3
KCDC 21 (with opening)				Excluding openings	
Decimeter (dm)	0.71	0.32	0.02	43.8	7
Centimeter (cm)	24.29	29.52	0.76	39.76	7
Millimeter (mm)	1948.51	1929.85	31.2	39.48	7
KCDC 21 (without opening)					
Decimeter (dm)	1.42	0.90	0.22	77.77	6
Centimeter (cm)	18.93	38.09	2.93	69.81	6
Millimeter (mm)	867.78	1806.78	>30	~ 70	6

and the execution of the backtracking algorithm with AC-3, as shown in Table 2. The results in Table 2 correspond to the panel placements shown in Fig. 3, Fig. 4, and Fig. 5, with each row in the table corresponding to a specific facade and scenario. The first set of rows corresponds to the FRP2 facade, shown in Fig. 3. The second set corresponds to the KCDC 11 facade, shown in Fig. 4. And the final set corresponds to the KCDC 21 facade, shown in Fig. 5. Each row in the table also presents the execution times, coverage percentages, and number of panels used for the different grid resolutions, reflecting the specific conditions for each facade in the corresponding figures.

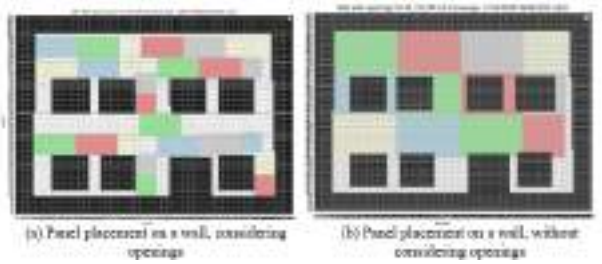


Figure 3. Panelized FRP2 with and without openings

The backtracking algorithm evaluates each panel in every cell's domain, leading to exponential complexity as grid resolution increases. At millimeter resolution, both initialization and valid grid computation times rise significantly compared to decimeter and centimeter scales, sometimes exceeding 30 hours even with AC-3 as shown in Table 2. To mitigate this, a 20% sampling rate is applied for constraint satisfaction, balancing efficiency and accuracy—higher rates caused excessive panel overlap, while lower rates added unnecessary overhead without notable accuracy gains.

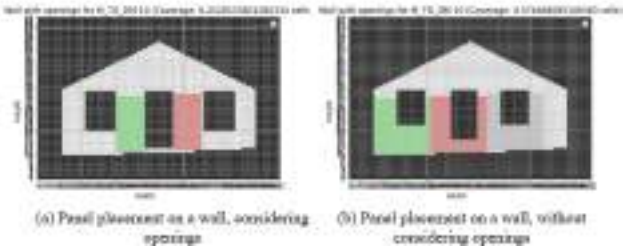


Figure 4. Panelized KCDC 21 with and without openings

The results from the algorithm were evaluated by focusing on two key attributes: the optimization objectives, such as total area coverage, the number of panels used, as outlined in Table 2 and the execution times, namely the time taken to initialize, compute the valid grid, and

execute the algorithm.

Coverage is the primary objective of the algorithm, with the aim of maximizing the valid area of the grid that corresponds exclusively to the facade. The algorithm ensures that as much of the facade as possible is covered effectively, while excluding non-relevant areas such as the attic as shown in Fig. 4, where wall insulation is not required. Additionally, to improve efficiency, the algorithm was designed to minimize the number of panels used. This dual-objective optimization assessment provides a comprehensive evaluation of the algorithm's capability in balancing these competing goals.

In Figures 3, 4, and 5 coverage is assessed with and without considering facade openings. In part (a), panels avoid overlapping openings based on constraint enforcement (Algorithm 1), resulting in more panels and gaps around these regions. In part (b), openings are ignored, leading to fewer panels and full facade coverage, including areas originally designated as openings.

This experiment revealed that maximizing coverage becomes notably more difficult when openings, such as windows and doors, are present. These openings create irregularities and constraints on the facade, which often result in uncovered regions due to the limitations of panel dimensions and placement options. This challenge is especially pronounced for facades with irregular shapes, which require the algorithm to navigate a more complex solution space. As a result, compromises must be made in either coverage or panel efficiency, as illustrated in the left plots of Fig.3 and Fig.4. These plots clearly show the trade-offs the algorithm faces when balancing the competing goals of maximizing coverage while minimizing the number of panels used, especially in the presence of openings and irregular facade shapes.

A key challenge in the results is the unoccupied panel areas, caused by placement restrictions around openings and fixed panel dimensions. While the algorithm optimizes coverage, irregular facades and panel limitations prevent perfect coverage. Future improvements could include more flexible panel shapes, such as modular panels, or allowing for smaller panel sizes that could better adapt to the irregularities of the facade.

Execution time, as a secondary evaluation metric, scales predictably with grid resolution. As the density of the grid increases, the execution time also increases, as shown in Table 2, reflecting the greater computational effort required to evaluate the finer cells of the grid. This trend highlights the increased computational burden when dealing with higher-resolution grids. In some millimeter-scale cases, the combination of higher grid granularity and the algorithm's complexity has led to execution times exceeding 30 hours, emphasizing the significant computational challenges posed by high-resolution analyses.

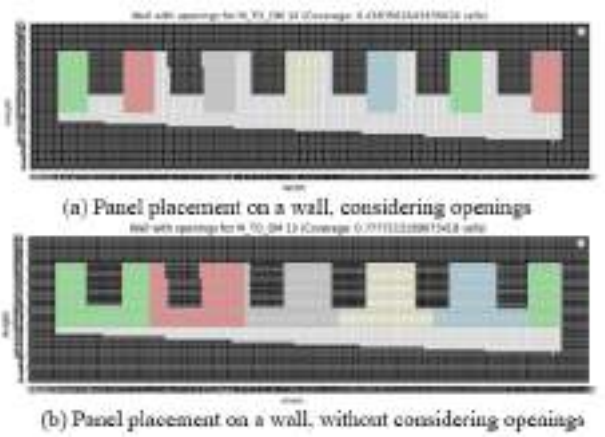


Figure 5. Panelized KCDC 11 with and without openings

In conclusion, the algorithm successfully addresses the dual objectives of maximizing the coverage of the facade and minimizing the usage of the panels, even when faced with the restrictions introduced by openings and irregular facade geometries. Looking ahead, the method could be applied to more complex building structures, such as multi-story buildings with varying facade shapes, by further enhancing the algorithm's ability to accommodate different panel configurations and structural constraints. Future improvements, such as the introduction of more diverse panel sizes or the further optimization of execution times, could further enhance its applicability to a wider range of facade types and building geometries.

4 Conclusion

Our research highlights a significant void in existing systems capable of retrofitting buildings through an end-to-end process—from virtualizing point cloud data to automatically generating panel layouts for each facade to facilitate panel installation. The goal of this study was to bridge the gap between building envelope digital twin generation for panelized systems with automatic tools for prefabricated panel installation using a novel digital tool for automated panel layout design. We propose a three step process to achieve the this: Facade Modelling, CSP Formulation, and Backtracking with AC-3. The proposed algorithm was demonstrated on a variety of facades from a simple rectangle to more complex polygons.

Operationally, the results of the algorithm are also quite robust, with panel layouts being generated in under two or three hours, which is a significant improvement compared to the days typically required for manual creation. Consequently due to the application of a backtracking algorithm for solving the meta-rectangle packing problem,

the user is also presented a variety of optimal solutions to choose from, depending on the optimization factor; rather than a single solution as is the case with other rectangle packing algorithms. While the proposed algorithm demonstrates efficient performance, there is always scope for improvement. Future work involves adapting the algorithm to incorporate a dynamic backtracking algorithm with memorization. This would allow the storage of results of previously calculated sub-problems, facilitating efficient domain pruning whilst reducing redundant computations. Such an approach has the potential to significantly improve the algorithm's efficiency and scalability, paving the way for the development of robust solutions in facade panelization and building retrofitting workflows.

Acknowledgement

This research was sponsored by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the U. S. Department of Energy, and used resources at the Building Technologies Research and Integration Center.

References

[1] Luis Pérez-Lombard, José Ortiz, and Christine Pout. A review on buildings energy consumption information. *Energy and buildings*, 40(3):394–398, 2008.

[2] Zhenjun Ma, Paul Cooper, Daniel Daly, and Laia Ledo. Existing building retrofits: Methodology and state-of-the-art. *Energy and buildings*, 55:889–902, 2012.

[3] Arva Arsiwala, Faris Elghaish, and Mohammed Zoher. Digital twin with Machine learning for predictive monitoring of CO2 equivalent from existing buildings. *Energy and Buildings*, 284:112851, 2023.

[4] Chukwuka Christian Ohueri, Md Asrul Nasid Masrom, and Taki Eddine Seghier. Digital twin for decarbonizing operating buildings: A systematic review and implementation framework development. *Energy and Buildings*, page 114567, 2024.

[5] Michel Aldanondo, Julien Lesbegueries, Andrea Christophe, Élise Vareilles, and Xavier Lorca. Building insulation renovation: a process and a software to assist panel layout design, a part of the ISOBIM project. In *ISARC 2023-40th International Symposium on Automation and Robotics in Construction*, pages 40–47, 2023.

[6] Bryan P Maldonado, Nolan W Hayes, and Diana Hun. Automatic point Cloud Building Envelope Segmentation (Auto-CuBES) using Machine Learning. In *Proceedings of the 40th International Symposium on Automation and Robotics in Construction*, pages 48–55, 07 2023. doi:10.22260/ISARC2023/0009.

[7] Nolan W. Hayes, Bryan P. Maldonado, Diana Hun, and Peter Wang. Real-time evaluator to optimize and automate crane installation of prefabricated components. In *Proceedings of the 40th International Symposium on Automation and Robotics in Construction*, pages 192–199, Chennai, India, July 2023. doi:10.22260/ISARC2023/0028.

[8] Diana Hun, Peter Lee-shein Wang, Nolan W Hayes, Bryan Maldonado Puente, Philip R Boudreaux, and Stephen M Killough. Real-Time Evaluator to Optimizing Prefab Retrofit Panel Installation, June 27 2024. US Patent App. 18/509,721.

[9] Andrea Lodi, Silvano Martello, Michele Monaci, and Daniele Vigo. Two-dimensional bin packing problems. *Paradigms of combinatorial optimization: Problems and new approaches*, pages 107–129, 2014.

[10] Richard E Korf. Optimal Rectangle Packing: Initial Results. In *ICAPS*, pages 287–295, 2003.

[11] B. P. Maldonado. Automatic Point Cloud Building Envelope Segmentation (AutoCuBES). Computer Software. Registration Number: TXu002412058, Dec 2023. URL `https://doi.org/10.11578/dc.20231214.1`.

[12] Eric Huang and Richard E Korf. Optimal rectangle packing: An absolute placement approach. *Journal of Artificial Intelligence Research*, 46:47–87, 2013.

[13] Andreas Bortfeldt. A genetic algorithm for the two-dimensional strip packing problem with rectangular pieces. *European Journal of Operational Research*, 172(3):814–837, 2006.

[14] Kevin Tole, Rashad Moqa, Jiongzhi Zheng, and Kun He. A simulated annealing approach for the circle bin packing problem with rectangular items. *Computers & Industrial Engineering*, 176:109004, 2023.

[15] Peter Van Beek. Backtracking search algorithms. In *Foundations of artificial intelligence*, volume 2, pages 85–134. Elsevier, 2006.

[16] Richard J Wallace. Why AC-3 is almost always better than AC-4 for establishing arc consistency in CSPs. In *IJCAI*, volume 93, pages 239–245, 1993.

[17] Sean Gillies, Casper van der Wel, Joris Van den Bossche, Mike W. Taves, Joshua Arnott, Brendan C. Ward, and others. Shapely, August 2024. URL `https://github.com/shapely/shapely`.

Pixel-Based Rust Severity Classification on Steel Surfaces Using HSI Features and Machine Learning Techniques

Shahinuzzaman[1], Po-Han Chen[2] , Emadaldeen Benshaaban[3]

[1,2,3] Department of Building, Civil and Environmental Engineering, Concordia University, Montreal, QC, Canada
s_sujon@live.concordia.ca, pohan.chen@concordia.ca, emadaldeen.benshaaban@concordia.ca

Abstract

Ensuring the safety and longevity of steel structures, like bridges and buildings, depends on the effective monitoring of rust and its severity. Traditional methods of rust detection often rely on visual inspections, which can be subjective and time-consuming. In this paper, we present a more efficient way to classify rust severity on steel surfaces using the HSI color space, along with pixel value analysis and computer vision techniques. By utilizing the different components of the HSI color model, we improve the accuracy of rust segmentation, which helps in better identifying the severity levels of rust. This method analyzes individual pixel values in images of steel surfaces, providing a detailed understanding of the rust's extent and intensity. Additionally, we conducted histogram analysis to calculate the mean and standard deviation of pixel values for assessing rust severity. Our machine learning framework, which includes Random Forest and OpenCV, processes high-resolution images to classify rust severity automatically through image segmentation and thresholding techniques. This method also incorporates the K-means algorithm with the Elbow Method to determine the optimal number of clusters and perform segmentation automatically. The results show that this approach offers a more precise and objective way of assessing rust damage according to the severity compared to traditional inspection methods.

Keywords –

Rust segmentation, Rust severity identification, steel surface, HIS color space, Pixel value analysis

1 Introduction

Rust on the steel surfaces of civil engineering structures is a prevalent problem that may pose significant safety hazards. [1], [2], [3]. The corrosion of steel structures poses a considerable challenge for engineers globally. [4], and prompt and precise monitoring of rust damage on steel surfaces is crucial to maintaining the structural integrity, safety, and durability of steel facilities [5],[6]. Conventional rust examination on steel surfaces, particularly for bridges, is laborious, time-consuming, and occasionally hazardous [6] . Since the late 1990s, vision-based image processing techniques have garnered significant attention for inspecting surface defects in infrastructure, such as rust and cracks [7], [8], [9], [10]. Corrosion, including rust, costs the global economy an estimated $2.5 trillion, equivalent to 3.4% of global GDP. Proper monitoring and preventive measures could reduce this cost by up to 35%, potentially saving US$875 billion [11]. In high-risk industries like construction, rust-induced structural failures are a significant cause of accidents. In 2022, for example, 1,056 fatalities occurred in the construction sector, with 423 deaths resulting from falls, slips, or trips [12].

Computer vision involves methodologies that allow machines to interpret the visual environment via digital photographs via machine learning algorithms. Visual media, including images and videos, are the predominant inputs for computer vision applications [13]. Computer vision-based structural health monitoring (CV-SHM) has garnered considerable interest in infrastructure management, including crack, rust, etc. Many computer vision techniques are available for surface defect detection, such as crack, which focus on the analysis of linear features and structural discontinuities, often using geometric modeling techniques. However, these techniques are less effective for rust detection, where the primary focus is on color variation, texture analysis, and corrosion patterns specific to rust. This methodology includes several techniques, including image classification, object detection, object segmentation, visual tracking, 3D reconstruction, image generation, image enhancement, pose estimation, action recognition, image captioning and optical flow analysis [14], [15]. The most common techniques are classification, detection, and segmentation. Segmentation is divided into two types: semantic segmentation and instance segmentation. Semantic segmentation labels every pixel of an object but doesn’t distinguish between individual items in the same group. On the other hand, instance segmentation not only labels the pixels but also identifies

and separates each instance of the same object (Figure 1) [16]. Computer Vision (CV) techniques are sophisticated instruments for assessing structural integrity, facilitating non-contact monitoring with great precision [17]. When combined with deep learning, computer vision markedly improves object detection abilities[17], [18].

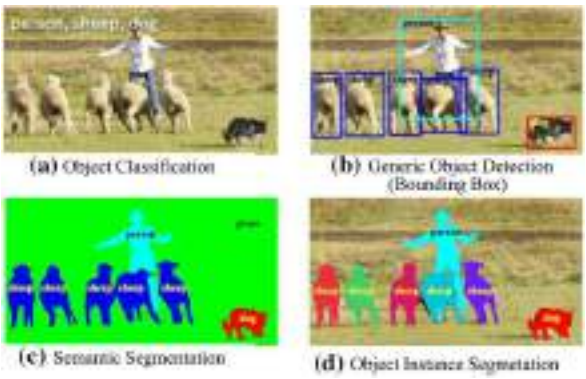


Figure 1. Computer vision techniques: a) image-level object classification, b) bounding box level generic object detection, c) pixel-wise semantic segmentation, d) instance-level semantic Segmentation [19].

2 Related work

We have organized our literature review into two main areas: (1) color space and (2) pixel value analysis

2.1 Color Space

Color properties are a fundamental part of image analysis because they help us understand the visual details in an image, such as pixel values and textures. Different color spaces, like RGB, HSI, and LAB, YCbCr each have their unique ways of representing color, using different shapes and ranges of pixel values. [20], [21]. The most commonly used color space is the three-channel Red, Green, and Blue (RGB) color space (Figure 2.) . In the RGB color space, each channel (Red, Green, and Blue) can represent 256 different values, ranging from 0 to 255, in an 8-bit image. This means that for each color channel, there are 256 possible intensity levels. For example, when the pixel values are (0, 0, 0), it represents pure black, with no color present. On the other hand, when the pixel values are (255, 0, 0), it indicates the maximum intensity of red. With 256 possible values for each of the three channels, the total number of possible colors in the RGB color space is 256 x 256 x 256, or about 16 million different colors.[20].

The conversion of images from RGB to grayscale is a common technique in image processing, as well as in rust identification. [2]. In 1995, FURUTAI et al. [22] Proposed a model for structural damage assessment caused by corrosion using the hue, saturation, and value (HSV) color space. This approach was first applied to color space with image processing. Bridge paint assessment through image processing proposed by Chen and Chang in 2003 [23]. The authors transformed all RGB photos into grayscale and delineated three regions according to the illumination conditions of the rust images. In 2004, Choi and Kim [24] employed the Hue (H), Saturation (S), and Intensity (I) HSI color space for corrosion detection and the identification of five unique corrosion types. In 2006, Lee et al. [25] proposed the utilization of color image rather than grayscale images to inspect steel bridge surface coatings. One of the initial methodologies in color space analysis for image processing was introduced by Chen et al. [26]. The authors examined 14 color spaces to identify the most effective one for addressing non-illumination issues. Similarly, Shen et al. [8] Investigated 19 color spaces, including grayscale, to address non-illumination problems in images.

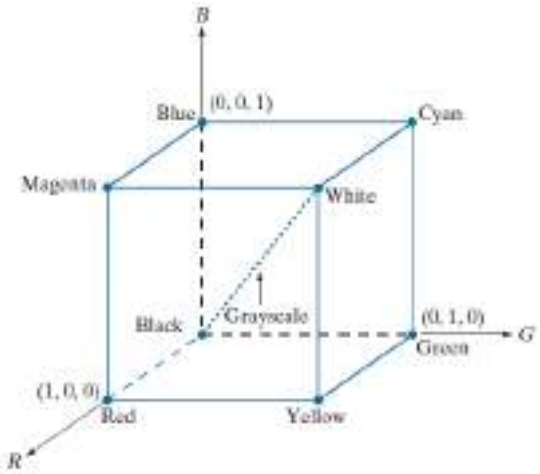


Figure 2. RGB three-dimensional Color cube [26].

Kuo-Wei Liao et al. [27] suggested an automated method for rust identification on a steel bridge utilizing digital image analysis. They employed three picture recognition methodologies: (1) K-means clustering in the hue channel, (2) double-center-double-radius (DCDR) method for RGB color space, and (3) DCDR algorithm for Hue-Saturation-Intensity (HSI) color space. The authors compared 12 color spaces (RGB, HIS, YPbPr, YCbCr, JPEG-YCbCr, YDbDr, YIQ, YUV, HSV, XYZ, LAB, CAT02 LMS) and determined that the Hue characteristic was essential and more efficacious. Yundong Li et al. [28] Converted the RGB image into the YCbCr color space and calculated only the Cr component using the Otsu method to identify rust.

2.2 Pixel Value Analysis

Pixel is the most commonly used term to refer to the components of a digital image [20]. In one of the earliest studies on rust identification, Chen et al. (2009) [26] converted RGB images into the Hue, Saturation, and Intensity (HSI) color space to analyze pixel values associated with rust. Their findings revealed that the mean pixel values in the RGB space were (R, G, B) = (173.22, 107.76, 60.54), while in the HSI space, the means were (H, S, I) = (24.4909, 0.4646, 0.4471). The corresponding standard deviations in the HSI space were (H, S, I) = (6.5469, 0.1498, 0.1442). In 2023, Sun et al. [3] analyzed corrosion using pixel scatter diagrams in the RGB color space. They calculated the mean pixel values as C (R, G, B) = (193.94, 123.57, 59.10) and the standard deviations as σ(R, G, B) = (29.88, 29.12, 17.92).

3 Methodology

Our methodology is divided into two segments. The first segment involves calculating the mean and standard deviation of HIS pixel values. The second segment focuses on training and predicting rust severity.

3.1 Statistical Analysis of HIS Pixel Values

3.1.1 Converting Image from RGB to HSI

Pixel values constitute the essential components and smallest units of a digital image [29]. For this study, we collected data from diverse and heterogeneous sources, including mobile phones, digital cameras, drones, and open-source repositories. This diversity ensures that the dataset reflects a wide range of real-world scenarios. Typically, digital cameras and mobile phones produce images in the Red, Green, and Blue (RGB) color model by default. To enhance the utility of the dataset, especially for addressing challenges like shadow removal and effectively identifying rust severity in the images, we converted all collected data from the RGB color space to the Hue, Saturation, and Intensity (HSI) color space. The range of RGB values is typically 0 to 255 for an 8-bit image [20], as discussed in the previous section. For this study, we normalized the RGB values to fit the HSI color model, where the intensity and saturation values are scaled to a range of 0 to 1, and the hue values are normalized to a range of 0 to 360 degrees using the following formula (Equations 1-4).

$$H\ (Hue) = \begin{cases} \theta & if\ B \le G \\ 360 - \theta & if\ B > G \end{cases} \quad (1)$$

with

$$\theta = Cos^{-1}\left\{ \frac{\frac{1}{2}\left[(R-G) + (R-B)\right]}{\left[(R-G)^2 + (R-B)(G-B)\right]^{\frac{1}{2}}} \right\} \quad (2)$$

$$S(Sat.) = 1 - \frac{3}{(R+G+B)}[\min(R,G,B)] \quad (3)$$

$$I\ (Intensity) = \frac{1}{3}\ (\mathrm{R} + \mathrm{G} + \mathrm{B}) \quad (4)$$

Hue represents the type of color Hue represents the type of color (e.g., red, blue, green) and is expressed in degrees (0°–360°). Saturation measures the purity of the color. It is computed as the proportion of the maximum deviation of the pixel color values from the neutral gray level. The intensity is the average of the three RGB components and represents the brightness of the pixel.

3.1.2 Optimal Number of Clusters Using Elbow Method and K-Means Clustering

Unsupervised machine learning, particularly the k-means clustering algorithm, is widely regarded as one of the most popular and efficient techniques for image clustering [30]. Despite its advantages, a significant limitation of k-means is that it requires the number of clusters to be predefined before the clustering process. This can be challenging when dealing with real-world data, where the optimal number of clusters is not known in advance. To resolve this issue, we utilized the Elbow Method, a commonly employed heuristic for identifying the ideal number of clusters. This method entails computing the within the sum of cluster squared (WCSS) distances (inertia) (Equation 5) between each data point and its respective cluster centroid across various cluster characteristics. The results are subsequently displayed to generate an inertia curve. The 'elbow' or 'knee' point of this curve signifies the appropriate number of clusters since it denotes the juncture when the addition of further clusters yields negligible enhancement in diminishing inertia, reflecting a balance between model complexity and efficacy.

To enhance precision and automation, we employed a knee locator algorithm that correctly determines the elbow position in the curve. By integrating k-means with the Elbow Method, we established a rigorous and data-driven strategy to ascertain the appropriate number of clusters, therefore enhancing the efficiency and precision of our picture clustering procedure.

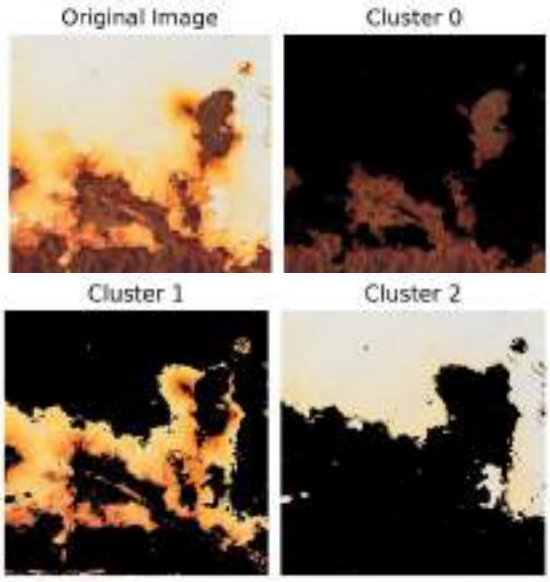


Figure 1. Rust image clustering using the k-means algorithm to visualize and evaluate rust severity in accordance with the AASHTO-2019 standards.

$$\text{Inertia (WCSS)} = \sum_{i=1}^{k} \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (5)$$

- k is the number of clusters
- Ci= The set of points assigned to the cluster i
- x= A data point in cluster Ci
- μ_i= The centroid of cluster Ci

In this paper, we consider three levels of rust severity: Severe, Major, and Minor. These categories were derived from the guidelines provided by the American Association of State Highway and Transportation Officials (AASHTO) in their 'Manual for Bridge Element Inspection 2019' [31]. This guidebook serves as a reliable resource for evaluating bridge conditions, including rust assessment. To comply with these requirements, we assessed rust severity by categorizing photos into separate clusters (Figure 1). Each cluster was subsequently aligned with one of the severity categories established by AASHTO. This methodology enabled us to categorize rust patterns according to picture attributes while maintaining alignment with industry standards. Utilizing these clusters, we can systematically identify and assess the degree of rust.

3.1.3 Histogram Analysis

A histogram is a graphical depiction that illustrates the frequency distribution of a dataset. In image analysis, histograms are frequently employed to examine the distribution of pixel values over various color spaces, including Hue, Saturation, and Intensity (HSI). Due to the absence of publicly available rust specific datasets, we collected a set of 30 real-world images of rusted steel surfaces. These images were analyzed using histogram-based techniques to calculate the mean and standard deviation (SD) of pixel values corresponding to varying rust severity levels. These pixel ranges were then used to train a Random Forest model, aimed at classifying rust severity based on the calculated values.

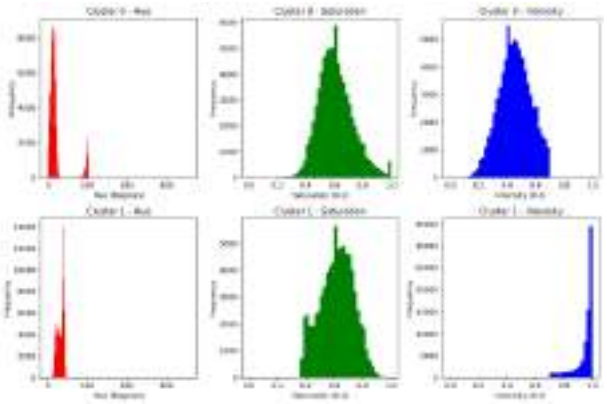


Figure 2. Histograms of H(Hue), S(Saturation), and I (Intensity) for Cluster 0 in HSI Color Space.

From Figure 1, we focused our analysis on Clusters 0 and 1, as Cluster 2 primarily represented the background and did not contain rust. We conducted a histogram analysis to determine the range of pixel values corresponding to different categories of rust severity.

The histograms, as shown in Figure 2, represent the distribution of HSI values for Cluster 0 and Cluster 1. In Cluster 0, the Hue values predominantly range from approximately 0 to 100. The mean Hue value for Cluster-0 is 20.11, with a standard deviation of 26.30. The saturation histogram displays a pyramid-shaped distribution with values ranging between 0.4 and 0.8, indicating a moderate degree of color vibrancy within this cluster. The mean saturation value is 0.61, with a standard deviation of 0.12. Similarly, the intensity histogram follows mid-intensity values of 0.2 to 0.7. The mean intensity value is 0.45, with a standard deviation of 0.11. These histograms provide a detailed overview of the color characteristics and brightness distribution within Cluster 0. A similar analysis was conducted for Cluster 1 further to compare the distinct HSI properties of the clusters.

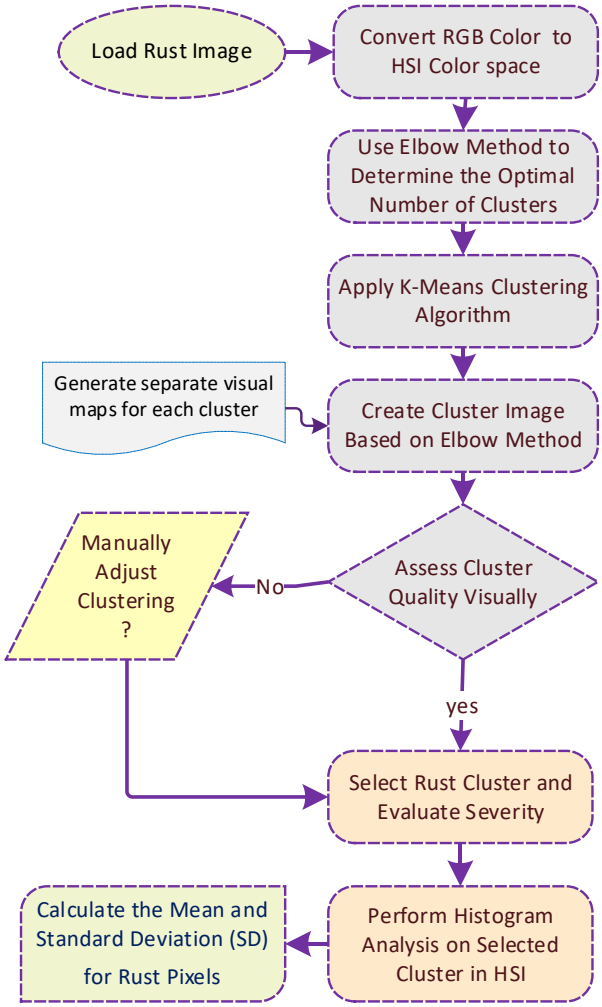


Figure 3. Workflow for Rust Severity Analysis Using Clustering and HSI Histogram Analysis.

We conducted this histogram analysis on 30 rust images and calculated the mean and standard deviation values for rust severity in the HSI color space (Table 1).

Table 1 Mean and Standard Deviation of HSI Pixel Values for Rust Severity Classification.

Severity Pixel	H	SD	S	SD	I	SD
Severe	18.4	16.5	0.65	0.13	0.44	0.14
Major	24.54	12.39	0.61	0.11	0.60	0.09
Minor	47.98	15.80	0.42	0.12	0.68	0.08

The methodology illustrated in Figure 3. provides a systematic process for analyzing rust in images. The process begins by loading the image and converting its colors from RGB to HSI, facilitating a focus on the significant color attributes associated with rust. The Elbow Method is employed to determine the optimal number of groups for segmenting the image, followed by the application of K-means clustering to form these clusters. Visual representations for each cluster are generated and amalgamated into a singular image to validate the coherence of the clustering. If the findings are inaccurate, manual changes might be implemented to enhance the segmentation. Upon accurate identification of the rust regions, the rust cluster is chosen for subsequent study. Analyzing the cluster's color distribution through a histogram and computing essential statistics such as the mean and standard deviation of the rust pixels facilitates the assessment of rust severity in the image. This approach guarantees a comprehensive and accurate analysis.

3.2 Rust Severity Training and Prediction

3.2.1 Training Random Forest Models with Synthetic Data

After calculating the pixel values, we created synthetic data using the mean and standard deviation values derived from our analysis. A Random Forest machine learning algorithm was trained on this synthetic dataset, enabling us to classify rust severity levels. Using the trained model, we generated a set of virtual images based on the input pixel characteristics and applied the trained model to classify external images effectively.

The proposed methodology involves a systematic approach to rust severity classification and segmentation. Synthetic training data is first produced by establishing criteria for hue, saturation, and intensity (HSI) values that align with designated rust severity levels: Severe, Major, and Minor. These criteria are established on the observed mean ± standard deviation values obtained from histogram analysis. A labeled dataset is generated by simulating HSI data points for each severity level based on these thresholds. This dataset is utilized to train a Random Forest classifier, a resilient and adaptable supervised machine learning technique. The dataset is divided into training and testing subsets, and the model's performance is meticulously assessed using metrics including accuracy and F1-score to verify its dependability.

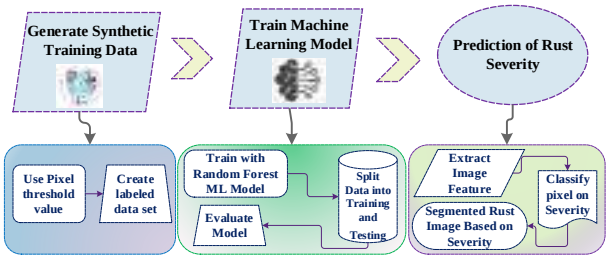


Figure 4. Workflow for Rust Severity Training and Prediction Using Machine Learning Models.

Upon completion of training, the model is utilized on actual photos. Pixel-wise HSI features are derived from the input image by transforming it from RGB to the HSI color space and restructuring the pixel values into a format suitable for the Random Forest model. The trained classifier forecasts the rust severity level Severe, Major, or Minor for each pixel in the image according to its HSI properties. The severity levels are represented by specific colors: Red for Severe, Orange for Major, and Yellow for Minor.

3.2.2 Prediction of Rust Severity

In the concluding phase, the trained Random Forest classifier was utilized to predict the rust severity category for each pixel in the input image, relying on its hue, saturation, and intensity (HSI) attributes. The anticipated labels were assigned numerical values that correspond to established severity levels: Severe, Major, Minor, and Background. The predictions were subsequently reformulated to match the original image dimensions, resulting in a segmentation map. The segmentation outcome, alongside the original image, offered a distinct representation of rust severity levels across the surface (Figure 5). This visual depiction enables accurate identification of rust damage and offers a quantitative framework for evaluating rust severity. This paper incorporates morphological operations techniques to fine-tune the segmented image. A 2×2 kernel was applied for erosion and dilation, refining the boundaries and reducing noise to enhance the segmentation quality.

4 Result

The K-means clustering algorithm was successfully applied, incorporating the elbow method to analyze histogram data. Based on this analysis, the mean and standard deviation of pixel values for each severity level were calculated, as detailed in Table 1. Using the calculated mean and standard deviation values, synthetic data was generated to represent numeric pixel values, which were then used to train a Random Forest machine learning model. The model's performance was evaluated, achieving an F1 score of 0.75, a precision of 0.74, and an accuracy of 0.81. The segmentation outcomes delineated distinct borders among the clusters, offering a visual depiction of rust severity over the surface. A pixel-based analysis was performed to measure the area occupied by each cluster. The image-1 comprised a total of 512*512= 262,144 pixels, with 69775 pixels (26.62%) indicative of Severe rust, 30929 pixels (11.80%) indicative of Major rust, and 52150 pixels (19.89%) indicative of Minor rust.

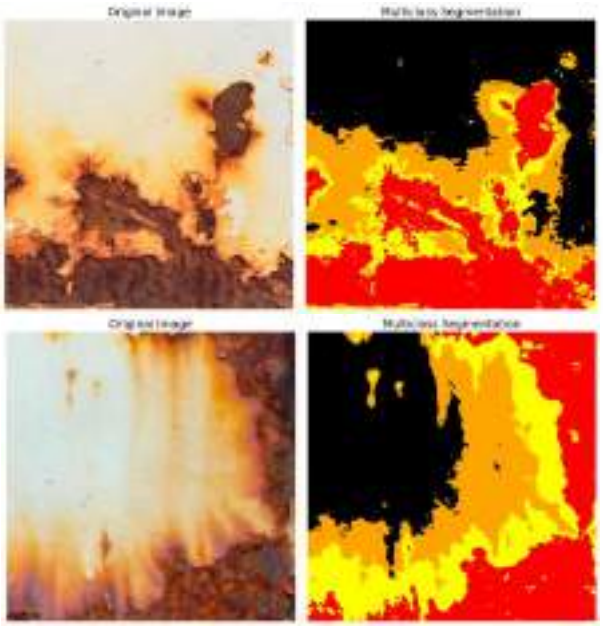


Figure 5. Original and Multiclass Segmentation image of Rust Severity (Severe: Red, Major: Yellow, Minor: Orange, Background: Black).

The Background comprised 109290 pixels (41.69%) of the image area. This quantitative research illustrates the efficacy of the clustering method in detecting and categorizing areas of differing rust intensity, offering significant insights for surface condition evaluation.

5 Discussion

The classification of rust severity is essential for corrosion monitoring and management, especially in industrial settings, where prompt detection can avert structural failures and minimize maintenance expenses. This study utilized an innovative method that integrates clustering-based analysis with machine learning to categorize rust severity into three unique classifications: severe, major, and minor. This workflow integrates unsupervised and supervised methodologies to attain precise and fast segmentation. The conversion of RGB photos to the HSI color space was crucial in the study. In contrast to RGB, the HSI model separates color intensity from chromaticity, offering a more comprehensible depiction of rust attributes under different lighting circumstances. This preprocessing procedure guaranteed that rust segmentation was resilient to shadows and inconsistent lighting, hence improving the dependability of future clustering and classification. The utilization of the Elbow method to ascertain the ideal number of clusters offered an unsupervised approach for segmenting the rusted areas based on pixel similarity. K-means clustering enhanced this segmentation by

categorizing pixels into discrete clusters, denoting areas with differing rust intensity. The computed mean and standard deviation (SD) values of pixel HSI components for each cluster facilitated the creation of synthetic data, which constituted the training basis for machine learning models.

A supervised machine learning method, specifically the Random Forest algorithm, was employed to categorize rust severity. This approach was selected for its resilience, capacity to manage high-dimensional data, and interpretability. Through training on synthetic data generated from computed HSI thresholds, the model demonstrated effective generalization to novel images. The evaluation criteria, including accuracy, precision, recall, and F1-score, validated the model's efficacy, with findings affirming its ability to effectively forecast rust severity across various test samples.

6 Conclusion and Future Research

This study significantly contributes by integrating histogram-based analysis with machine learning. The technology integrates traditional image processing with data-driven approaches by utilizing statistical features of pixel distributions, including mean and standard deviation. This hybrid approach guarantees that the classification of rust severity is both comprehensible and scalable for practical applications. Moreover, the visual depiction of segmented output images provides actionable insights, with pixels color-coded by severity. Morphological operations with a 2×2 kernel further refined the segmentation by enhancing boundary accuracy and reducing noise, enabling users to assess the degree and spread of rust intuitively. This quantitative approach provides a dependable decision-making instrument for rust monitoring, facilitating tailored interventions to reduce rust damage successfully.

Notwithstanding its advantages, the proposed methodology does have several limitations. One significant limitation is the reliance on synthetic data specifically, pixel value ranges (mean and standard deviation) derived from a small set of 30 real-world images. Due to the lack of publicly available rust datasets, we used these 30 real-world images to generate the necessary training data. While this approach provides a controlled dataset, the small size and lack of diversity may limit the model's generalization ability, especially with varying rust conditions. Furthermore, external factors like illumination fluctuations and surface texture variations can influence the model's predictions.

Despite these limitations, the approach remains a valuable starting point. Future work will aim to tackle these challenges by expanding the dataset, incorporating a wider variety of real-world images, and integrating adaptive thresholds to account for varying conditions. Additionally, exploring advanced techniques such as deep learning could enhance model generalization and accuracy. This study's methodology, while in need of further refinement, offers a realistic and scalable solution for classifying rust severity. By combining clustering, machine learning, and quantitative analysis, we establish a solid framework that can be further developed for industrial applications, improving both efficiency and precision in rust severity management systems.

7 Reference

[1] I. Katsamenis, N. Doulamis, A. Doulamis, E. Protopapadakis, and A. Voulodimos, "Simultaneous Precise Localization And Classification of metal rust defects for robotic-driven maintenance and prefabrication using residual attention U-Net," *Autom Constr*, vol. 137, May 2022, doi: 10.1016/j.autcon.2022.104182.

[2] A. Rahman, ; Zheng, Y. Wu, M. Asce, and R. Kalfarisi, "Semantic Deep Learning Integrated with RGB Feature-Based Rule Optimization for Facility Surface Corrosion Detection and Evaluation," 2021, doi: 10.1061/(ASCE).

[3] L. Sun, Y. Li, X. Li, and C. Liu, "Corrosion defect segmentation method based on superpixel feature cascade," *Ain Shams Engineering Journal*, vol. 15, no. 2, Feb. 2024, doi: 10.1016/j.asej.2023.102425.

[4] H. U. Sajid and R. Kiran, "Influence of corrosion and surface roughness on wettability of ASTM A36 steels," *J Constr Steel Res*, vol. 144, pp. 310–326, May 2018, doi: 10.1016/j.jcsr.2018.01.023.

[5] Y. Zhang, B. M. Ayyub, and J. F. Fung, "Projections of corrosion and deterioration of infrastructure in United States coasts under a changing climate," *Resilient Cities and Structures*, vol. 1, no. 1, pp. 98–109, Mar. 2022, doi: 10.1016/j.rcns.2022.04.004.

[6] C. Y. Liu and J. S. Chou, "Bayesian-optimized deep learning model to segment deterioration patterns underneath bridge decks photographed by unmanned aerial vehicle," *Autom Constr*, vol. 146, Feb. 2023, doi: 10.1016/j.autcon.2022.104666.

[7] S. Jiang, Y. Wu, and J. Zhang, "Bridge coating inspection based on two-stage automatic method and collision-tolerant unmanned aerial system," *Autom Constr*, vol. 146, Feb. 2023, doi: 10.1016/j.autcon.2022.104685.

[8] H. K. Shen, P. H. Chen, and L. M. Chang, "Automated steel bridge coating rust defect

recognition method based on color and texture feature," *Autom Constr*, vol. 31, pp. 338–356, 2013, doi: 10.1016/j.autcon.2012.11.003.

[9] S. Lee, L. M. Chang, and M. Skibniewski, "Automated recognition of surface defects using digital color image processing," *Autom Constr*, vol. 15, no. 4, pp. 540–549, Jul. 2006, doi: 10.1016/j.autcon.2005.08.001.

[10] N. Dhanachandra, K. Manglem, and Y. J. Chanu, "Image Segmentation Using K-means Clustering Algorithm and Subtractive Clustering Algorithm," in *Procedia Computer Science*, Elsevier, 2015, pp. 764–771. doi: 10.1016/j.procs.2015.06.090.

[11] E. Bowman, N. Thompson, D. Gl, O. Moghissi, M. Gould, and J. Payer, "International Measures of Prevention, Application, and Economics of Corrosion Technologies Study," 2016. Accessed: Sep. 30, 2024. [Online]. Available: National Association of Corrosion Engineers (NACE)

[12] Bureau of Labor Statistics, "NATIONAL CENSUS OF FATAL OCCUPATIONAL INJURIES IN 2022," Dec. 2023. [Online]. Available: www.bls.gov/iif

[13] A. A. I. M. Ali *et al.*, "Computer Vision and Image Processing Approaches for Corrosion Detection," Oct. 01, 2023, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/jmse11101954.

[14] Y. Wang, U. Ahsan, H. Li, and M. Hagen, "A Comprehensive Review of Modern Object Segmentation Approaches," 2022, *Now Publishers Inc*. doi: 10.1561/0600000097.

[15] C. Z. Dong and F. N. Catbas, "A review of computer vision–based structural health monitoring at local and global levels," Mar. 01, 2021, *SAGE Publications Ltd*. doi: 10.1177/1475921720935585.

[16] picsellia.com, "Segmentation vs Detection vs Classification in Computer Vision: A Comparative Analysis." Accessed: Feb. 29, 2024. [Online]. Available: https://www.picsellia.com/post/segmentation-vs-detection-vs-classification-in-computer-vision-a-comparative-analysis

[17] B. Kim, N. Yuvaraj, H. W. Park, K. R. S. Preethaa, R. A. Pandian, and D. E. Lee, "Investigation of steel frame damage based on computer vision and deep learning," *Autom Constr*, vol. 132, Dec. 2021, doi: 10.1016/j.autcon.2021.103941.

[18] Y. J. Cha, K. You, and W. Choi, "Vision-based detection of loosened bolts using the Hough transform and support vector machines," *Autom Constr*, vol. 71, no. Part 2, pp. 181–188, Nov. 2016, doi: 10.1016/j.autcon.2016.06.008.

[19] L. Liu *et al.*, "Deep Learning for Generic Object Detection: A Survey," *Int J Comput Vis*, vol. 128, no. 2, pp. 261–318, Feb. 2020, doi: 10.1007/s11263-019-01247-4.

[20] R. C. . Gonzalez and R. E. . Woods, *Digital image processing*. Pearson, 2018.

[21] F. - Ponce, "Computer Vision - A Modern Approach."

[22] H. Furutai, T. Deguch12, and M. Kushida3, "Neural Network Analysis of Structural Damage Due to Corrosion," Osaka, Japan, Sep. 1995. doi: 10.1109/ISUMA.1995.527678.

[23] P. H. Chen and L. M. Chang, "Artificial intelligence application to bridge painting assessment," *Autom Constr*, vol. 12, no. 4, pp. 431–445, Jul. 2003, doi: 10.1016/S0926-5805(03)00016-5.

[24] K. Y. Choi and S. S. Kim, "Morphological analysis and classification of types of surface corrosion damage by digital image processing," *Corros Sci*, vol. 47, no. 1, pp. 1–15, Jan. 2005, doi: 10.1016/J.CORSCI.2004.05.007.

[25] S. Lee, L. M. Chang, and M. Skibniewski, "Automated recognition of surface defects using digital color image processing," *Autom Constr*, vol. 15, no. 4, pp. 540–549, Jul. 2006, doi: 10.1016/j.autcon.2005.08.001.

[26] P. H. Chen, Y. C. Yang, and L. M. Chang, "Automated bridge coating defect recognition using adaptive ellipse approach," *Autom Constr*, vol. 18, no. 5, pp. 632–643, Aug. 2009, doi: 10.1016/j.autcon.2008.12.007.

[27] K. W. Liao and Y. T. Lee, "Detection of rust defects on steel bridge coatings via digital image recognition," *Autom Constr*, vol. 71, no. Part 2, pp. 294–306, Nov. 2016, doi: 10.1016/j.autcon.2016.08.008.

[28] Y. Li, A. Kontsos, and I. Bartoli, "Automated Rust-Defect Detection of a Steel Bridge Using Aerial Multispectral Imagery," *Journal of Infrastructure Systems*, vol. 25, no. 2, Jun. 2019, doi: 10.1061/(asce)is.1943-555x.0000488.

[29] Y. Cai, "How Many Pixels Do We Need to See Things?," 2003.

[30] K. P. Sinaga and M. S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.

[31] "Manual for Bridge Element Inspection from IHS Markit," 2019.

The Artificial Intelligence (AI) Impact on Construction Project Management

Kwame B.O. Amoah[1]
Department of Civil and Architectural & Construction Management, University of Cincinnati, OH, USA.
amoahkb@ucmail.uc.edu

Abstract – The study aims to explore the degree of Artificial Intelligence (AI) integration in construction projects and analyze its effects on cost-effectiveness, safety, and productivity. The objective is to investigate the relationship between the application of AI and its integration in the construction sector, evaluating the possible changes and adjustments in the labor force brought about by AI technological incorporation. The author employed a systematic review methodology. The review approach provides comprehensive insights into the application of AI in real-world construction projects inform data evaluation trends of AI application in the industry. The preliminary results highlight AI's enormous potential for streamlining repetitive processes, maximizing resource use, and improving construction safety standards. Integrating AI into construction projects may result in significant gains in productivity and safety, but doing so requires a careful strategy in balancing the workforce's changing requirements and technology improvements, and the role of the industry is vital in this process.

Keywords: Artificial Intelligence; AI Project Management; AI powered Construction: Construction Project Management: Machine Learning.

1 Introduction

Especially compared to more digitalized industries like manufacturing, the construction industry faces ongoing obstacles that hinder its growth and competitiveness despite its critical role in global development [1]. The industry is characterized by a culture resistant to change and firmly embedded manual procedures. As a result, it needs to catch up with new technologies, which leads to reduced inefficiency [2]. These difficulties, the industry's workforce problems, the pressing need for sustainable infrastructures, and time management highlight the importance of embracing digitalization and using artificial intelligence's (AI) disruptive potential [3].

The application of Artificial Intelligence (AI) in construction projects is not just a trend, but a transformative wave poised to revolutionize the industry's operational landscape. Traditionally known for its labor-intensive practices, the construction sector faces challenges such as low productivity, inefficiencies, safety concerns, and cost overruns. AI's potential offers solutions to age-old problems that have hampered growth and innovation within the construction domain.

The construction industry needs more productivity, waste management, safety concerns, and cost overruns. Despite technological advancements, experts in construction projects must explore AI integration. The industry's historical reliance on conventional methods and resistance to change has hindered its adoption of AI-driven solutions. Challenges include high initial costs, limited expertise, trust deficits, and cultural barriers. However, the critical need for comprehensive studies is clear.

Procore and Autodesk BIM 360 are AI-powered construction project management software that analyses data to forecast hazards and improve timelines. They free up time for making decisions by automating processes like scheduling and documentation. These technologies let teams collaborate by visualizing project progress through 3D models and reports [4]. They improve overall project efficiency and results by using machine learning to find patterns in data for improved resource allocation and cost control. The main hypotheses of the paper include: 1. Explores the AI application processes in construction project management, 2. Addresses AI integration into construction project and assesses how the integration affects productivity, safety concerns and cost effectiveness, and 3. Provides opportunities to mitigate AI applications' challenges in the construction industry.

The study's unique contribution provides insights into key AI applications and integration as they apply to construction-specific challenges and the pathway to realizing AI's measurable benefits in the construction industry.



Figure 1: AI applications in different fields of construction.

2 Artificial Intelligence Relevance in Project Management

The advent of AI dates back to 1950 when the British mathematician Alan Turing asked whether machines could think [4]. After experiencing ups and downs over 60 years, AI has regained technological importance due to rapid developments in computing, big data, artificial neural networks, deep learning, and other new technologies [5, 6]. AI is a computer system that perceives visual perception, recognizes speech, and can translate between languages. It solves complex decision-making processes that mathematics cannot solve directly by understanding project information and the environment [6]. Machine learning is a subsidiary of AI, using algorithms and statistical models to study data and make decisions 8, 7]. With the advancements in machine learning and big data technologies, AI has become a significant technological opportunity globally. The potential of this technological advancement also triggers the need for AI use in the AEC industry. On a limited scope, the need for AI use in construction arises from the desire to improve cost management, enhance quality control, increase efficiency and productivity, and address complex issues in project management. The potential areas that the AEC industry can benefit from AI use are enormous.

As construction projects become complex and large-scale in the AEC industry, the number of participants and the volume of construction data, including project plans, schedules, and performance metrics, increase. In this regard, AI stands out as one of the digital technologies with significant potential to leverage the vast amount of available big data for problem-solving and enhancing decision-making within construction management [10-9]. AI can use machine learning algorithms to inform decision-making, optimize resource allocation, and improve project outcomes by identifying patterns, trends, and insights. As a result, artificial intelligence can offer continuous real-time monitoring and analysis of construction operations, enabling project managers to make well-informed decisions based on data [11] promptly. This process improves project control, reduces risks, enhances overall project performance, and prevents poor decision-making through the project management processes.

The use of AI in the AEC industry also has the potential to sustain project success. Project success is mainly defined by the project management triangle, consisting of time, cost, and quality pillars [12]. Planning a project is one of the most critical processes in project management, but the importance of designing software needs to be investigated [13]. Construction cost control is crucial, as delays in any project stage can cause cost overruns due to increased artistry, working time, or material usage [14]. AI has the potential to forecast cost overruns by considering project scale, contract type, and the competence of the project manager [15]. By implementing AI in construction scheduling, managers can monitor schedules more efficiently by estimating construction project completion and delay times [16]. AI can be applied in real-time, enabling project managers to swiftly and knowledgeably decide how to allocate resources as the project unfolds [17]. This can help prevent project delays and lead to more realistic timelines for future projects.

Furthermore, AI can provide a clear and realistic view of construction site activities to top-level management and engineers, improving construction efficiency and quality [18]. AI can also analyze data from various sources, such as sensors and cameras, to monitor real-time construction quality. It can identify defects, deviations from specifications, and potential issues, enabling early intervention and improving overall quality control.

Safety is a paramount and indispensable aspect of the construction industry. AI's potential to anticipate and prevent occupational accidents and equipment-related safety issues during construction projects is a significant development. AI-powered technologies, such as drones, robots, and wearable technologies, can be used for site inspections, monitoring hazardous areas, identifying potential safety risks, and monitoring workers' behavior. AI algorithms can analyze real-time data on-site from sensors to detect unsafe conditions, helping to prevent accidents and improve safety on construction sites. Additionally, pattern recognition-based AI technologies are used for data and system integration to enhance safety management. These technologies become even more potent when paired with virtual reality as they assure real-time personnel safety.

As the main field of construction management, AI can monitor, recognize, evaluate, and predict potential risks in terms of safety, quality, efficiency, and cost across teams and work areas, even under high uncertainty [19, 20]. AI-based risk analysis can provide insights to help project managers quickly prioritize potential risks and identify proactive actions rather than risk mitigation responses [21]. Machine learning and natural language processing are being applied in construction for risk detection and assessment to issue early warnings [20]. Consequently, AI is expected to play a significant role for project managers in risk assessment, generating decision support, automation of risk monitoring, and simulation and scenario analysis. Additionally, AI technologies, such as Building Information Modeling (BIM) and virtual reality, allow for better visualization, clash detection, and coordination among different disciplines, reducing errors and rework during construction.

In summary, the introduction of AI systems in the construction industry represents a significant leap

forward. These systems leverage advanced technologies, data analysis techniques, and automation to address specific challenges and improve efficiency, productivity, safety, decision-making, and overall project outcomes. AI systems bring new capabilities, insights, and efficiencies to the construction industry, enabling practitioners to leverage data-driven intelligence for better project execution. By embracing AI technologies, the AEC industry can drive innovation, increase productivity, support decision-making arising from complex project management challenges, and achieve better project outcomes. The potential for AI to drive innovation in the AEC industry is vast, inspiring a new era of technological advancement and efficiency.

Along with these contributions that artificial intelligence will provide to project management, there has been a notable rise in research on the application of artificial intelligence in the construction industry. It is evident from the literature that the use of AI techniques in the construction industry has become a trending topic.

3. Research methodology

This section outlines the research approach for the study, emphasizing the systematic and rigorous nature of the author's methodology. The author considered the approach based on the recommendations made by Kitchenham and Charters [22]. The case investigation 'Software Scent Detection Techniques: A Systematic Literature Review (SLR) '[23] served as a crucial tool for establishing the systematic literature review process we employed to investigate earlier research on our chosen subject, 'The Impact of Artificial Intelligence (AI) on Construction Projects.' Following the SLR methodology, our inclusion criteria guided the selection process, prioritizing papers with relevant abstracts and titles, while our exclusion criteria methodically eliminated irrelevant articles.

A systematic search across reliable resources guarantees a thorough assessment of the content. The author learned about standard procedures to conduct this detailed research, which gave me a solid basis for the subsequent projects. A paper titled "Software Smell Detection – Systematic Literature Review" [24] influenced the initial portion of the study as it provided knowledge about software smells and a solid basis for the author's work. The report utilized these data sources for our investigation: Electrical and Electronic Engineers (IEEE), Google Scholar, and ACM.

The study's topic, 'The Impact of Artificial Intelligence (AI) in Construction Projects,' was closely related to specific papers in which the author thoroughly examined the library's resources. The following research questions were formulated to guide our study and provide further details on the fields of study connected to our chosen topic and how the inclusion and exclusion criteria are implemented. These questions, carefully crafted to focus our search and ensure we only included relevant papers in our review, played a crucial role in shaping the direction of our study.

3.1. Search Process

The selections are limited to studies between 2010 and 2024, constituting the current literature. The databases used for searching are IEEE Xplore Digital Library, Web of Science (WOS), and Scopus. The keywords used for searches are "**AI** Project Management," "machine learning," "artificial intelligence," "AI-powered Construction," "Machine Learning," and "construction project management.".

The author's approach to selecting studies was to define the parameters for inclusion or exclusion criteria, determine the impacts of artificial intelligence applications and integration on construction projects, and select the trends and quality of the studies. The study's time frame was restricted to between 2010 and 2024 to keep up with current studies, with older studies being excluded under the assumption that they may be insufficient or irrelevant. The exclusion criteria were also applied to studies not closely related to the subject matter. Figure 2 illustrates the literature search approach.

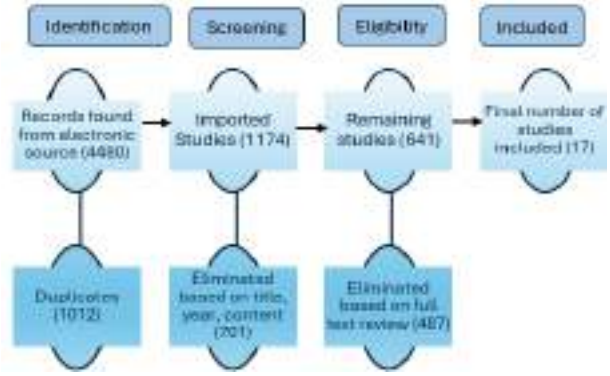


Figure 2: Literature search process

4. AI Applications in different construction sectors.

Artificial Intelligence (AI) is transforming the construction industry from the ground up, primarily by significantly improving productivity. It does this by streamlining processes and resource allocation using sophisticated algorithms and machine learning. This enables precise task scheduling, anticipation of potential hazards, and wiser resource allocation, all of which contribute to shorter project turnaround times and fewer delays. In addition, AI-powered Building Information Modelling (BIM) solutions provide more intelligent designs, collision detection, and thorough simulations, which expedite the planning and design stage and reduce mistakes. The main benefit of AI in the construction industry is its ability to significantly improve

productivity, thereby transforming the sector into many aspects.

AI-driven technologies are revolutionizing and significantly enhancing safety protocols in the construction industry. Using computer vision algorithms to examine data from sensors, drones, and cameras, AI can identify potential risks on construction sites, thereby improving safety procedures and preventing accidents. Furthermore, AI's ability to predict equipment faults makes it easier to perform maintenance, ensuring prompt repairs and reducing downtime. The role of AI in selecting environmentally friendly materials and optimizing energy use is also crucial, as it is leading to the adoption of greener and more sustainable building techniques. Incorporating AI is changing and significantly improving the construction industry, making operations safer, more environmentally friendly, and more efficient. Examples include autonomous cars and robotics performing mundane jobs and data-driven insights altering decision-making.

AI is revolutionizing and plays a pivotal role in waste reduction, safety, and project management in the construction industry. It is used in advanced software like Procore and Autodesk BIM 360 for resource allocation, scheduling, and communication. AI-driven wearables, computer vision systems, and IoT sensors enhance safety measures, while AI-powered trash tracking and recycling technologies improve waste management and recycling effectiveness. The comprehensive impact of AI in these areas is transforming the construction industry, making it more efficient and sustainable.

4.1. AI tools in construction

Several artificial intelligence (AI) solutions have been developed expressly for the construction sector, focusing on waste management, safety protocols, and project management. Here are a few instances:

Procore: This AI-powered tool manages documents, tracks timelines, keeps an eye on budgets, and streamlines communication to assist with project management. It provides insights to streamline operations and use machine learning algorithms to forecast possible project hazards.

Building information modeling (BIM): This tool is made easier using Autodesk BIM 360, which offers a centralized platform for project teamwork with artificial intelligence (AI) systems that examine data from several sources to facilitate better decision-making, identify conflicts and optimize construction scheduling.

Smartvid.io: This program uses artificial intelligence (AI) for safety to identify possible dangers and hazards by analyzing photos and videos taken on-site. By proactively addressing problems, it can identify patterns in safety occurrences and contribute to reducing accidents.

Predictive Solutions: This platform uses machine learning to forecast possible safety events by examining past data, spotting trends, and spotting patterns that might result in mishaps. It makes it possible to implement preventative safety measures.

Eco Domus: This technology optimizes waste management during construction by fusing AI and BIM. It follows the life cycle of building materials and analyses data to reduce material waste, allowing for effective recycling and reuse techniques.

4.2. Machine learning

The creation and application of computer systems that may be specifically intended to learn from prior data or experience to model, control, or forecast using statistical approaches is known as machine learning. Among the machine learning techniques are: (A) Supervised machine learning: this area of research focuses on how computers decide what to do after learning from labeled datasets or pairs of input and intended output. It is divided into regression and classification categories. (B) Unsupervised Machine Learning: this approach teaches machines the fundamental organization seen in unlabeled datasets. Techniques for dimension reduction and grouping are divided into two categories. (C) Reinforcement learning (RL) is the process of "learning a mapping from situations to actions to maximize a reinforcement signal or scalar reward." This computational method involves deducing knowledge from the results of interactions with the surroundings; also, (D) Deep Learning: this is the most advanced machine learning methodology available today, and it has been shown to provide predictions that are more accurate than those made using traditional methods.

4.3. Automated Planning and Scheduling

In artificial intelligence, known as planning, activities are carefully chosen and arranged according to their anticipated results to enable intelligent systems to accomplish specified goals or objectives. Determining plans and allocating the time and resources required to achieve the intended purposes using the available resources constitute scheduling. Planning and scheduling techniques create complicated application solutions that better meet user requirements and issue limitations. Due to its complexities, expense, and duration, planning is employed only in circumstances where its advantages surpass the disadvantages. Standard techniques and algorithms include genetic algorithms, search approaches, and optimization techniques for planning and scheduling.

4.4. Robotics

Robots are highly automated machines that do real-world physical tasks. Engineering that crosses many disciplines is called robotics, and it deals with creating, building, and maintaining robots and other computer programs that simulate real-world human movements.

Robots are meant for highly specialized activities; they don't always take on human-like forms but adopt shapes best suited to their purposes. They communicate with their surroundings using sensors and movements. The system outlined how reinforcement machine learning issues make up the bulk of robotics learning difficulties.

4.5. Safety Measures

Real-time Monitoring: AI-driven cameras and sensors monitor building sites around the clock. They enable quick response by identifying safety issues, illegal individuals, or risky behaviors.

Wearables and IoT Devices: To ensure worker safety and avoid accidents, artificial intelligence (AI) analyses data from wearables and IoT devices worn by workers to monitor vital signs, movement patterns, and tiredness levels.

5. Case Studies

Case Study 1: Integrating AI-Based Tool to Streamline Workflows

The case study highlights the integration of AI-based tools such as "Procore Copilot," which aims to improve communication and provide actionable insights across construction projects.

Procore's Copilot AI is a game-changer in the construction industry, addressing critical issues such as labor shortages and manual data processing. By automating tasks and accelerating data access, it expedites decision-making. The suite of AI tools, including Procore Copilot, AI Locations, and Procore Maps, revolutionizes project management by providing vital insights and seamlessly integrating with popular collaboration tools.

Procore's AI-driven tools, unveiled at the 2024 Innovation Summit, are designed to make project data more accessible and user-friendly for construction teams. Procore Copilot's integration with Microsoft Teams allows for natural language interaction with project data, making it easy and intuitive to access information such as Requests for Information (RFIs) and submittals. AI Locations simplify project organization by automatically generating location lists from project drawings, making it easier for teams to structure and locate items. Procore Maps introduces spatial tracking, enabling users to view photos mapped by location and filtered by date, providing a clear visual overview of project progress and facilitating workflow optimization.

Procore's AI implementation has not just improved project management, it has revolutionized it. By automating tasks such as location tracking and submittal management, Procore's AI tools have significantly reduced the time required for manual data entry, thereby enhancing workflow efficiency. The integration of Procore Copilot with Microsoft Teams has streamlined communication, enabling teams to access and discuss project information seamlessly. Procore Maps has enhanced data utilization by visualizing work progress with mapped photos, aiding decision-making and reducing delays. This efficiency gain is a testament to the potential of AI in the construction industry.

Procore's AI advancements, while promising, come with their own set of challenges. These include data privacy and security concerns, especially with third-party integrations like Microsoft Teams. Safeguarding sensitive construction project data is crucial for user trust. Additionally, users need training to adopt new AI tools effectively. Developing advanced features like Procore Copilot's natural language processing involves technical complexities and significant resources. By being aware of these challenges, the construction industry can better prepare for the future of AI.

The advantage

Procore Copilot integrates directly into Microsoft Teams, allowing users to access project data and context in natural language. Its features, such as AI Locations and Procore Maps, streamline visual data, enabling users to monitor and optimize project progress more efficiently. Also, IProcore uses AI to surface critical insights from Requests for Information (RFIs) and submittals, helping project teams make data-informed decisions.

Challenges and Barriers

While beneficial, Procore's AI advancements pose challenges. Data privacy and security are crucial, especially with third-party integrations like Microsoft Teams, to safeguard construction project data and maintain trust. Advanced AI tools also require user training for effective adoption. Developing features like Procore Copilot's natural language processing also demands substantial resources and expertise.

Case Study 2: BIM 360 and Smartvid.io AI

Autodesk BIM 360 introduction of AI agents has the potential to elevate the capabilities of Autodesk BIM 360 to enhance decision-making and operational efficiency.AI agents can transform project management and automation by integrating with Autodesk BIM 360.

AI integration in Autodesk BIM 360 can greatly enhance workflows. Though the platform lacks built-in AI, integrating AI solutions can improve efficiency by analysing data, automating tasks, and offering insights through analytics and machine learning, thus supporting smarter project management decisions. Automation for Architectural and Design Efficiency The key examples of AI-Driven automation include:

Automated Document Management: AI can process project documents and organize them for easier access., **Predictive Risk Management**: By using historical data, AI can identify potential project risks

based on previous projects and **Enhanced Collaboration**: AI agents can improve communication by summarizing project updates and informing team members of changes.

AI agents simplify workflows in complex settings like construction or architecture, accelerating processes and reducing manual effort.

6. BIM-AI Framework for Construction Integration

The framework, which focuses on data interoperability, AI algorithms, and application areas, is designed to improve efficiency, reduce errors, and optimize building performance across the project lifecycle. Figure 5 illustrates a continuous process of capturing the project team's knowledge into BIM and AI practices in the four (4) toolbox knowledge areas. The framework's use case not only presents processes and corresponding appropriate BIM and AI tools but also emphasizes the adaptability of the framework to different phases of the project life cycle, making it a versatile and attractive solution for stakeholders.

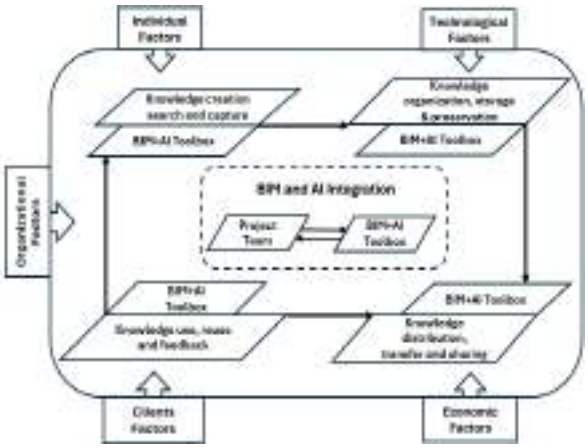


Figure 5. Conceptual knowledge framework for the integration of knowledge into BIM-AI practices

The framework is further subcategorized into four (4) classifications:

1. Data Interoperability and Preparation:

Standardized Data Formats: Establish clear protocols for data exchange between BIM models and AI systems, using Industry Foundation Classes (IFC) or other relevant standards.

Data Extraction and Transformation: Develop methods for extracting relevant data from BIM models (geometrical, semantic, and parametric) and transform it into formats suitable for AI algorithms.

Data Quality and Management: Ensure data accuracy, consistency, and completeness to support reliable AI analysis.

2. AI Algorithms and Applications:

Machine Learning: Utilize machine learning algorithms for tasks such as clash detection, cost estimation, schedule optimization, and performance simulation.

Image Recognition and Processing: Employ AI algorithms for visual inspection, quality control, and progress monitoring on-site.

Predictive Analytics: Leverage AI for predicting potential problems, optimizing resources, and making informed decisions throughout the project lifecycle.

3. Application Areas:

Design Optimization: Use AI to explore design alternatives, optimize building performance, and identify sustainable solutions.

Construction Scheduling and Management: Automate scheduling tasks, predict delays, and optimize resource allocation using AI.

Cost Management: Analyze cost data, predict project costs, and optimize resource allocation using AI.

Quality Control and Safety: Implement AI-powered systems for visual inspection, safety hazard detection, and quality control.

Facility Management: Use AI to monitor building performance, optimize maintenance schedules, and improve energy efficiency.

4. Challenges and Mitigation Strategies:

Data Availability and Quality: Address data gaps and inconsistencies by implementing robust data management practices and standards.

Algorithm Accuracy and Reliability: Validate AI algorithms and ensure their accuracy and reliability before deploying them in real-world projects.

Interoperability Issues: Develop standardized data formats and protocols to facilitate seamless integration between BIM and AI systems.

Lack of Expertise: Invest in training and education to equip construction professionals with the necessary skills to use BIM and AI effectively.

Ethical Considerations: Address potential biases in AI algorithms and ensure responsible use of AI in construction.

7. Traditional Approach to AI process in construction sector Integration

Automated decision support and artificial intelligence (AI)-driven solutions empower you with real-time data insights, automating tasks like scheduling and documentation to bolster your decision-making process.

Improved collaboration cloud-based platforms allow teams to work together in real-time while accessing 3D models and reports, enhancing communication and helping with project management.

Real-time project visualization technologies provide

a dynamic, real-time perspective of the project's state, enabling more informed decision-making. They do this by utilizing 3D models and progress reports.

Data-driven resource allocation and machine learning algorithms analyze extensive data to find trends, which helps with targeted cost control plans and efficient resource allocation.

Efficiency compared to manual techniques, Procore and Autodesk BIM 360 considerably increase efficiency by streamlining procedures and providing real-time data. Waste management integration with AI undergoes a few steps to convert the scrap, decrease inefficiencies, and promote sustainability throughout project lifecycles, revolutionizing waste management in the construction industry. AI is essential for predictive analytics because it can estimate possible waste generation using past project data. As a result, less waste is produced overall, and proactive efforts to optimize material utilization and purchasing are made possible. Moreover, by precisely recognizing and classifying diverse waste products, AI-powered systems with cameras and sensors enable effective sorting and recycling procedures. Artificial intelligence (AI)-powered real-time monitoring systems monitor waste production on building sites and provide prompt decision-making to reduce trash accumulation. Complete lifecycle analysis is made possible by integrating AI with Building Information Modelling (BIM), which helps architects create buildings with the least amount of waste possible during their lifespan.

AI is crucial in project planning, from requirement gathering to plan selection and implementation. Leveraging AI can help choose the best plan, improving project efficiency. AI's benefits in project management include improved project planning, real-time project monitoring and alerts, enhanced efficiency, communication and collaboration, risk management, predictive analytics, and better decision-making. Drones, wearable tech, autonomous equipment, and 3D printers contribute to improved safety measures in construction. Drones enable efficient site inspections, real-time safety monitoring, and emergency response, reducing worker exposure to hazardous areas. Wearable tech provides real-time safety alerts, biometric tracking, and immersive training experiences. Autonomous equipment reduces worker exposure to high-risk tasks, minimizes errors, and operates 24/7, potentially reducing fatigue-related accidents. 3D printers allow for safer prototyping, reducing the need for workers to handle sharp materials. These technological advancements contribute to mitigating risks, enhancing monitoring capabilities, and reducing exposure to hazards in construction environments.

8. Results

Table 1: Results [3][25][26]

	Traditional method	AI methods
Staff evaluation	Project planning with long period involves in recruiting.	Automatic skill matching and team making.
Work assignment	Planning tools, team meetings, strategy making, business alignment, one-on-one meetings, time consuming.	Auto tasks assigned and auto availability detection as needed.
Performance monitoring	Qualitative and quantitative data feedback	Automated monitoring
Safety measure	Emergency Response Planning, Site Inspections and Hazard Identification, Safety Education and Training, and Personal Protective Equipment (PPE), etc.	Wearable Tech, 3D printers, autonomous equipment's and drones etc., can be used.
Cost analysis	Excessive purchase is made, and cost cannot be analysed prior to the time, Estimation Techniques.	No excess purchases are made, planned accordingly.
Waste management	Safe disposal and recycling methods are not adopted	Recycling and safe disposal of scrap
Project planning	Information issues, resource productivity, Gantt charts, CPM, scheduling, and resource allocation.	Planning accordingly, considering material availability, time and cost management.

9. Conclusion

The results in Table 1 demonstrate how artificial intelligence (AI) may transform building methods by tackling long-standing problems such as poor productivity, safety, and overspending. The arrival of artificial intelligence (AI) in project management software, such as Autodesk BIM 360 and Procore, is demonstrating uses that improve decision-making, optimize workflows, and allocate resources as efficiently as possible.

Furthermore, this report highlighted how AI is revolutionizing the construction industry. The study projects that increasing the integration of AI would improve operational efficiency while also requiring a shift in the qualifications and abilities of the workforce. Ultimately, the data supports more AI applications in construction projects despite difficulties. The argument for industry-wide adaptation is substantial, given the possible advantages of increased production, safety, and cost-effectiveness. To fully realize AI integration's potential and advance the construction sector toward a

more inventive and sustainable future, coordinated efforts to remove obstacles and develop knowledge in this area are essential.

Future research should explore AI's sustainability and dispute management domains, address resistance to technology adoption, and increase the availability of datasets for AI applications in construction project management.

References

[1] S.O Abioye, LO Oyedele, L Akanbi, 2021 - Journal of Building Engineering. Artificial intelligence in the construction industry: A review of present status, opportunities, and future challenges.

[2] E Dobrucali, S Demirkesen, E Sadikoglu, C Zhang, A Damci 2022 *Engineering, Construction and Architectural Management*. Investigating the impact of emerging technologies on construction safety performance.

[3] F. Niederman 2021, *Information Technology & People*: Project management: openings for disruption from AI and advanced analytics.

[4] B Delipetrev, C Tsinaraki, U Kostić - Joint Research, 2018, AI Watch Historical Evolution of Artificial Intelligence.

[5] Wang XL. Application of artificial intelligence in oil and gas industry. Mod Inf Technology 2017; 3:117−119.

[6] Shi T, Wu J. Application of Artificial Intelligence in Water Conservancy Project Management. Proceedings - 2021 2nd International Conference on Big Data and Artificial Intelligence and Software Engineering 2021;556−559.

[7] Bingöl K, Akan AE, Örmecioglu HT, Er A. Artificial intelligence applications in earthquake resistant architectural design: Determination of irregular structural systems with deep learning and Image AI method. J Fac Eng Archit Gazi Univ 2020; 35:2197−2209.

[8] Lu P, Chen S, Zheng Y. Artificial Intelligence in Civil Engineering. Math Probl Eng 2012;1−22

[9] Karan E, Asgari S, Mohammadpour A. Applying artificial intelligence within the AEC industry: Collecting and interpreting data. Construction Res Congr 2020; 7:809−818.

[10] Regona M, Yigitcanlar T, Xia B, Li RYM. Opportunities and adoption challenges of AI in the construction industry: a PRISMA review.

[11] Fulford R, Standing C. Construction industry productivity and the potential for collaborative practice. Int J Proj Manag 2014; 32:315−326.

[12] Kabirifar K, Mojtahedi M. The impact of engineering, procurement and construction (EPC) phases on project performance: A case of large-scale residential construction project. Buildings 2019; 9:15.

[13] Okudan O, Çevikbaş M, Işık Z. An exploratory study on the critical features of construction project planning software. Sigma J Eng Nat Sci 2022; 41:781−792.

[14] Liu J. Bayesian network inference on risks of construction schedule-cost. Proceedings - 2010 International Conference of Information Science and Management Engineering, ISME. 2010; 2:15−18.

[15] Kong F, Wu X, Cai L. International Symposium on Computational Intelligence and Design. 2008;21−24.

[16] Ensafi M, Alimoradi S, Gao X, Thabet W. Machine learning and artificial intelligence applications in building construction: present status and future trends. Constr Res Congr 2022;964−973.

[17] Sravanthi J, Sobti R, Semwal A, Shravan M, Al-Hilali AA, Alazzam MB. AI-Assisted Resource Allocation in Project Management. In: 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE). 2023. P

[18] Shi T, Wu J. Application of Artificial Intelligence in Water Conservancy Project Management. Proceedings - 2021 2nd International Conference on Big Data and Artificial Intelligence and Software Engineering 2021;556−559

[19] Afzal F, Yunfei S, Nazir M, Bhatti SM. A review of artificial intelligence-based risk assessment methods for capturing complexity-risk interdependencies: Cost overrun in construction projects. Int J Manag Proj Bus 2021; 14:300−328.

[20] Chenya L, Aminudin E, Mohd S, Yap LS. Intelligent risk management in construction projects: Systematic Literature Review. IEEE Access 2022; 10:72936-72954.

[21] Pan Y, Zhang L. Roles of artificial intelligence in construction engineering and management: A critical review and future trends. Autom Constr 2021; 122:103517.

[22] B Delipetrev, C Tsinaraki, U Kostić - Joint Research, 2018, AI Watch Historical Evolution of Artificial Intelligence.

[23] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436.

[24] Amjad AbuHassan, Mohammad Alshayeb, and Lahouari Ghouti. 2021. Software smell detection techniques: A systematic literature review. J. Softw. Evol Process 33, 3 (March 2021). https://doi.org/10.1002/smr.2320.

[25] A Srivastava, S Jawaid, R Singh, A Gehlot, SV Akram, N Priyadarshi, B Khan 2022 Advances in Civil Engineering, Imperative role of technology intervention and implementation for automation in the construction industry.

[26] M Bolpagni, I Bartoletti 2021- Proc. of the Conference CIB W78, researchgate.net Artificial Intelligence in the Construction Industry: Adoption, Benefits and Risks.

Integrating Ontology and LLMs for Diagnosis and Repair of Concrete Surface Defects

F. Bahreini[1] and A. Hammad[1]

[1]Concordia Institute for Information Systems Engineering, Concordia University, Canada
fardin.bahreini@concordia.ca, amin.hammad@concordia.ca

Abstract –

This paper explores the integration of a novel Ontology for Concrete Surface Defects (OCSD) with a Large Language Model (LLM), specifically GPT-4o (omni), to enhance defect diagnosis and repair strategies in concrete structures. While LLMs independently offer significant reasoning and natural language capabilities, this study demonstrates the value of combining their interpretative power with the structured knowledge representation provided by OCSD. By enabling adaptive reasoning, where the LLM relies on the ontology's domain-specific relationships and thresholds, and updates its conclusions for specific defect types, the system flexibly adjusts its decision-making based on context. This integration improves diagnosis accuracy, reasoning transparency, and decision-making efficiency. The proposed method is validated through a case study, highlighting the synergy of OCSD and GPT-4o in addressing challenges in defects diagnosis and repair.

Keywords –

Large Language Models (LLMs), GPT-4o, AI-based Reasoning, Concrete surface defects, Ontology

1 Introduction

The durability and safety of concrete structures depend on accurate and timely diagnosis of defects such as cracks, spalling, and delamination. These defects compromise the structural integrity and pose significant maintenance and financial challenges. While traditional methods for defect diagnosis rely heavily on manual inspection and expert judgment, advancements in artificial intelligence (AI) have introduced powerful tools for enhancing this process. On the other hand, ontologies formalize domain knowledge into structured representations, enabling systematic diagnosis and informed decision-making [1–4]. Ontology for Concrete Surface Defects (OCSD), developed in our previous work [5], provides a comprehensive framework for modeling knowledge about inspection, diagnosis, and 3R (Repair, Rehabilitation, and Replacement) processes for concrete defects. However, ontologies, by their static nature, may fall short when applied to context-specific scenarios. Recent advancements have demonstrated that integrating domain-specific ontologies with AI models enhances interpretative accuracy and contextual reasoning [6, 7]. Large Language Models (LLMs), such as GPT-4o (omni), have emerged as transformative AI systems capable of interpreting complex, unstructured data and providing natural language insights [8–11]. LLMs possess exceptional ability in analyzing textual data, which triggered the opportunity of using ontologies alongside such systems. On the other hand, ontologies provide a robust foundation for mitigating challenges like hallucinations in LLM reasoning, ensuring alignment with domain-specific standards [10].

This study presents an integrated framework for diagnosis and repair of concrete surface defects by combining OCSD [5] with the interpretative capabilities of LLMs. By leveraging the ontology's explicit semantic relationships, this integration ensures logical consistency, alignment with domain standards, and a systematic approach to defect classification, causal analysis, and repair recommendations. This synergy enhances both diagnosis accuracy and reasoning transparency, effectively addressing challenges in defect analysis through ontology-driven reasoning and advanced LLM-based interpretative capabilities. The objectives of this study are: (1) to demonstrate how integrating ontology knowledge enhances the LLM reasoning for concrete defect diagnosis and repair; and (2) to showcase the contributions of this synergy in improving response clarity and actionable recommendations. Through a detailed case study focusing on spalling defects, this research highlights the transformative potential of combining ontological frameworks with advanced LLMs, addressing critical challenges in structural inspection and repair. By bridging the gap between structured knowledge representation and adaptive reasoning, this approach provides a robust solution for structural diagnosis.

The paper is structured as follows. Section 2 reviews related works. Section 3 explains the proposed method.

Section 4 presents the implementation. Section 5 presents the testing and evaluation. The last section provides the conclusions and outlines potential areas for future research.

2 Literature Review

Tupayachi et al. [1] introduced AI-powered ontology construction to enhance urban decision-support systems, demonstrating how integrating structured ontologies with LLMs improves reasoning over complex data. Saeedizade and Blomqvist [2] highlighted how semantic frameworks, including ontologies, enhance LLM reasoning layers by formalizing domain-specific requirements. Similarly, Momcilovic et al. [3] illustrated that formal ontologies guide AI reasoning processes, ensuring logical consistency. Existing studies on LLM and ontology integration can be grouped into three primary categories: **(1) LLMs and ontologies in engineering and analytics:** Several works have focused on the integration of LLMs and ontologies in engineering and analytics. For instance, Allemang and Sequeda [12] used ontology-based query checks to improve LLM accuracy in question answering, addressing hallucinations, while Song and Yoon [13] employed ontology-guided LLMs in building performance simulation. Additionally, Wang et al. [7] integrated ontological reasoning into LLM workflows for predictive maintenance and demonstrated how ontologies enhance LLM contextual understanding. Nakajima and Miura [14] incorporated ontological reasoning into LLM workflows for robotic tasks and showcased their ability to improve precision in robotic reasoning. Furthermore, Babaei Giglou et al. [15] investigated LLM-driven ontology learning and evaluated LLMs in extracting structured knowledge. Baldazzi et al. [16] explored domain-specific LLM fine-tuning and emphasized ontological reasoning as a foundation for fine-tuning LLMs in financial tasks. Additionally, Margineda Mangues [17] utilized LLMs with ontologies for procedural content generation, Doumanas [18] applied iterative ontology-driven learning in engineering analytics, and Liu et al. [19] combined domain-specific ontologies with LLM reasoning for dynamic fault diagnosis. **(2) LLMs and ontologies in conversational AI:** Some studies have demonstrated the role of ontologies in enhancing conversational AI through the integration with Large Language Models (LLMs). For instance, Zhang et al. [4] employed knowledge graphs to improve conversational workflows with LLMs, while Palagin et al. [9] used ontology-driven structured prompts to enhance chatbot adaptability. These works illustrate the utility of ontologies in providing context and reducing ambiguity. **(3) LLM for ontology quality assurance:** Research has also explored the application of LLMs in ontology quality assurance. For example, Meyer et al. [8] utilized LLMs for engineering tasks within knowledge graphs, and Tsaneva et al. [20] verified ontology restrictions to address quality issues. Both studies highlight the importance of ontologies in structuring knowledge and ensuring the integrity and accuracy of AI-driven systems.

These studies emphasize that combining domain-specific ontologies with advanced LLM models can provide structured guidance, improving reasoning accuracy, domain alignment, and context-aware decisions. They highlight the value of ontologies as a stable backbone that ensures consistent semantics and prevents hallucination. Our approach distinctively employs specific ontology thresholds to enable flexible reasoning using LLM tailored to structural defect diagnosis and repair scenarios.

3 Proposed Method

LLMs like GPT-4o utilize both their pre-trained knowledge and domain-specific inputs to generate contextually relevant outputs. In this approach, the ontology serves as a structured repository of domain-specific knowledge, providing explicit relationships and predefined thresholds for classification, causal analysis, and repair recommendations. These thresholds standardize defect analysis, ensuring consistency and domain relevance. When interacting with the ontology, the LLM uses its language understanding capabilities to interpret relationships, contextualize attributes, and fill gaps where the ontology lacks explicit details. For example, the LLM complements the ontology by interpreting preventive measures not explicitly defined as a class, leveraging its understanding of related concepts like the 3R approaches. This combination ensures that the LLM's outputs are both guided by the ontology and enriched by its generalized knowledge, improving accuracy, relevance, and interpretability.

The method has the following main modules: **(1) Ontology-driven reasoning module:** Leveraging the OCSD ontology [5], this module maps defect attributes to semantic relationships for classification, causal analysis, and severity assessment; and **(2) Defect analysis and recommendation module:** Integrating the semantic attributes and ontology reasoning with GPT-4o, this module generates actionable insights, including specific repair strategies using the 3R processes. Adaptive reasoning is the ability of an AI system, like an LLM, to flexibly adjust its decision-making based on context and evolving inputs. In this framework, when a user specifies a defect type, the ontology filters and provides the most relevant triplets (i.e., subject–predicate–object relationships), guiding the LLM's reasoning with authoritative standards. The LLM then interprets these insights, filling any gaps with its pre-trained knowledge and updating results as new data or feedback emerge. Through this context-aware interaction,

the system demonstrates true adaptive reasoning, ensuring accurate and tailored guidance for complex, real-world scenarios.

3.1 Ontology-driven Reasoning Module

The OCSD ontology provides a semantic framework for mapping defect attributes to structured classes and relationships. By defining explicit thresholds and categorizing defects into severity levels (e.g., *Light Spalling*, *Medium Spalling*, *Severe Spalling*), the ontology guides logical classification, causal analysis, and context-specific recommendations. Leveraging these ontology-derived relationships ensures that each defect is analyzed consistently and aligned with authoritative domain standards.

A key aspect of this process is the use of dynamically inserted variables, which are generated by querying the ontology for the most relevant defect-specific information. Rather than relying solely on user inputs, the ontology-driven reasoning module filters and selects the appropriate triplets related to the specified defect type. These ontology-derived details, such as severity thresholds, corresponding causes, or recommended repair pathways, are then programmatically inserted into the reasoning prompt at runtime. For example, if the user specifies a spalling defect, the ontology returns the relationships that define the severity classifications, potential root causes, and applicable 3R strategies leveraging the OCSD for structured and domain-specific analysis. These ontology-derived variables (i.e., *{ontology_info}*) appear as placeholders in the prompt and are replaced by the actual, contextually relevant data before the LLM is called. By doing so, the LLM's outputs reflect the most up-to-date, domain-specific information, enhancing the accuracy, relevance, and credibility of the resulting recommendations. The details of the model prompts are provided in Table 1. This table compares the baseline model and the OCSD-enhanced model prompts. The prompt of the baseline model relies only on defects attributes. The OCSD-enhanced model uses the same prompt as the baseline model but adds the provided ontology relationships.

Iterative reasoning further refines the analysis by allowing the system to repeatedly engage with the ontology and LLM as new data or feedback emerge. If the initial recommendations need adjustment, due to changes in defect attributes, updated inspection results, or evolving maintenance strategies, the ontology-driven reasoning module retrieves the related relationships and thresholds. The LLM then reinterprets this ontology-derived information, ensuring that guidance remains current and context aware. In addition to textual outputs, the approach incorporates interactive visualization to depict the relationships and semantic pathways underlying each recommendation. By visually representing ontology-derived nodes (e.g., defect classes) and relationships, the user gains a clearer understanding of how the ontology informs LLM outputs. This graphical representation makes the reasoning process more transparent and intuitive, helping the user rapidly understand why certain causes are inferred or why specific repair actions are proposed.

3.2 Defect Analysis and Recommendation Module

Building on the ontology-driven reasoning described in Section 3.1, this module employs GPT-4o to produce scenario-specific recommendations. By processing the severity classification and ontology-derived details, the model confirms the defect's classification, identifies underlying causes, and suggests actions aligned with authoritative standards. The defined 3R interventions guides the selection of repair strategies, ensuring that each recommendation is contextually relevant and rooted in established domain knowledge.

In this module, the structured prompt integrates the previously retrieved ontology relationships, severity classifications, and any available defect attributes. This

Table 1. Comparison of prompts of baseline model and OCSD-enhanced model

Baseline model prompt	Enhanced model with OCSD prompt
You are a structural engineer specializing in concrete defects. You have been provided with the following information: Defect Description: - Type: '{defect_description}' Defect Attributes: {attributes_info} Using the provided defect attributes, please perform the following tasks: 1. Severity classification 2. Diagnosis of potential causes. 3. Recommendation of remedial actions. 4. Preventative measures Please ensure that your analysis is thorough and that you logically integrate the defect attributes to support your conclusions. Present your response in a clear and structured manner.	You are a structural engineer specializing in concrete defects. You have been provided with the following information: Defect Description: - Type: '{defect_description}' Defect Attributes: {attributes_info} **Relevant Ontology Relationships.** **{ontology_info}** Using the provided **ontology relationships** and defect attributes, please perform the following tasks: 1. Severity classification 2. Diagnosis of potential causes. 3. Recommendation of remedial actions. 4. Preventative measures Please ensure that your analysis is thorough and that you logically integrate the **ontology information** and defect attributes to support your conclusions. Present your response in a clear and structured manner.

combined input enables GPT-4o to generate logically consistent outputs tailored to the defect's unique characteristics. The resulting analysis provides clear justifications for proposed repair actions, linking them directly to the identified causes and ensuring a transparent, actionable reasoning process.

4 Implementation

This section details the practical implementation of a prototype system based on the proposed method developed using Python. The application integrates various libraries and tools to process input data and interact with the ontology and LLM. It uses *rdflib* to parse and query the ontology, *networkx* and *pyvis* for building and displaying interactive relationship graphs, and *OpenAi* for calling the GPT-4o model's API. The developed prototype system based on the method supports automated data ingestion, dynamic ontology querying, and adaptive reasoning calls to GPT-4o, ensuring an integrated and reproducible environment for experimentation and analysis. Figure 1 shows the developed user interface.

In this implementation, supplementary image analysis was conducted using GPT-4o API for extracting defect attributes. The image of a concrete surface with dimensions of $1m \times 2m$ was uploaded into the developed application for analysis. The GPT-4o model was accessed through the OpenAI API and tasked with identifying the major defect type, classifying individual defects, and extracting their attributes considering the surface dimensions. Figure 1(a) shows an example of an image that was fed to the system. Five spalls have been detected, and their attributes have been extracted automatically. Figure 1(b) shows the extracted defects attributes from the image analysis. For instance, Defect 4 has the width of about 20 cm, height of about 2 cm, and depth of about 2 cm, with features including being the widest defect, and having uneven erosion and discoloration. The defect attributes are used as input for the prompt as explained in Section 3.1. These attributes could also be derived from point cloud semantic segmentation results, as demonstrated in prior work [21], or obtained through on-site inspections and entered to the system by the user as shown in Figure 1(c). As the type of the detected defect is spalls, Figure 1(d), enlarged in Figure 2, shows only 38 classes and 9 instances for the threshold's related to spalls, which were automatically extracted from the complete ontology which has a total of 335 classes.

5 Testing and Evaluation

Table 2 provides two example comparing responses from the baseline and enhanced models for two defects (Defects 1 and 4). Condition assessment utilizes predefined severity thresholds encoded in the OCSD, which are compared with the defect dimensions, ensuring consistency in defect evaluation. Furthermore, by leveraging the OCSD ontological relationships, our reasoning module identified plausible cause-and-effect relationships between specific defect characteristics and

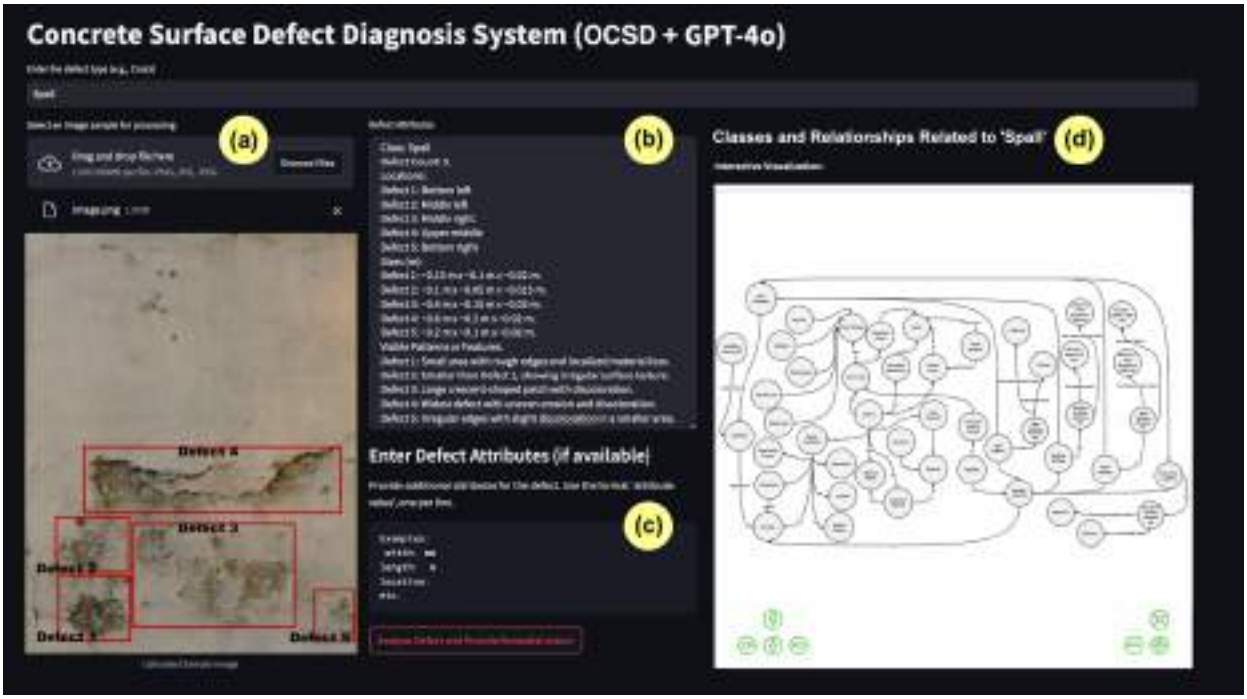


Figure 1. User interface of the prototype system

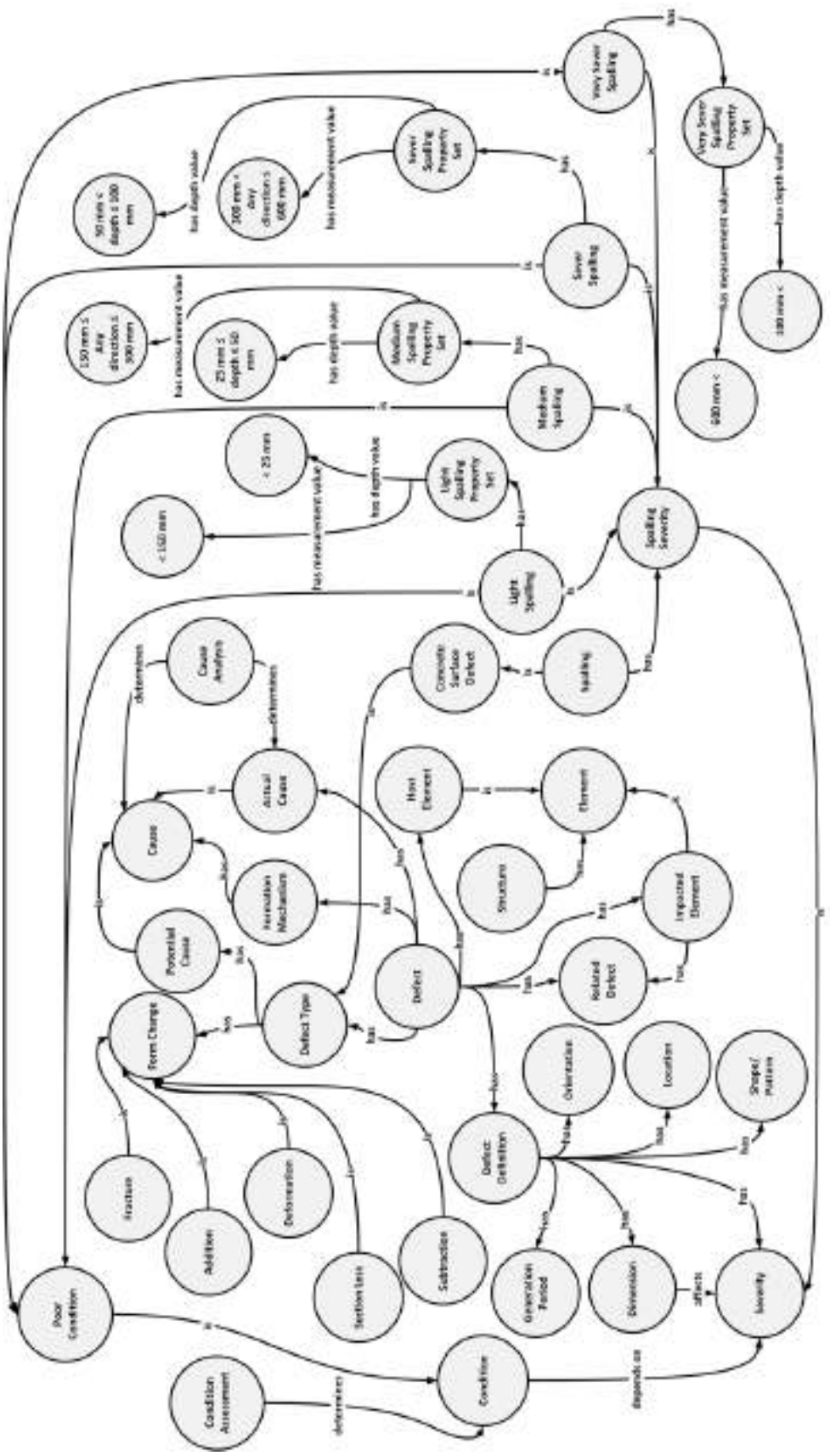


Figure 2. An example of interactive visualization of OCSD-derived thresholds for condition assessment of spalling

their underlying factors. For Defect 1, the baseline model classified Defect 1 as low severity due to its irregular texture and smaller size, attributing it primarily to minor surface erosion. However, the OCSD-enhanced model identified it as *medium spalling*, emphasizing moisture ingress and chemical exposure as contributing factors. By leveraging ontology-driven thresholds, the enhanced model provided a more refined diagnosis, recommending surface treatments aligned with spalling severity classifications. This result, representative of the model's performance across five spalling defects, illustrates how incorporating authoritative knowledge sources leads to more contextually accurate and actionable guidance. The enhanced model, informed by ontology-derived insights, provided a precise classification for Defect 4 (*severe spalling*) rather than a generic label, identified clearly documented causes, and recommended standards-based repair and preventative measures. For example, severe spalling in Defect 4 indicated more advanced chemical degradation, potentially from de-icing agents, and the impacts of freeze-thaw cycles. Informed by the ontology-derived severity classifications and guided by the 3R interventions, GPT-4o provided well-structured repair strategies for each identified defect. For Defect 4, characterized by severe spalling, the system advised partial or complete replacement, reflecting the severity of material loss and structural compromise. Preventative strategies here focused on advanced sealants and thorough, regular inspections to forestall recurrence. Across all defect cases, GPT-4o's recommendations demonstrated logical consistency with the severity levels, causal insights, and the repair and preventative strategies defined by the ontology.

The effectiveness of integrating the ontology-based framework with LLMs in diagnosing and repairing spalling defects is evaluated focusing on four key metrics: severity classification, cause identification, remedial actions, and preventative measures. Partial matches were counted to accommodate natural variations in phrasing, ensuring a comprehensive assessment. These metrics and their associated terms are drawn from recognized standards and guidelines, as defined in *Concrete Structures Protection, Repair and Rehabilitation* [23] and the *Ontario Structure Inspection Manual (OSIM)*

Table 2. Comparison of responses of baseline model and OCSD-enhanced model for Defects 1 and 4

Baseline model response	Enhanced model with OCSD response
Defect 1 - Location (Bottom-left), Smaller Area with Rough Edges *1. Severity Classification:* Low Severity: The defect has irregular surface texture, indicating initial material surface degradation. *2. Potential Causes:* Surface Erosion: Possibly due to weather exposure or inadequate curing. Concrete Quality: Inadequate surface finishing leading to early wear. Reasoning: Smaller size with irregular texture suggests less severe environmental action, likely a surface treatment issue *3. Remedial Actions:* Surface Sealing: Apply a protective surface sealant to prevent moisture ingress and further surface degradation. Rationale: Sealing restores the integrity and protects against environmental exposure. *4. Long-term Preventative Measures:* Proper Curing: Ensure adequate curing times to enhance surface strength and durability. Surface Coatings: Use protective coatings to shield surfaces from weather and chemical attacks.	**Defect 1:** Location: (Bottom-left) Size: Medium size with irregular surface texture and discoloration. *1. Severity Classification: MediumSpalling* Justification: Medium defect size and discoloration indicate a severity higher than LightSpalling. OCSD's MediumSpalling class involves more notable material loss and visual deterioration. *2. Potential Causes Inherited from OCSD:* (a) Moisture ingress with discoloration: Medium-severe discoloration effects. (b) Chemical exposure: Such as chloride ingress from de-icing agents. *3. Remedial Actions (3R Approach):* Surface treatments or coatings: Justification: Mitigation needed for discoloration and moderate material loss. OCSD guidelines suggest comprehensive surface treatment for medium spalling. *4. Long-term Preventative Measures:* (a) Improved concrete mixes: Utilize mixtures that resist spalling agents. (b) Surface cleaning protocols: Avoid accumulation of harmful substances. *5. Summary:* Medium spalling needs enhanced materials addressing both aesthetic and structural perspectives, guided by OCSD, while providing clarity on moderate deterioration and avoiding overly extensive solutions.
Defect 4 - Location (Upper-middle), Widest, Uneven Erosion, Discoloration *1. Severity Classification:* High Severity: With uneven erosion and discoloration, it heavily impacts the structural performance. *2. Potential Causes:* (a) Severe Moisture Exposure: Possible long-term exposure to water and chemical agents. (b) Aggregate Reaction: Alkali-silica reaction could cause both erosion and discoloration. Reasoning: The extent and characteristics suggest significant chemical and physical reactions over time. *3. Remedial Actions:* (a) Comprehensive Repair: Extensive removal, including damaged rebar sections, repair or replacement of rebar, and concrete restoration. (b) Rationale: Ensures safety by restoring both external and internal structural components. *4. Long-term Preventative Measures:* (a) Enhanced Concrete Mixes: Use mixes resistant to alkali-silica reactions and high moisture environments. (b) Improved Design Drainage: Ensure design includes proper drainage systems to minimize water retention on the surface.	**Defect 4**: Location: (Upper-middle) Size: Widest defect with uneven erosion and discoloration. *1. Severity Classification:* Severe Spalling Justification: The largest area with significant erosion matches OCSD's Severe Spalling definitions due to extensive material loss and discoloration. *2. Potential Causes inherited from OCSD:* (a) Aggressive chemical attack: May be from industrial exposure or pollutants leading to severe erosion. (b) Freeze-thaw cycles: Intensifies with large and deep spalling areas. *3. Remedial Actions (3R Approach):* Partial or complete replacement: Justification: Severe Spalling defects compromise integrity, necessitating comprehensive intervention. Aligns with OCSD suggesting replacement for severe cases. *4. Long-term Preventative Measures:* Advanced sealants and coatings: Prevent further degeneration. Regular, in-depth inspections: Address early signs of similar severe deterioration. *5. Summary:* Severe defects need major structural interventions guided by OCSD, providing well-planned steps for long-term effectiveness over ad-hoc methods.

[24]. By using widely accepted terminology drawn from these references, the evaluation ensures that the chosen terms reflect common engineering practice, rather than proprietary ontology constructs. Specifically, for severity classification, the terms *Light Spalling*, *Medium Spalling*, and *Severe Spalling* were applied. To represent common underlying causes, *moisture penetration*, *freeze-thaw cycles*, *chemical attack*, and *reinforcement corrosion* were selected. For remedial actions, *patch repair*, *surface coating*, and *corrosion inhibitors* were used, while *waterproofing*, *high-performance concrete mixes*, and *regular inspections* served as preventative measures.

Figure 3 shows a bar chart summarizes how frequently each model correctly employs the established key terms. Along the horizontal axis are the four metrics, and the vertical axis indicates the count of recognized terms. The enhanced model's higher bars in categories like severity classification and remedial actions emphasise its stronger alignment with domain standards.

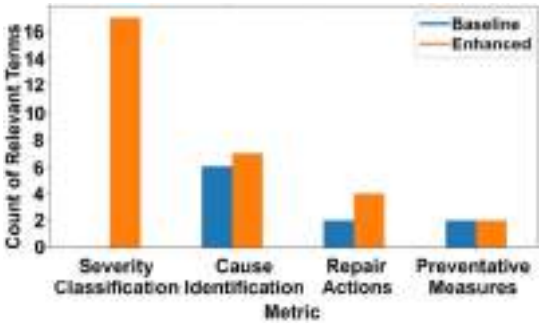


Figure 3. Frequency of correctly identified terms across the four defined metrics

The radar plot shown in Figure 4 provides a holistic view by plotting each metric on a different axis, forming a polygon for each model. The enhanced model's polygon extends further along all axes, especially severity classification, indicating a more balanced and comprehensive understanding of the key concepts.

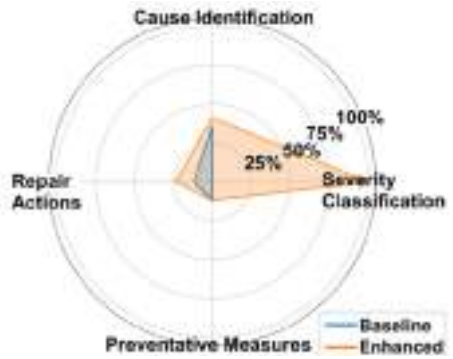


Figure 4. Normalized visual comparison of the models across the four defined metrics

Overall, by grounding the evaluation in accepted industry references [17] and OSIM, and using figures that clearly illustrate comparative performance, the analysis demonstrates that ontology integration significantly enhances the LLM's ability to produce structured, accurate, and actionable responses for spalling defect diagnosis and repair.

6 Conclusions and Future Work

This paper introduced an integrated framework for diagnosing and analyzing concrete surface defects by combining ontology-driven reasoning with the advanced interpretative capabilities of an LLM, GPT-4o. The proposed method is implemented in a prototype system integrating OCSD with GPT-4o API. By leveraging GPT-4o for ontology-based reasoning, the system provided a comprehensive and context-aware workflow for defect classification, causal analysis, and repair planning. A case study involving spalling defects demonstrated the efficacy of this approach. The proposed framework offers several advantages over traditional methods: (1) The use of OCSD ensures that diagnosis outputs are grounded in structured domain knowledge, enhancing the reliability of the results; and (2) The integration of defect attributes and ontology relationships enables specific and context-aware repair strategies. The evaluation results indicate that integrating domain-specific ontologies, such as the OCSD, significantly enhances the depth, accuracy, and actionability of LLM-generated outputs for defect diagnosis and repair.

Despite the promising results, the following limitations were identified: (1) While the OCSD provides a structured framework for defect classification, it may not comprehensively cover all defect types, causes, and repair strategies. (2) The quality of the LLM's outputs depends on prompt clarity and data specificity, making it susceptible to errors if attribute extraction is incomplete or ambiguous. (3) Reliance on the OpenAI API for GPT-4o introduces computational costs, and scaling up the system for broader deployments could increase these expenses, posing challenges in resource allocation. To address these limitations, future research can expand the OCSD by incorporating additional classes and relationships, thereby improving coverage and versatility. Fine-tuning domain-specific LLMs and integrating advanced 3D semantic segmentation techniques can enhance the precision of attribute extraction and severity assessment. Moreover, exploring strategies to optimize computational costs, such as deploying more efficient models or refining interaction protocols, would support sustainable, large-scale implementation.

References

[1] Tupayachi, J., Xu, H., Omitaomu, O.A., Camur, M.C., Sharmin, A. and Li, X., 2024. Towards next-

generation urban decision support systems through AI-powered construction of scientific ontology using large language models—A case in optimizing intermodal freight transportation. *Smart Cities*, *7*(5), pp.2392-2421.

[2] Saeedizade, M.J. and Blomqvist, E., 2024, May. Navigating Ontology Development with Large Language Models. In *European Semantic Web Conference*, pp. 143-161, Cham: Springer Nature Switzerland.

[3] Momcilovic, T.B., Buesser, B., Zizzo, G., Purcell, M. and Balta, D., 2024. Towards Assurance of LLM Adversarial Robustness using Ontology-Driven Argumentation. *arXiv preprint arXiv:2410.07962*.

[4] Zhang, B., Carriero, V.A., Schreiberhuber, K., Tsaneva, S., González, L.S., Kim, J. and de Berardinis, J., 2024. OntoChat: a Framework for Conversational Ontology Engineering using Language Models. *arXiv preprint arXiv:2403.05921.*

[5] Bahreini, F. and Hammad, A., 2024. Developing an Ontology for Concrete Surface Defects to Enhance Inspection, Diagnosis and Repair Information Modeling. *Infrastructures*, *9*(12), p.220.

[6] He, Y., Chen, J., Dong, H. and Horrocks, I., 2023. Exploring large language models for ontology alignment. *arXiv preprint arXiv:2309.07172*.

[7] Wang, P., Karigiannis, J. and Gao, R.X., 2024. Ontology-integrated tuning of large language model for intelligent maintenance. *CIRP Annals*, 73(1), pp. 361-364.

[8] Meyer, L.P., Stadler, C., Frey, J., Radtke, N., Junghanns, K., Meissner, R., Dziwis, G., Bulert, K. and Martin, M., 2023, June. LLM-assisted knowledge graph engineering: Experiments with ChatGPT. In *Working Conference on Artificial Intelligence Development for a Resilient and Sustainable Tomorrow*, pp. 103-115, Wiesbaden: Springer Fachmedien Wiesbaden.

[9] Palagin, O., Kaverinskiy, V., Litvin, A. and Malakhov, K., 2023. OntoChatGPT information system: Ontology-driven structured prompts for ChatGPT meta-learning. *arXiv preprint arXiv:2307.05082*.

[10] Qiang, Z., Taylor, K., Wang, W. and Jiang, J., 2024. OAEI-LLM: A Benchmark Dataset for Understanding Large Language Model Hallucinations in Ontology Matching. *arXiv preprint arXiv:2409.14038*.

[11] Norouzi, S.S., Mahdavinejad, M.S. and Hitzler, P., 2023. Conversational ontology alignment with chatgpt. *arXiv preprint arXiv:2308.09217*.

[12] Allemang, D. and Sequeda, J., 2024. Increasing the LLM Accuracy for Question Answering: Ontologies to the Rescue!. *arXiv preprint arXiv:2405.11706*.

[13] Song, J. and Yoon, S., 2024. Ontology-assisted GPT-based building performance simulation and assessment: Implementation of multizone airflow simulation. *Energy and Buildings*, *325(1)*, p.114983.

[14] Nakajima, H. and Miura, J., 2024. Combining Ontological Knowledge and Large Language Model for User-Friendly Service Robots. *arXiv preprint arXiv:2410.16804*.

[15] Babaei Giglou, H., D'Souza, J. and Auer, S., 2023, October. LLMs4OL: Large language models for ontology learning. In *International Semantic Web Conference*, pp. 408-427, Cham: Springer Nature Switzerland.

[16] Baldazzi, T., Bellomarini, L., Ceri, S., Colombo, A., Gentili, A. and Sallinger, E., 2024. *Fine-tuning large language models for financial markets via ontological reasoning*, No. 44, Bank of Italy, Directorate General for Markets and Payment System.

[17] Margineda Mangues, X., 2024. Development of a PCG with a LLM as an inference engine and an ontology to aid the creative process, *Bachelor's Thesis*, Universitat Politècnica de Catalunya.

[18] Doumanas, D., Soularidis, A., Kotis, K. and Vouros, G., 2024, June. Integrating LLMs in the Engineering of a SAR Ontology. In *IFIP International Conference on Artificial Intelligence Applications and Innovations*, pp. 360-374, Cham: Springer Nature Switzerland.

[19] Liu, B., Wu, X., Pan, D., Chen, Y., Huang, J., Liao, F. and Lin, S., 2024, August. Enhancing Large Language Models with Graph-Based Node Sampling for Fault Attribution in Power Distribution Networks. In *International Conference on Intelligent Computing*, pp. 444-455, Singapore: Springer Nature Singapore.

[20] Tsaneva, S., Vasic, S. and Sabou, M., 2024. LLM-driven Ontology Evaluation: Verifying Ontology Restrictions with ChatGPT, In *Data Quality Meets Machine Learning and Knowledge Graphs: DQMLKG Workshop at ESWC,* pp. 1-16, Hersonissos, Greece.

[21] Bahreini, F. and Hammad, A., 2024, Dynamic graph CNN based semantic segmentation of concrete defects and as-inspected modeling. *Automation in Construction*, *159*, p.105282.

[22] Woodson, R.D., 2009. *Concrete structures: protection, repair and rehabilitation*. Butterworth-Heinemann.

[23] Ontario Ministry of Transportation Ontario, 2018, Ontario Structure Inspection Manual (OSIM*), Ontario ministry of transportation, Provincial highways management division, Highway standards branch, Bridge office, Ontario, Canada.*

3D Full-body Pose Estimation of Construction Machines Using Deep Neural Network and Stereo Vision

Han Luo[1], Mingzhu Wang[2,*], Peter Kok-Yiu Wong[3], Pak Him Leung[3], Jingyuan Tang[4], Jack C. P. Cheng[3]

[1] China Three Gorges Investment Management Co., Ltd, China
[2] Department of Architecture and Civil Engineering, City University of Hong Kong, Hong Kong, China
[3]Department of Civil and Environmental Engineering, the Hong Kong University of Science and Technology, Hong Kong, China
[4] School of Management Science and Real Estate, Chongqing University, Chongqing, China
*corresponding author
han.luo@connect.ust.hk; m.wang@cityu.edu.hk; kywongaz@connect.ust.hk; phleungae@connect.ust.hk; jtangap@cqu.edu.cn; cejcheng@ust.hk

Abstract –

The movement and posture variability of construction machines is a significant contributor to safety hazards on construction sites. Even when a machine's location is fixed, its moving parts may collide with on-site personnel or objects, leading to injuries or production loss. Accurate estimation of 3D full-body poses of machines can enhance safety by providing more precise spatial information. This paper proposes a framework to estimate 3D full-body poses of construction machines using deep neural networks (DNNs) and stereo vision. The proposed framework employs an entropy-based active learning method to select informative images for fine-tuning the DNN model for 2D pose estimation. 3D poses are estimated through stereo camera calibration, coarse-to-fine stereo matching, and triangulation. Experimental validation using an excavator model achieved an average error percentage (AEP) of 12.11%, demonstrating the framework's feasibility for enhancing safety management.

Keywords –

Computer vision; Construction machine; Deep learning; Deep active learning; Pose estimation, Stereo vision; Construction safety

1 Introduction

Construction sites are inherently hazardous environments, with a significant portion of accidents caused by heavy machinery. In the U.S., the construction sector employed nearly 7.5 million workers in 2018 [1], contributing to 19.77% of all occupational fatalities. Similarly, Hong Kong reported 3,726 work-related injuries in the construction sector in 2018 [2]. Heavy machinery movement is a major source of these hazards, and the interaction between workers and machines exacerbates the risk. Therefore, monitoring machine movements is critical for improving safety. Although surveillance cameras are increasingly used, safety monitoring is often still performed manually, which can be inconsistent and subject to human limitations. Automated monitoring offers a promising solution, but most current systems focus on 2D pose estimation, which is insufficient for evaluating 3D safety conditions. For example, 2D pose estimation cannot accurately capture the spatial relationships between machine components and nearby workers or objects, leading to potential safety hazards. In real-world construction sites, machines like excavators have complex 3D movements, and their moving parts (e.g., booms, arms, and buckets) can pose significant risks to workers even when the machine itself is stationary. The current reliance on 2D monitoring fails to provide the comprehensive spatial information needed to assess these risks accurately. This limitation increases the likelihood of accidents, injuries, and production losses in the construction industry. Therefore, there is a critical need for 3D full-body pose estimation to enhance safety monitoring and mitigate these risks.

Previous studies have utilized devices such as Inertial Measurement Units (IMUs) [3] and Global Positioning Systems (GPS) [4], along with surveillance cameras [5-7], to estimate machine poses, yielding satisfactory results. However, systems relying on pre-installed devices often require multiple installations, making them impractical. In contrast, surveillance cameras can be installed once and provide continuous monitoring. While computer vision techniques are becoming popular for machine pose estimation, most existing studies focus on 2D poses, which are insufficient for evaluating 3D safety conditions. 3D pose estimation can offer critical spatial

information about the machine's working zone, which can then be used to assess potential accident risks more accurately. However, research in 3D pose estimation for construction machines remains limited.

This study proposes a framework for 3D full-body machine pose estimation using deep neural networks (DNNs) and stereo vision to provide comprehensive information for safety assessment. The framework uses deep active learning to fine-tune a pre-trained model, minimizing the need for large annotated datasets. The system estimates 2D full-body poses through stereo vision, and 3D poses are derived via stereo matching and triangulation. The proposed method is validated using an excavator model, demonstrating its potential for enhancing safety monitoring on construction sites.

2 Related Works

The safety of construction sites has garnered increasing attention due to its significant potential for spatial-temporal safety analysis and proactive hazard warning. Research in this area focuses on various techniques, including machine pose estimation, transfer learning, and stereo vision, to enhance the monitoring of construction machines and workers.

2.1 Pose Estimation of Construction Machines

Pose estimation is critical for determining the precise working areas of construction machinery, which is essential for safety monitoring. Traditionally, sensors such as Inertial Measurement Units (IMUs) [3] and Global Positioning Systems (GPS) [4] have been employed for estimating machine poses. Tang et al. [8] used IMUs to estimate 3D poses by measuring angles between machine components. While these approaches provide useful data, they are limited by the need for additional hardware installation, increasing costs. GPS and Ultra-Wide Band (UWB) systems [9] offer some advantages but also require extra infrastructure.

In contrast, computer vision has emerged as a cost-effective alternative, utilizing cameras already installed on construction sites. Early computer vision approaches focused on 2D pose estimation. Azar and McCabe [10] proposed a method based on surveillance video analysis using support vector machines (SVMs) and histogram of oriented gradients (HOG). This was later improved by integrating machine skeleton extraction [5] and real-time location systems (RTLS) [11], achieving more accurate 2D pose estimations. However, these methods relied on handcrafted features that required careful design, which limited their generalization capability. More recently, deep neural networks (DNNs) have been used for 2D pose estimation due to their ability to handle complex patterns in data [7]. But it relied solely on 2D image processing, lacking the ability to provide depth information for safety monitoring.

3D machine pose estimation is essential for spatial-temporal safety analysis, as it provides information about the machine's working zone. While stereo vision and DNNs have been combined to extend pose estimation to 3D, existing studies have primarily focused on partial poses, such as the boom, bucket, and stick of excavators [11,12]. These methods fail to provide a comprehensive understanding of the full-body poses of machines and their dynamic working regions, which are crucial for safety evaluations.

2.2 Transfer Learning for On-Site Monitoring

A challenge in machine pose estimation is the need for large, annotated datasets to train reliable DNN models. Manual annotation is time-consuming and costly, particularly for dynamic environments like construction sites. To address this, transfer learning has been explored as a method to apply knowledge gained from one domain to another, reducing the amount of new data required. Transfer learning has been used in construction safety applications, such as Kolar et al.[13], who applied VGG-16 for safety guardrail detection, and Kim et al. [14], who utilized pre-trained models for machine pose estimation.

However, transfer learning approaches often overlook the quality of the annotated data. A smaller high-quality dataset may yield better results than a large poorly annotated one. To improve data efficiency, deep active learning (DAL) has been introduced. DAL allows the model to iteratively select the most informative images for manual annotation, thereby minimizing the amount of data required. Kim et al. [15] demonstrated the effectiveness of DAL in on-site machine detection, showing that it improves DNN performance while reducing manual labeling efforts. While DAL has shown promise, its application in construction machine pose estimation remains underexplored.

2.3 Stereo Vision for On-Site Monitoring

Traditional single-camera systems often struggle with capturing 3D information due to projection and occlusion effects. Stereo vision, which uses multiple cameras, offers a promising solution by reconstructing 3D information from 2D images. Stereo vision has been applied in construction monitoring to estimate the 3D locations of workers and machines [16]. However, previous studies have only provided rough estimations of 3D locations, often ignoring variations in poses during machine operation. Han and Lee [17] explored 3D worker pose estimation using stereo vision, and Yuan et al. [12] applied stereo vision for 3D machine pose estimation by detecting joints and triangulating the corresponding keypoints.

Despite its potential, stereo vision often relies on

conventional computer vision techniques, such as HOG-based detection, which have limitations in processing complex 2D images and extracting useful features. Recent advancements suggest that DNNs can overcome these limitations, as demonstrated by researchers in the robotics field [18], who used DNNs for 3D movement tracking of robotic arms. The integration of stereo vision and DNNs for construction machine pose estimation is a promising avenue for improving 3D safety monitoring, but further research is needed to fully explore this approach.

3 The Proposed Framework for 3D Full-body Pose Estimation of Construction Machines Using Deep Neural Network and Stereo Vision

This study proposes a framework for automatically estimating 3D full-body poses of construction machines using deep neural networks (DNNs) and stereo vision, based on images and videos. As illustrated in Figure 1, the framework consists of two sequential modules: (1) 2D machine pose estimation and (2) 3D machine pose estimation.

The process begins by capturing stereo images (e.g., *N* images) using a multi-camera system. Each image is processed independently by a DNN-based 2D pose estimation model to extract keypoints of the machine's pose, along with the confidence probability for each keypoint. Given that collecting a large annotated dataset is time-consuming, deep active learning is utilized to select images with the most informative pose keypoints for fine-tuning the pre-trained 2D pose estimation model.

After fine-tuning, the model estimates 2D machine poses, which are then passed to the 3D pose estimation module. This module first performs stereo camera calibration to obtain intrinsic and extrinsic parameters for triangulation. Using the results from 2D pose estimation, coarse-to-fine stereo matching is applied to refine pose keypoints. Finally, triangulation calculates the 3D coordinates of the machine's poses based on keypoint correspondences and stereo camera calibration results.

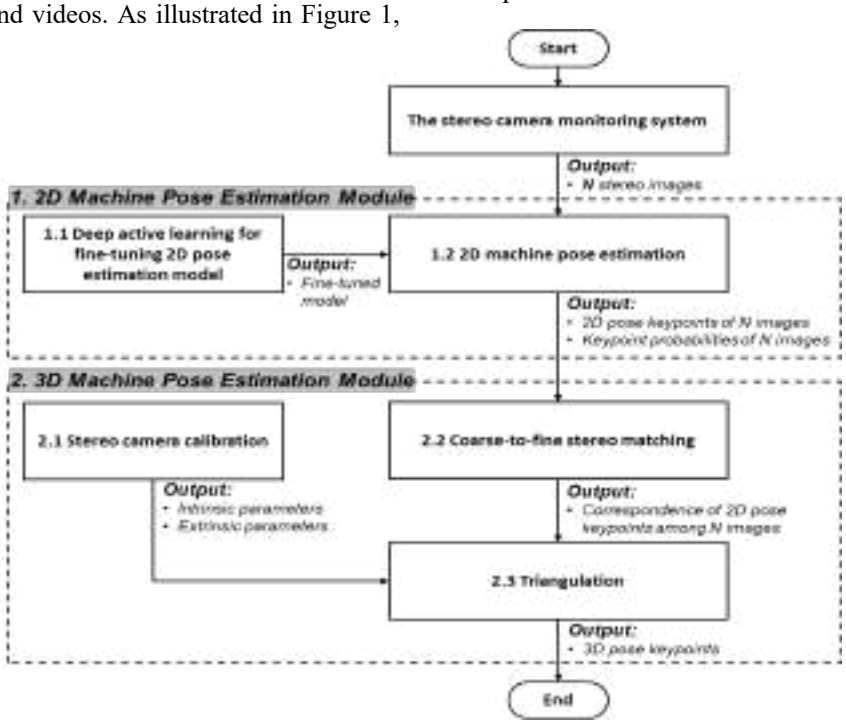


Figure 1. The overall framework for 3D full-body pose estimation of construction machines using DNN and stereo vision.

3.1 2D Machine Pose Estimation Module Based on Deep Active Learning

The goal of this step is to fine-tune a DNN model for 2D machine pose estimation using a small set of manually annotated images collected from new construction sites. Deep active learning (DAL) is employed to reduce the number of training images required by selecting the most informative images for manual annotation. This iterative process involves transferring knowledge from a pre-trained model, which is initially trained on a general dataset of construction sites, to new construction scenes. Figure 2 presents the typical flow of deep active learning for this process.

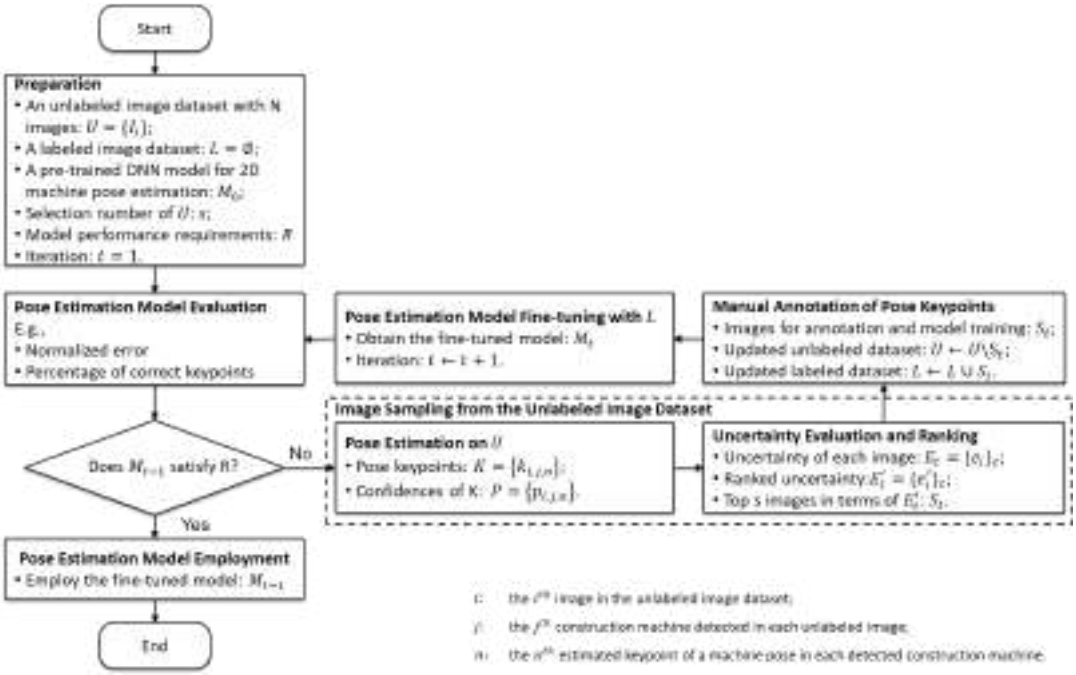


Figure 2. Workflow of deep active learning for fine-tuning the 2D pose estimation model.

Preparation. The first step involves preparing an unlabeled image dataset (U), containing images collected from a new construction site. The dataset should include a variety of machine operations to ensure the model can generalize to new images. The pre-trained DNN model (denoted as M_0) is used as the starting point to avoid training from scratch. Pre-defined performance evaluation metrics, such as Normalized Error (NE) and Percentage of Correct Keypoints (PCK), are also defined for model performance assessment.

Pose Estimation Model Evaluation. The performance of the current model is evaluated using the NE and PCK metrics. If the model does not meet the required performance thresholds, the process moves to the next step.

Pose Estimation on U. The 2D pose estimation model is applied to all images in the unlabeled dataset, and the model estimates keypoint coordinates along with the confidence score for each keypoint.

Uncertainty Evaluation and Ranking. The uncertainty of each image is evaluated using entropy-based uncertainty evaluation, where higher entropy indicates more uncertainty. Images with higher uncertainty are considered more informative and are selected for manual annotation.

Manual Annotation of Pose Keypoints. Expert annotators manually label the selected images, providing keypoint annotations. These annotated images are then incorporated into the labeled dataset.

Pose Estimation Model Fine-Tuning with L. The labeled dataset is used to fine-tune the DNN model. The model weights are updated using forward inference and backpropagation.

Pose Estimation Model Employment. Once the fine-tuned model satisfies the performance requirements, it is employed for 2D pose estimation on images and video frames from the target construction site, completing the deep active learning process.

3.2 3D Pose Estimation of Construction Machines Using Stereo Vision

The 2D poses estimated in Section 3.1 are used in the 3D pose estimation module. Stereo camera calibration is firstly performed to obtain necessary camera parameters, and then coarse-to-fine stereo matching is applied to the 2D pose keypoints. This module generates 3D poses by triangulating the keypoints from the stereo images.

3.2.1 Stereo Camera Calibration

Stereo camera calibration is an offline process that provides intrinsic and extrinsic parameters for the triangulation process. The intrinsic parameters describe internal camera characteristics, such as focal length and image center, while extrinsic parameters describe the relative position and orientation of the cameras in the stereo system.

3.2.2 Coarse-to-Fine Stereo Matching

The goal of stereo matching is to find corresponding points from images with different viewpoints. The proposed coarse-to-fine stereo matching strategy includes:

(1) *Keypoint-Based Coarse Stereo Matching:* This step matches pose keypoints based on semantic information (e.g., keypoint type) from 2D pose estimation, as shown in Figure 4.

(2) *Patch-Based Stereo Matching Refinement:* This step refines the stereo matching by using a sliding window approach to compare template patches with search regions to find the best match for each keypoint and therefore improve keypoint correspondence accuracy (as shown in Figure 5).

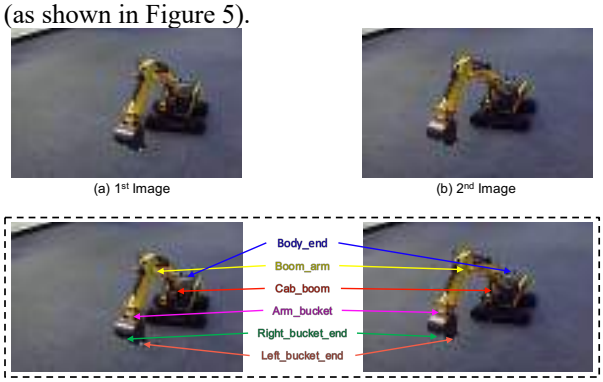


Figure 4. Keypoint-based coarse stereo matching using two stereo images as an example.

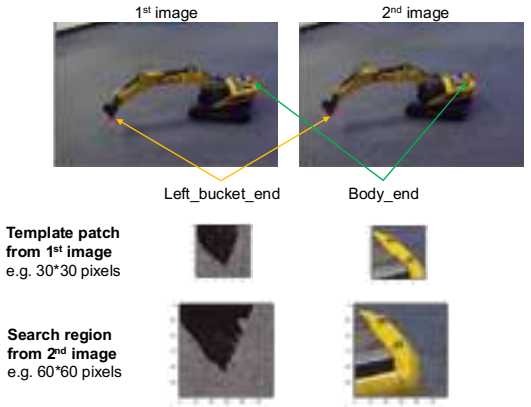


Figure 5. Illustrative examples of template patches and search regions.

3.2.3 Triangulation

The triangulation process uses the refined stereo matching results and camera parameters to compute the 3D position of each keypoint. Linear triangulation is used to solve the equation and determine the 3D coordinates.

4 Validation of the Proposed Framework

To validate the feasibility of the proposed framework, this study uses excavators as illustrative examples due to their numerous movable components and complex poses during construction tasks. The pose keypoints of excavators are defined as shown by the dark dots in Figure 6.

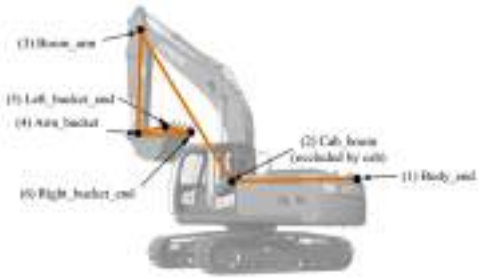


Figure 6. Pose keypoints of construction machines defined in [7].

4.1 Experiment Data Preparation

Due to the lack of a dataset for 3D machine pose estimation using stereo vision, an excavator model was used to create a controlled dataset. The excavator model, a 1:14 scale full-metal replica, performs various tasks like digging, dumping, and swinging during earthmoving operations. The actual lengths of the movable components were measured (see Table 1). A ZED stereo camera was used to capture stereo images, as illustrated in Figure 7. For real-world deployment, cameras can be mounted on fixed infrastructure to ensure a stable and unobstructed view of the construction site. The height and angle of the cameras should be optimized to minimize occlusions and maximize coverage of the working area.

Table 1. Component lengths of the adopted scaled-down excavator model

Component name	Cab	Boom	Arm	Bucket
Length (mm)	180	300	155	80



Figure 7. ZED camera fixed in a holder to capture stereo images of working construction machines.

During data collection, the excavator model mimicked real-world activities such as digging, swinging, and dumping. The dataset consists of 1069 pairs of stereo images, which was split into test and non-test classes using stratified random sampling, with 20% (214 images) allocated for testing and the remaining 80% (855 images) for fine-tuning the DNN models.

4.2 Implementation Details

Two main experiments were conducted to evaluate the framework's performance:

(1) 2D Machine Pose Estimation: This experiment fine-tunes a pre-trained DNN-based model for 2D pose estimation using deep active learning. The model was trained using stereo images captured from Camera 1. The DNN model was fine-tuned using two different sampling strategies: deep active learning (as outlined in Section 3.1) and random sampling.

(2) 3D Full-Body Machine Pose Estimation: After the 2D model was fine-tuned, 3D poses were estimated using both stereo images from Camera 1 and Camera 2. Stereo camera calibration was performed manually, with an overall calibration mean error of 0.051 pixels. The 3D poses were estimated using the proposed coarse-to-fine stereo matching and triangulation method.

The experiments were implemented using Python on Ubuntu 16.04, with an Intel Core i7-6700 CPU @ 3.40 GHz×8 and a NVIDIA GeForce GTX 1080 GPU. The DNN model was built using PyTorch.

4.3 Evaluation Metrics

2D Machine Pose Estimation: The performance of the 2D machine pose estimation model was evaluated using Normalized Error (NE), which calculates the average normalized distance between estimated and actual positions of pose keypoints. The NE formula is defined by Equations (1) and (2).

$$NE = \frac{1}{N}\sum_i \frac{\sum_n \left(\frac{d_{i,n}}{s_i} \cdot \delta(v_{i,n} \geq 0)\right)}{\sum_n \left(\delta(v_{i,n} \geq 0)\right)} \quad (1)$$

$$s_i = \max(h, w) \quad (2)$$

where, s_i is the distance normalization parameter in the i^{th} image; h is the height of the bounding box of a detected construction machine in the i^{th} image; w is the width of the bounding box of the detected machine in the i^{th} image; N is the total number of test images as defined in Figure 1; $d_{i,n}$ is the Euclidean distance between the estimated and the actual position of the n^{th} keypoint in the i^{th} image; $v_{i,n}$ means the visibility of the n^{th} keypoint in the i^{th} image.

3D Machine Pose Estimation: The accuracy of 3D pose estimation was evaluated using Average Error Percentage (AEP), comparing the estimated and actual component lengths using Equation (3). A smaller AEP indicates better accuracy in 3D pose estimation.

$$AEP = \frac{1}{N}\sum_i \sum_{mc} \left(\frac{\left|L_{i,mc}^{Estimated} - L_{i,mc}^{Actual}\right|}{L_{i,mc}^{Actual}}\right) \quad (3)$$

where, $L_{i,m}$ is the estimated or actual component length of the m^{th} component of a detected construction machine in the i^{th} image; mc is the indexing of machine component, which can be Cab, Boom, Arm and Bucket.

4.4 Experiment Results and Discussion

4.4.1 2D Pose Estimation Using Deep Active Learning

Table 2 shows the NE results of the 2D pose estimation DNN model, which was fine-tuned using deep active learning. With 25 iterations and 750 images, the fine-tuned model achieved a 77.82% reduction in average NE compared to the pre-trained model. This method significantly reduced the number of images needed for training, demonstrating its efficiency for on-site safety monitoring.

The experiment also analyzed the impact of different sampling strategies for deep active learning. As shown in Figure 8, the NE value decreases significantly as more images are used for fine-tuning. According to Figure 9, the entropy-based uncertainty evaluation method outperforms random sampling in most cases, showing a more effective strategy for selecting images for training.

Table 2. Normalized error (NE) results of the adopted DNN model for 2D pose estimation

Keypoint	Body_end	Cab_boom	Boom_arm	Arm_bucket	Left_bucket_end	Right_bucket_end	Average
Pre-trained (10^{-3})	88.8644	51.1234	48.5361	59.4094	46.8722	43.9860	56.4653
Final fine-tuned (10^{-3})	17.0047	20.1196	6.2324	8.8859	10.6549	12.2473	12.5241
NE reduction	80.86%	60.65%	87.16%	85.04%	77.27%	72.16%	77.82%

Table 3. Results of the proposed 3D pose estimation framework

No	Keypoint-based coarse stereo matching	Patch-based stereo matching refinement		Average error percentage (AEP) (%)					Estimation speed (ms/sample)
		Basic setting (template patch from a random camera)	Confidence-based template patch selection	Cab	Boom	Arm	Bucket	Avg.	
1	√	√	√	**12.86**	10.64	**8.68**	16.26	**12.11**	133.21
2	√	√	×	14.13	10.22	9.16	**15.80**	12.33	132.60
3	√	×	×	18.25	**9.12**	9.71	19.91	14.24	131.61

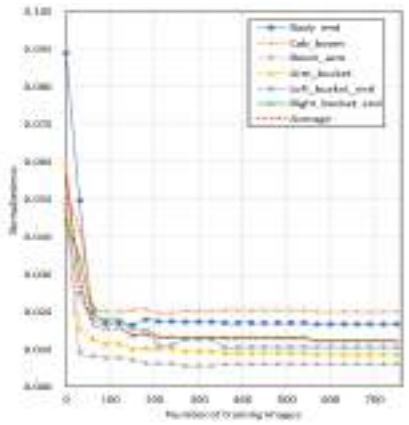


Figure 8. NE curves during deep active learning for 2D pose estimation with different pose keypoints.

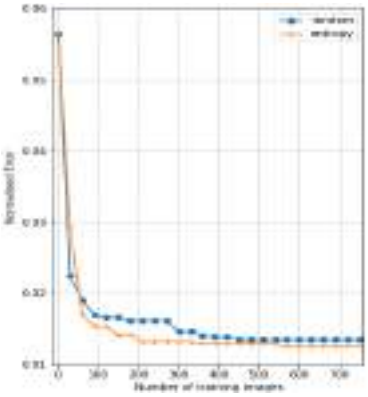


Figure 9. NE curves during deep active learning comparing entropy-based sampling and random sampling.

Table 4. Comparison between the results of the proposed framework and those of previous studies.

No	Method	Experimental condition	Average error percentage (AEP) (%)					Estimation speed (ms/sample)
			Cab	Boom	Arm	Bucket	Avg.	
1	Stereo vison for 3D pose estimation (Ours)	Controlled environment	12.86	10.64	8.68	16.26	**12.11**	133.21
2	Deep learning for 3D pose estimation [6]	Real-world construction site	-	39.10	30.70	58.80	42.80	832.00
3	Stereo vision with location data fusion [11]	Real-world construction site	-	9.50	30.00	-	19.75	≈2000
4	Stereo vision without location data fusion [11]	Real-world construction site	-	36.38	74.21	-	55.29	≈2000

4.4.2 3D Pose Estimation Based on Stereo Vision

Table 3 summarizes the results of 3D pose estimation using the proposed framework. The average AEP for all machine components is 12.11%. The Boom length achieved the smallest error (8.68%), while the Bucket length had the highest error (16.26%). The estimation speed was 133.21 ms per stereo image sample.

The results indicate that the combination of keypoint-based coarse stereo matching and patch-based stereo matching refinement yields the best AEP. Notably, the Boom length was more accurately estimated in the trial with no refinement, likely due to occlusion issues with the Cab_boom keypoint. Regarding the significant error of bucket length, it is computed using Arm_bucket keypoint and the middle point between Left_bucket_end and Right_bucket_end. The 2D pose estimation results of the three keypoints may not be accurate simultaneously, and such errors will be transmitted to 3D machine pose estimation. Hence, the error may exceed expectations, especially when the 2D estimated result of one of the three keypoints deviates from its actual position.

Table 4 presents a comparative analysis of the proposed framework against prior studies, including Ref. [11], which integrates 2D estimation with stereo vision and location data fusion for 3D pose estimation. This comparison serves as a valuable benchmark for evaluating 3D pose estimation performance in the context of construction machinery. The proposed framework demonstrates a substantial improvement in AEP over earlier methods. Specifically, it achieves an AEP of 12.11%, significantly surpassing previous studies that reported AEPs of 42.8%, 19.75%, and 55.29%. Moreover, the estimation speed of the proposed framework is much faster compared to previous studies.

5 Conclusions

Construction site safety is increasingly important due to the hazardous and labor-intensive nature of the industry. Monitoring 3D full-body poses of heavy construction machines is essential for mitigating potential hazards, as it provides comprehensive spatial information about the area occupied by on-site machines. This study proposes a framework for 3D full-body pose estimation of construction machines using deep neural networks (DNNs) and stereo vision. By fine-tuning a pre-trained model with deep active learning and using a coarse-to-fine stereo matching approach, the framework efficiently estimates 3D machine poses based on 2D images. Key contributions include:

(1) Achieved an average AEP of 12.11% for estimating full 3D poses, offering a more comprehensive safety evaluation than partial poses in prior studies;

(2) Applied DNNs to extract features for triangulation, improving efficiency over traditional methods;

(3) Developed a deep active learning strategy with entropy-based uncertainty evaluation, reducing the need for large training datasets;

(4) Proposed a coarse-to-fine stereo matching strategy

to minimize 2D pose estimation errors in 3D estimation.

The framework shows strong potential for jobsite application. However, its real-world deployment faces challenges such as visual complexities, occlusions, lighting variations, and environmental factors. Future work will focus on scaling up the framework for real-world deployment through pilot testing, developing data collection protocols for heavy machinery, and enhancing robustness with techniques like multi-view systems. The framework will undergo rigorous testing in complex jobsite environments and be integrated with real-time safety alerts to detect unsafe actions of construction machines and notify relevant workers and managers instantly.

Acknowledgments

The authors are grateful for the financial support provided by the National Nature Science Foundation of China (Grant number: 72301043).

References

[1] U.S. Bureau of Labor Statistics, "*Graphics for economic news releases - employment levels by industry,*" 2020, Available at: https://www.bls.gov/charts/employment-situation/employment-levels-by-industry.htm (accessed: 13 January 2021).

[2] Occupational Safety and Health Branch, "*Occupational Safety and Health Statistics 2018,*" 2019, Available at: http://www.labour.gov.hk (accessed: 13 January 2021).

[3] J. Tang, H. Luo, P. K.-Y. Wong, and J. C. Cheng, "Study of IMU installation position for posture estimation of excavators," in *Proceedings of the International Conference on Computing in Civil and Building Engineering*, 2020, pp. 980-991: Springer.doi:10.1007/978-3-030-51295-8_68

[4] Leica, "Leica iCON iXE3 - 3D system," 2019.

[5] M. M. Soltani, Z. Zhu, and A. Hammad, "Skeleton estimation of excavator by detecting its parts," *Automation in Construction,* vol. 82, pp. 1-15, 2017.

[6] C.-J. Liang, K. M. Lundeen, W. McGee, C. C. Menassa, S. Lee, and V. R. Kamat, "A vision-based marker-less pose estimation system for articulated construction robots," *Automation in Construction,* vol. 104, pp. 80-94, 2019.

[7] H. Luo, M. Wang, P. K.-Y. Wong, and J. C. Cheng, "Full body pose estimation of construction equipment using computer vision and deep learning techniques," *Automation in Construction,* vol. 110, p. 103016, 2020.

[8] J. Tang, H. Luo, W. Chen, P. K.-Y. Wong, and J. C. Cheng, "IMU-based full-body pose estimation for construction machines using kinematics modeling," *Automation in Construction,* vol. 138, p. 104217, 2022.

[9] F. Vahdatikhaki, A. Hammad, and H. Siddiqui, "Optimization-based excavator pose estimation using real-time location systems," *Automation in Construction,* vol. 56, no. aug., pp. 76-92, 2015.

[10] E. R. Azar and B. Mccabe, "Part based model and spatial–temporal reasoning to recognize hydraulic excavators in construction images and videos," *Automation in Construction,* vol. 24, no. Jul., pp. 194-202, 2012.

[11] M. M. Soltani, Z. Zhu, and A. Hammad, "Framework for location data fusion and pose estimation of excavators using stereo vision," *Journal of Computing in Civil Engineering,* vol. 32, no. 6, p. 04018045, 2018.

[12] C. Yuan, S. Li, and H. Cai, "Vision-based excavator detection and tracking using hybrid kinematic shapes and key nodes," *Journal of Computing in Civil Engineering,* vol. 31, no. 1, p. 04016038, 2017.

[13] Z. Kolar, H. Chen, and X. Luo, "Transfer learning and deep convolutional neural networks for safety guardrail detection in 2D images," *Automation in Construction,* vol. 89, no. MAY, pp. 58-70, 2018.

[14] H. Kim, H. Kim, Y. Hong, and H. Byun, "Detecting construction equipment using a region-based fully convolutional network and transfer learning," *Journal of Computing in Civil Engineering,* vol. 32, no. 2, 2018.

[15] J. Kim, J. Hwang, S. Chi, and J. O. Seo, "Towards database-free vision-based monitoring on construction sites: A deep active learning approach," *Automation in Construction,* vol. 120, 2020.

[16] Y. J. Lee and M. W. Park, "3D tracking of multiple onsite workers based on stereo vision," *Automation in Construction,* vol. 98, no. FEB., pp. 146-159, 2019.

[17] S. U. Han and S. H. Lee, "A vision-based motion capture and recognition framework for behavior-based safety management," *Automation in Construction,* vol. 35, pp. 131-141, 2013.

[18] M. Abdelaal, R. M. A. Farag, M. S. Saad, A. Bahgat, H. M. Emara, and A. El-Dessouki, "Uncalibrated stereo vision with deep learning for 6-DOF pose estimation for a robot arm system," *Robotics and Autonomous Systems,* vol. 145, 2021.

Deployment for Additive Construction through Digital Platforms: Gananoque Project as a Case Study

Pablo Banda[1], Ioulios Georgiou[1], Patricio Emanuel Carrasco Pérez[2], Romanella Soria[1], Zayed Elbadri[1], Alejandro Gonzalez[1], Daniel Cantor[1] and Nhung Nguyen[1]

[1] Horizon Legacy Construction Automation Corp, Canada
[2] Bío-Bío University, Chile
pbanda@horizonlegacy.com, igeorgiou@horizonlegacy.com, patriciocarrascoperez@gmail.com, rsoria@horizonlegacy.com, zelbadri@horizonlegacy.com, agonzalez@horizonlegacy.com, dcantor@horizonlegacy.com, nnuygen@horizonlegacy.com

Abstract –

This work presents the 3D Printed Construction Digital Workflow (3DPCDW) as a novel platform for integrating AEC with 3DPC, Digitalizing the 3DPC flow from basic elements to execution with robotic manipulators, while augmenting design through inherent tectonic expression of this technology. This platform was tested producing programs for the Gananoque Project, a 26 houses development in Canada by Horizon Legacy© (HL), delivering a successfully printed envelope. Thereafter, Gananoque Project positioned 3DPCDW as a foundational framework for future 3D developments.

Keywords –

Digital Design; 3D Printed Construction; Robotic Manipuator; Architectural Projects.

1 Introduction

The construction industry has seen significant technological advancements, with the integration of digitalization and automation within the framework of Industry 4.0 [1-2] conceiving a massive implementation of robotics and 3D printing[3].

In this context, Additive Manufacturing (AM) materializes a volume stacking layers by deposition, is primarily valued for its ability to enable complex designs and structural optimization [4-5]. The application of AM with cementitious materials was originally first explored by Khoshnevis [6], and it is characterized by implementation of integrating computational, 3D printing specific and material knowledge[7-9].

1.1 The Gananoque Project

The Gananoque project is a housing development consisting of two townhouse blocks of 26 houses in total. The perimeter walls of the first floor were 3D-printed, and the second storey was conventionally built with timber. For HL this project represents the first onsite proof-of-concept deployment of 3DPC.

The Gananoque project invokes a series of challenges for Digital Design Workflow (DDW), particularly on the delivery of large walls and a narrow footprint. 3DPCDW assisted successfully those challenges by retroactively interacted with 3DPC team during this early exposition to the construction site.

1.2 Digital Design Workflows

DDW's for 3DPC are parametric CAD software platforms that develop printable building data from AEC sources, principally architecture. The minimum feature is the *slicing* process, or the process of describing a 3D object into evenly displaced 2D horizontal layers along the Z-axis. At its maximum, DDW's features can become a holistic development platform, in terms of building design orientation to 3DPC, producing a *Digital Conception* of the model through parametric or generative means. These can include evaluations and simulations of all the pertinent building performances, as well as *AEC-3DPC Integration* of building geometry, material information and machinery details. This bonding of Digital Design, AEC and 3DPC Knowledge into different *Parametric Modeling* (PM) environments can potentially transform DDWs in a state-of-the-art holistic CAD platform for exploring 3DPC capability to set a new standard in buildings[11]; however, that will depend also in DDW's *Generalization* and *Flexibility* aspects in terms of scope of use and adaptation to building context vicissitudes.

Current DW use different manifestations of the BIM paradigm[12], that will tend to conventional but highly AEC-integrated development software such as *Revit*® (*BIM Systems*), and other *Free-Form Parametric CAD* software such as *Rhino*® with *Grasshopper*®.

1.2.1 BIM Systems

BIM software are AEC digital tools that optimize management and collaboration in construction projects, while enhancing efficiency, reducing errors, and improving communication among key stakeholders[12]. 3DPC implementations in BIM are characterized by a powerful AEC-Integration into the design development and process, as well as a *Generalization* of the DW task because of their more robust, consolidated and construction-oriented *Parametric Objects* (PO).

Some remarkable BIM-DW include integration of the *triad*: building geometry, material information and machinery details, to produce efficient and accurate printing codes; while accessing into BIM's *Industry Foundation Classes* [8-9]. This, on the contrary of *brute-force* slicing approach, is an access to the core of BIM PO showing more integration and digitalization.

1.2.2 Free-Form Parametric CAD Approaches

Other DW systems, based on PM platforms such as *Grasshopper©*, are more suitable for producing an impact on Design Conception of the 3DPC project. This lays in their power of *Digitalization* based on the control of complex geometry and performance, through a myriad of plugins for addressing performance, interoperability with robots, etc. Nevertheless, few of the registered DWs for 3D are holistic but project-specific in terms of Design Conception. One of a few examples is Cascone et. al. that developed an entire residential multi-story building streamlining *User Mapping*, *Structural Optimization* and *Daylight Analysis* while outputting printable programs[13-14]. Other authors focus on building specific elements for 3DPC, including in their DW's Finite Element Analysis (FEA) Topology Optimization (TO) for iterate in a generative buckle that ends on 3DPC program delivery [15-16]. Breseguello et. al. using FEA for directly generate toolpath, expressing 3DPC performance-oriented tectonics [17].

2 3DPC Digital Workflow (3DPCDW)

3DPCDW is a workflow and a platform that develops building envelope walls throughout a sequence of tasks for transforming architectural geometry to 3DPC program for execution (see Figure 3), while integrating 3DPC subsystems such as pump, robot manipulator and *Wall Construction*. As a different approach from [7-9] *triad*, In the context of 3DPCDW, *Wall Construction* correspond to an AEC 3DPC Knowledge cluster that integrates constructive, structural and material variables, constraints and recipes.

2.1 Challenges for DW in the context of the Gananoque Project.

The references mentioned in the introduction set some starting positions for creating a DW. We have further acknowledged them as 5 major challenges to face and implement with 3DPCDW, to tackle the challenges of the Gananoque Project.

- **Integration:** Integrating 3DP expertise with digital design and the use of robotics in AEC. This approach minimizes manual intervention and improvisation while delivering supplementary materials efficiently.
- **Digitalization**: Effective digital representation of the 3DPC construction. Controlling the process for the digital development of walls, detailing key design features, execution phases and program delivery.
- **Generalization**: Produce a persistent and robust Workflow and Platform for future projects.
- **Flexibility**: Support for fast onsite adjustments and algorithmic adaptations when facing real contextual construction vicissitudes.
- **Digital Conception:** 3DPC powered shaping of the building project, including performance building aspects beyond 3DPC.

2.2 3DPC Systems and Subsystems comprehended

The 3D printing system implemented for the Gananoque project consists mainly of a 7thaxis manipulator called VAL 2.0 and a concrete pumping system (see Figure 1); as well as a series of diverse elements for cleaning, curing and on- the-fly editing.

Material variables such as Layer Height and Setting Time were comprehended inside Wall Construction.

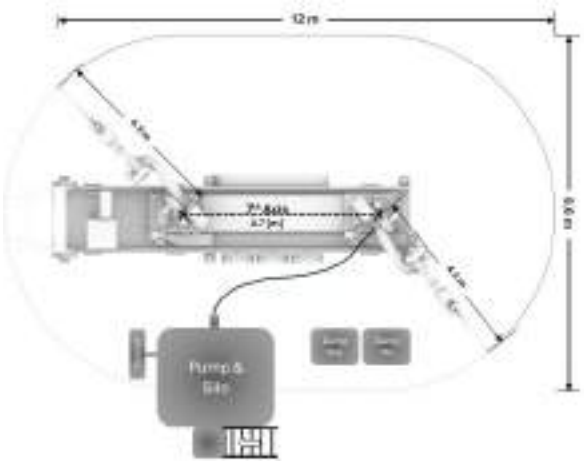


Figure 1. 3DPC System implemented in Gananoque.

2.2.1 VAL 2.0: Large AM Robotic System.

VAL 2.0 is a mobile 7-axis robotic system engineered at HL for CA, based on the ABB IRC 6700 manipulator. Gananoque's EOAT frontal reach is 4.3 [m] maximum and 3.59 [m] for wall heigh, which is amplified to a 12 [m] reach, when combined with the 4.7 [m] linear rail (see Figures 1 and 2). This enables it to execute large rows of walls by concrete and mortar deposition.

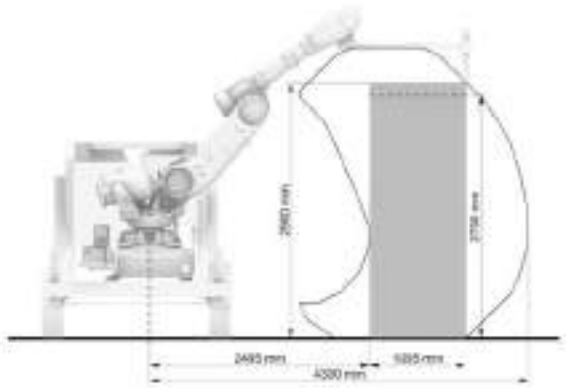


Figure 2. VAL 2.0 vertical reach.

2.2.2 Pumping System

Gananoque Project implemented a double chamber mixer specifically designed for 3DPC featuring a maximum flow of 15 [lt/min], with a 3[m^3] silo for hosting premixes., pushing a min speed of 200 [mm/sec] (see Figure 1).

2.2.3 Wall Construction

In the case of the Gananoque Project, walls consist of a complex ensemble of 3D-printed concrete and rebar, serving both as a self-supporting structure and slab interface. Two wall types were defined: W (normal wall) and *C* (corner wall).

From the perspective of 3DPCDW, wall construction is a 3DPC subsystem with integrated AEC Knowledge from the different parts involved in the Gananoque Project. It is mainly represented by 4 features:

Table 1. Features representing Wall Construction.

Wall Feature	Characteristics	Subtask
Wall Basic Morphology	Exterior, interior, lateral	Architectural Translation
Wall Segregation	Vertical segregation	Toolpath Programming
Wall Complements	Rebars and other structural components	Architectural Translation
Wall Pattern Allocation	Perimeter Projection	Wall Tectonics

3 Digital Workflow

3.1 Architectural Drawing Translation:

The architectural project is initially integrated by processing 2D wall outlines in accordance to *Wall Basic Morphology* to create *Primitive Wall Objects* (PWO). It also generates plans deriving knowledge from *Wall Complements.*

PWOs further simplify wall programming and serve as an interface for future 3DPCDW versions to receive and process other architectural projects.

3.2 3DPC Programming Tectonics

3.2.1 Wall Tectonics

This task augments architectural walls by exploring tectonics unique to 3DPC building components, enabling the Gananoque Project to showcase 3DPC's material expressiveness and structural potentials. As a limited augmentation, it only modifies the projection of the main wythe or the faces of the walls.

Wall Tectonics take the *Perimeter Selection* from the PWOs to add *Wall Patterns* with *Maximum Projection*. This proceeds directly onto toolpath layers as *polylines*, offering more accurate control over the target number economy and enhancing the control of 3DPC tectonic features.

In detail, *Wall Patterns* consist of the allocation of a certain 3D *Relief Pattern* and an associated *Dripping Pattern*. A relief pattern is defined as a volume projected from a supporting background, while a dripping pattern is a tectonic feature that uses a detour at the wythe's intermediate layers with extra peak-point projection, utilizing gravity for the controlled falling of material. The wall pattern modifies this façade gradually, with a maximum nominal value of 55mm to 75mm for wall relief projection, and a 15mm extra projection for dripping patterns. As options, these patterns can create different conjugations or types: *NN* (no pattern), *TN* (wall with relief), *NT* (wall with dripping), *TT* (wall with relief and dripping), that can be suitable for other projects.

3.2.2 Toolpath Designing

This step implements Wall Construction's *Wall Segregation* recipes, defining the suitable toolpath strategies and creating various toolpath types. Toolpath features such as types are essential in programming 3DPC tasks, as they align with the 3DPC system's actual capabilities and limitations, to enhance productivity. They allow for the printing of large amounts of walls and tool changes during the print process.

From this process 3DPCDW delivers four toolpath types: *Sp* (simple print), *Gp* (group print), *DtSp* (double tool simple print) and *DtGp* (double tool group print), according to current *Selected Walls*. While these

strategies offer positive effects, they also introduce complexity and require more control during the print process.

3.3 Toolpath Intervention and Diagnostics

This subtask will evaluate printing programs against VAL 2.0 position for manipulator reaching and printing variables compliance, leading to *Program Delivery*, or requesting changes in previous subtasks such as *Toolpath Designing* or rejecting a current wall group.

This is relevant because it ensures and details reachability, while achieving stable speed that are critical for maintaining precision and quality during printing.

3.3.1 VAL 2.0 Allocation and Movement Control

This subtask will evaluate toolpaths into context, for manipulator reaching, correct movement and printing variable compliance. It consists of an automatic or interactive simulation buckle that allocates the VAL 2.0 for reaching target walls and, if necessary, produce tool rotations for adapting VAL 2.0 range, enabling the inclusion and printing of perpendicular wall parts without collision. If this is still unsuccessful, changes in wall grouping and spilling locations need to be made.

3.3.2 Toolpath Analysis

This step delivers a diverse set of diagrams (see Table 2.) to verify wall and program integrity, along with recommendations for appropriate variable ranges. It serves both the 3DPCDW user and robot technician for identifying potential issues that could disrupt the manipulator's smooth movement and affect quality.

Table 2. Different *diagram* reports for code analysis.

Diagram	Description
Code Speeds	Controls bead width, avoid bulkiness.
Zone Values	Sharpness and aesthetic patterns.
Setting Times	Diagrammed for each layer, section (*feet*) and walls.
Rotation vs. Length	Gradients for speed issues from 6th Axis rotation coordination.
Code Parts and Segregation	Program compliance regarding location, start, finish, and *robotStops* for the whole task.

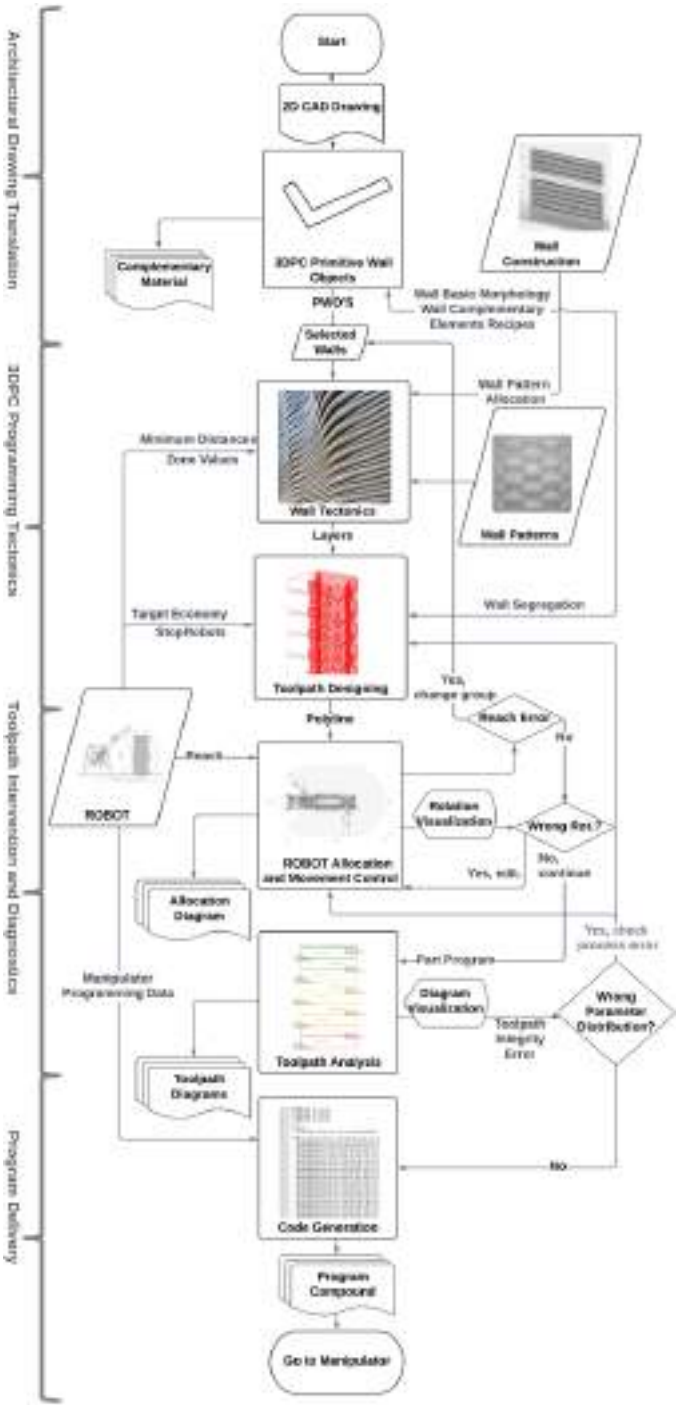


Figure 3. 3DPCDW structure.

Figure 4. Perspective view of 3DPC Walls.

3.4 Program Delivery

This task generates a *Program Compound* consisting of 4 codes for completing a 3DP task through pre-evaluation and printing; first one consists of *Dry-run* and *Probe* programs that enable the robot technician to ensure material and printing task program compliance respectively, where the second one executes the actual printing task (see Table 3).

Table 3. Content of a *Program Compound* delivered by 3DPCDW.

Type	Description
Main	Main routine, for running program parts.
Part	Program part, involving total or a wall(s) portion modulated by sections.
Dry-run	Dry-run file for program compliance through simulation.
Probe	Program with different probe shapes for structural testing.

4 Results

A total of 25 walls were produced to host 26 townhouses in Gananoque. These walls included both straight and corner walls. Results are addressed in terms of: Preparation, Printing Task and Printed Results.

4.1 Preparation: Digital 3DPCDW Outcomes

There were effectively 19 program compounds required to complete the 25 3DPC walls of the project, specifically: {*StSp*=9, *StGp*=2, *DtSp*=4, *DtGp*=8}, Notably *StSp* prints were preferred in the last month due to the capabilities achieved, practicality and compatibility to the period mix. The main task was employed exclusively on prototyping, while the probe was used in 1 of every 5 prints. *Dry-run* and *Parts* were always used.

4.1.1 3DPC as a Platform

As a platform, there were several updated versions (20) of the 3DPDW, most of which were actualized on site during a prototyping period prior to actual printing. This was critical because of unforeseen situations, such as accommodation of slab and key tolerances, quality concerns and integration of observations that enhanced execution. The tasks were represented by:

- **Integration:** AEC Integration through Wall Construction.
- **Digitalization:** Wall to 3DPC program. Representation subtasks.
- **Generalization:** Through Parametric Wall Objects (PWO), the production of wall and group programs.
- **Flexibility:** Through 3DPCDW parametric interaction, ease of 3DPCDW and algorithm edition.
- **Design Conception:** Through Wall Design Augmentation and PWO compliance, generative building envelope solutions can be produced.

Program versioning (20) represented critical modifications of 3DPCW mostly before the definitive printing period. These involved, for example, adding multi tool capabilities, toolpath adaptation to fit the key, including *probe* in program compound, etc. Other non-vital complements such as adding more *diagrams* to *Toolpath Intervention and Diagnostics* were made in parallel on the definitive print period. Second half of the definitive period only required parametric change.

4.2 Printing Task

Changes in wall grouping and allocation from previous codes were necessary for executing the 3DPC task, mainly due to alterations in the printing sequence, which were constrained by confronting the slab and VAL 2.0 possible movement ranges while moving.

Code reaching errors were found where tolerance with actual project were more than 10 mm from the actual VAL2.0 position. This happened typically on large wall groups where maximum reach was needed.

Changing pattern parameters was necessary when

changed mix between Block 1 and Block 2, reducing wall pattern projection.

4.2.1 3D Printing Task

The Printing task integrates 3DPCDW digitalization. Numbers 4 to 7 are repeated 9 times to get an actual height of 2.75 [m] (see Table 4.):

Table 4. 3D Printing Subtasks.

#	3DP Subtask	3DPCDW Related Output
1	Pump preparation	-
2	Calibration & *Dry-run*	*Alloc. Diagram* *Probe* *Dry-run*
3	Material setup	-
4	Printing 1 feet high	*Part*
5	Change of wall or stop	*Part/robotStop*
6	Install rebar compound	*RebarDiagram*
7	Continue printing next wall or next feet	*Part*
8	Stop printing	*Part/robotStop*
9	watering & covering	-
10	Pump cleaning	-
11	Move into next position or close	*Alloc.Diagram*

- **2:** Calibration and Dry-run time. This 3DP task demanded a few program compound modifications.
- **6:** Installing rebar compound. The *Rebar Diagram* gave correct dimensions for rebar cutting and positioning, while considering high tolerances
- **7:** As printing tasks consisted of many *parts*, loading of next program part, on stop robot location. No considerable time was expended from *part*-to-part loading.
- **11:** Robot allocation was made using the *Allocation Diagram*, where differences of less than 10mm were admitted. Time here was considerably high (10 mins min, 1 hour max) especially when going out of the wall row.

4.2.2 Material Waste

Material waste for group prints (*DtGp* and *StSp*) toolpath strategy waste less material, because of the alternation between walls eliminates waiting spilling time. In the case of individual wall printings, robot stops in spillage zones produced waste while waiting on rebar allocation. Results on material use show an average of 15% of material loss.

4.3 Printed Results

4.3.1 Global Dimensional Results

Walls exhibit rough compliance with length and width measurements, with the majority closely matching the required dimensions, resulting in deviations of 0 to 20mm. These results depend on many factors, non standing changes from Block1 to Block2 established significant difference.

4.3.2 Dimensional Results on Beads

Bead widths show some variation but remain within acceptable tolerances. Although the centres were conceived to be 75 mm ± 5 mm, tolerances can go up to 100 mm because of the *Dripping Pattern*. The wall width and side beads were designed to be 85 ± 5 mm; however, the maximum use of the rail increased the bead width because of speed dampening effects.

4.3.3 Aesthetic Results

Aesthetic results were found expressive, with clear Design Pattern gradients for the *Relief* and *Dripping Patterns*. Different mixes yielded different surface finishes, so a lower Projection Parameter for Relief and Dripping was counterweighted with harsher texture, and softer surface finishes were expressed better with a more higher *Projection Parameter* value (see Figure 6).



Figure 5. Perspective detail view of Wall Patterns.

5 Discussion

3DPC Task execution is highly complex and relays mostly on implicit *knowhow* of the professionals involved [18], so 3DPCDW made a great part of this *knowhow* explicit, integrated and functional.

As 3DPCDW also increased its capabilities over time, retrofitting from the context and then be just operated by parameter change, its *Flexibility* and *Generalization* effectively supported Gananoque Project linkage between materiality by means of *Wall Construction*, design and robotics.

5.1 Printing Task Discontinuity and Times

Regarding times expended in the Printing Task, less than 60% is used on the actual 3D print (see Table 4). Number 1 to 3 and 9 to 11 require up to 40% of the Printing Task. There is some potential in automation of subtask 2 and 11 that require more than the 20% of the Printing Task. On both (2 and 11), the use of diagrams and analogue measurements were performed. What was expressed through diagrams in Gananoque may be automated in the future for a more productive performance. Implementation of *Computer Vision* (CV), positioning automation and other technologies such as *Digital Twins* can be implemented for efficiency, however, what's closer to implement depends on multiple factors.

Printing task discontinuity is not desirable. The two major factors are individual prints execution and rebar positioning, both of which are related. The first one can be tackled by printing only wall groups, however changes in mixing setting time expressed bulking, especially in short walls (see Results: Printed Result). The second factor can be solved by robotic *on the fly* rebar positioning or using better rebar design.

5.2 Printed Results Discussion

Results on material use show a general 15% of material loss. This is almost equivalent to constructing four new walls of 3 mt, including losses. A major amount of this material waste can be mitigated through enabling group printing on 3DPCDW, however, this is caused by the pump's inability to produce extrusion stops.

Another pump enhancement regarding quality issues is the coordination with robot speed, or its control directly by the robot program, for solving vertical bulkiness and horizontal extrusion variations.

6 Conclusion

We presented a platform for creating 3D- printed envelopes by means of Automation of 3DPC. The actual successful implementation of this was demonstrated through the printing of the Gananoque project´s wall units. Raising the technological level will hopefully increase 3DPC team and tasks productivity, producing also more precise and cost-effective products.

PM offer precise control over project walls, enabling unique tectonic expressions through toolpath strategies based on 3DPC knowledge. Insights into manipulator rotations and controlled material falls can enrich 3DPC oriented design by shifting the focus from representation to generation, challenging traditional design approaches.

Generalization, through using both persistent geometry and workflow, enables 3DPCDW with the ability for solving a wide range forthcoming 3D projects. Increment in wall object intelligence regarding performative and practical 3DPC oriented design must keep *Flexibility* of 3DPCDW.

3DPCDW Design Conception missed generative means that are ageless on CAD culture, however, less conventional shapes in architectural design weren't possible until *Digital Fabrication* and 3DPC arrived. Frameworks and the notion of *Design Spaces* for exploration linked with actual predictive performance of the 3DPC physical phenomena can empower different and more interesting architectural spaces, 3DPC tectonics, building performance, ergonomics and other explorative but useful built expressions.

Acknowledgements

We would like to thank the ABC Team for their assistance in construction, Frederic Wright for his helpful insights and actions, and Matthew Hill for all the fun and hard work onsite.

References

[1] N. Perrier et al., Construction 4.0: A Survey of Research Trends, *Journal of Information Technology in Construction*, 25: 416–437, 2020.

[2] E. Forcael, I. Ferrari, and A. Opazo-vega, Construction 4.0 : A Literature Review, *Sustainability*, 12: 1-28, 2020.

[3] E. Forcael, M. Medina, A. Opazo-Vega, F. Moreno, and G. Pincheira, Additive manufacturing in the construction industry, *Automation in Construction*, 169:1-21, 2025.

[4] F. Mondragón-Serrano, A. Padilla-Viveros, G. Hernández, and J. V. Calderón-Salinas, The Conceptual System of Additive Manufacturing via a Quantitative-Qualitative Approach: A Bibliometric Analysis, *Journal of Scientometric Research*, 12(3): 657–669, 2023.

[5] M. K. Niaki, S. A. Torabi, and F. Nonino, Why Manufacturers Adopt Additive Manufacturing Technologies: The Role of Sustainability, *Journal*

of Cleaner Production, 222: 381–392, 2019.

[6] D. Hwang and B. Khoshnevis, An innovative construction process-contour crafting (CC), In *22nd International Symposium on Automation and Robotics in Construction*, pages 1-6, Ferrara, Italy, 2005.

[7] A. K. Al-Tamimi, H. H. Alqamish, A. Khaldoune, H. Alhaidary, and K. Shirvanimoghaddam, Framework of 3D Concrete Printing Potential and Challenges, *Buildings*, 13(3): 1-27, 2023.

[8] O. Davtalab, A. Kazemian, and B. Khoshnevis, Perspectives on a BIM-integrated software platform for robotic construction through Contour Crafting, *Automation in Construction*, 89: 13–23, 2018.

[9] P. Peralta Abadia, S. Heine, H.-M. Ludwig, and K. Smarsly, Printing Information Modeling (PIM) for Additive Manufacturing of Concrete Structures, *Applied Sciences*, 13:1-27, 2023.

[10] P. P. Abadia and K. Smarsly, An algorithmic BIM approach to advance concrete printing, In *28th International Workshop on Intelligent Computing in Engineering (EG-ICE)*, pages 1-10, Berlin, Germany, 2020.

[11] N. Labonnote, A. Rønnquist, B. Manum, and P. Rüther, Additive Construction: State-Of-The-Art, Challenges and Opportunities, *Automation in Construction*, 72: 347–366, 2016.

[12] C. Eastman, P. Teicholz, R. Sacks, and K. Liston, *BIM Handbook, 2nd Edition*, Wiley, New Jersey, 2011.

[13] P. Cascone, M. Laddaga, and F. Forestiero, Digital Construction: 3D Printing for Performative Houses, In *2nd RILEM International Conference on Concrete and Digital Fabrication, Digital Concrete 2020*, pages 804–813, Eindhoven, The Netherlands, 2020.

[14] P. Cascone, F. Galdi, A. Giglio, and E. Ciancio, Architectural Self-Fabrication, In *International Journal of Parallel, Emergent and Distributed Systems*, 32: S39–S53, 2017.

[15] M. Alabbasi, H. Chen, and A. Agkathidis, Developing a Design Framework for the 3DPrinting Production of Concrete Building Components, In *The 9th Arab Society for Computer Aided Architectural Design*, pages 713–726, Cairo, Egypt, 2021.

[16] O. Kontovourkis and G. Tryfonos, Integrating Parametric Design with Robotic Additive Manufacturing For 3D Clay Printing: An Experimental Study, In *35th International Symposium on Automation and Robotics in Construction*, pages 918-925, Berlin, Germany, 2018.

[17] L. Breseghello and R. Naboni, Toolpath-Based Design for 3D Concrete Printing of Carbon-Efficient Architectural Structures, *Additive Manufacturing*, 56: 1-12, 2022.

[18] F.P. Bos, C. Menna, M. Pradena, E. Kreiger, W.R. Leal da Silva, A.U. Rehman, D. Weger, R.J.M. Wolfs, Y. Zhang, L. Ferrara, V. Mechtcherine, The Realities of Additively Manufactured Concrete Structures in Practice, *Cement and Concrete Research*, 156: 1-14, 2022.

Indoor Comfort Assessment Based on a Digital-Twin Platform

Chia-Ying Lin[1] and I-Chen Wu[2]

[1] Department of Civil Engineering, National Kaohsiung University of Science and Technology, Taiwan
[2] Department of Civil Engineering, National Kaohsiung University of Science and Technology, Taiwan
F112141104@nkust.edu.tw, kwu@nkust.edu.tw

Abstract –

Indoor comfort is shaped by many factors, such as thermal comfort, air quality, and lighting. Traditional building-management systems often struggle to bring all these elements together and make real-time adjustments. To tackle this challenge, this study develops a digital-twin platform that combines indoor-comfort assessment with intelligent environmental control to realize real-time monitoring, accurate analysis, and efficient management of various environmental parameters. The system is based on building information modeling (BIM) and incorporates real-time sensors and camera data, which are processed to assess and adjust core comfort factors. These include thermal indices (predicted mean vote and predicted percentage of dissatisfied), air-quality metrics (CO2 levels), and lighting conditions like illuminance and color temperature. The platform utilizes Unity for environmental simulation and visualization, integrating smart lighting and air-quality sensors to deliver real-time feedback and control. To demonstrate its application, the study focuses on three common scenarios in educational settings: classroom mode, relaxation mode, and presentation mode. The system can automatically fine-tune environmental settings to meet specific user needs and enhance comfort. The results show that the system successfully bridges monitoring and control across thermal comfort, air quality, and lighting, offering a practical and efficient solution for smart-building management.

Keywords –

Digital Twin; Building Information Modeling; Thermal Comfort; Indoor Air Quality; Visual Comfort.

1 Introduction

With the rise of modern smart-building technologies, there has been growing attention on creating human-centric indoor environments that prioritize comfort. This includes addressing key aspects like thermal comfort, air quality, and lighting. However, traditional building-management systems often fall short when it comes to real-time monitoring and control of these factors.

Research shows that integrating digital-twin (DT) technology with smart devices can effectively address the limitations of traditional management systems. For example, Tan et al. [1] developed a digital-twin Lighting (DTL) system that merges building information modeling (BIM) and computer-vision technology, enabling efficient energy use alongside real-time environmental control. Their testing demonstrated an impressive 95.15% decision-making accuracy and energy-cost savings of around 79%. Similarly, Leplat et al. [2] applied DT models for multivariate simulations, achieving a balance between visual needs, ecological concerns, and lighting design, further showcasing the versatility of DT technology.

To overcome these limitations, this study proposes an integrated system based on DT technology for assessing indoor comfort and enabling intelligent environmental control. By integrating BIM, Internet of Things (IoT) architecture, and smart devices, the system enables dynamic management and optimization of multiple environmental parameters, ultimately enhancing overall comfort and user experience.

The essence of DT technology lies in seamlessly connecting physical objects, virtual models, and data to simulate and enhance the operational efficiency of physical systems [3]. While BIM has been extensively used during the design and operational phases of buildings, it struggles with integrating dynamic data and enabling adaptive control, particularly when coordinating multiple systems like HVAC, lighting, and fire safety. Additionally, challenges persist in real-time data processing and achieving standardization [4]. As modern buildings place greater emphasis on health and efficiency, the need for dynamic management of thermal comfort, air quality, and lighting has become increasingly important in smart building design. For example, maintaining optimal temperature and air

circulation can significantly boost physical well-being, while lighting tailored to user needs can improve mental health, reduce stress, and enhance productivity in both learning and work environments [5].

International standards play a crucial role in enhancing indoor environmental quality. The EN 12464-1 standard offers detailed guidelines for lighting across various spaces, focusing on achieving a balanced approach to health, psychological comfort, and visual needs [6]. Appropriate adjustment of environmental parameters is essential for ensuring comfort in different scenarios, as ideal color temperature and illuminance levels can vary greatly between relaxation and work settings [7].

Building on this foundation, this study incorporates smart-lighting and environmental-sensing technologies to create a DT-based system designed for the dynamic assessment and control of thermal comfort, indoor air quality, and lighting comfort. The system simulates three common scenarios: classroom mode, relaxation mode, and presentation mode, all aligned with the EN 12464-1 standard. Through experimental evaluations, it assesses the system's effectiveness in enhancing comfort and promoting health. This research offers a practical solution for managing indoor environments in smart buildings and serves as a reference for future human-centered architectural design.

2 System Framework

The system framework is shown in Figure 1. This study adopts the EN 12464-1 lighting standard to design three typical scenario modes: classroom mode, relaxation mode, and presentation mode. Lighting parameters were customized to align with the specific requirements of each scenario. The system integrates BIM with a DT platform to enable comfort simulation and environmental control. For data processing, IoT sensors gather real-time data on temperature, humidity, and carbon-dioxide (CO2) concentration, which are further complemented by image recognition technology for detailed analysis. All data are stored in a PostgreSQL database, supporting data visualization, dynamic updates to environmental models, and intelligent decision-making capabilities. The DT platform dynamically evaluates thermal comfort, indoor air quality, and lighting comfort, focusing on predicted mean vote (PMV), predicted percentage of dissatisfied (PPD) analysis, CO2 concentration heatmaps, and intelligent lighting control. PMV and PPD are important indicators in the ISO 7730 and ASHRAE 55 standards for assessing indoor thermal comfort. PMV represents the average thermal sensation of individuals in a given environment, while PPD estimates the percentage of people likely to experience discomfort. To maintain a comfortable indoor climate, PMV values are generally recommended to stay between −0.5 and 0.5.

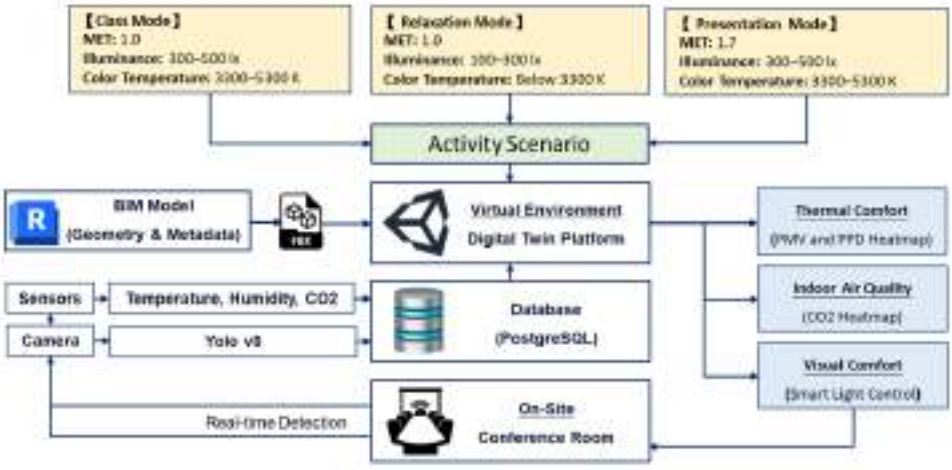


Figure 1. System framework for this research.

2.1 Conventional vs. Proposed Data Acquisition

Traditional DT platforms primarily rely on sensor data and manually entered information to construct digital models. However, this approach can lead to delays in data updates and human errors, making it difficult to reflect real-time changes in the indoor environment.

The proposed method improves automation and data timeliness by using sensors to continuously transmit environmental data, such as temperature, humidity, and CO2 concentration. Additionally, it integrates image-recognition technology to automatically estimate the number of occupants and their clothing insulation levels. These values are then used to calculate thermal comfort indicators like the PMV index. This reduces manual input errors and allows for real-time adjustments to environmental control strategies, making the DT model more responsive to actual conditions. A comparison of

these two approaches is presented in Figure 2.

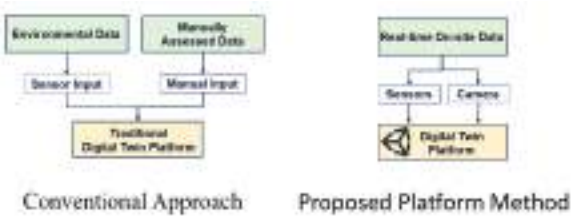


Figure 2. Comparison of two data-acquisition methods.

2.2 Data Processing and Integration Workflow

First, this study uses Revit 2023 to create a BIM model that accurately reflects the characteristics of the experimental environment. To preserve material appearance during the modeling workflow, 3ds Max was used for material conversion. Autodesk-native materials were transformed into Physical Materials to avoid material loss during export. Following this step, the model was exported in FBX format for seamless integration into the Unity platform, supporting the development of the DT.

The system integrates mass data from diverse data sources, including temperature and humidity sensors, CO2 sensors, image recognition tools, and smart lighting systems. Real-time environmental data are obtained from temperature and humidity sensors, while image recognition is applied to detect room occupancy and evaluate clothing insulation levels (Icl), which are then used to calculate PMV and PPD values. Indoor air quality is assessed by monitoring and calculating CO2 concentrations using dedicated sensors. Moreover, the smart-lighting system manages dynamic lighting control based on predefined comfort parameters. This integrated dataset is processed to generate visual outputs, such as PMV and PPD heatmaps, CO2 concentration heatmaps, and lighting simulation scenarios. These visualization results can help users better understand environmental conditions and make informed decisions about adjustments.

The system utilizes continuous data streaming and real-time feedback mechanisms to dynamically adjust indoor environmental parameters, addressing the need for real-time monitoring and enhancing comfort in smart buildings. The following sections will provide a detailed overview of the analysis and implementation methods for thermal comfort, indoor air quality, and visual comfort.

2.3 Thermal Comfort Analysis

This study provides a detailed analysis of indoor thermal comfort primarily by using the PMV and PPD indices to assess comfort under various environmental conditions. The PMV index, introduced by Danish engineer Fanger [8], measures human thermal comfort in a specific setting. It takes into account several environmental factors, such as air temperature, radiant temperature, air velocity, and relative humidity, as well as personal factors like clothing insulation and metabolic rate. The calculation formula is as follows:

$$PMV = (0.303e^{-0.0036M} + 0.028) \times \{ M - 3.05 \times 10^{-3} \times (5733 - 6.99M - p_a) - 0.42 \times (M - 58.15) - 1.7 \times 10^{-5} \times M \times (5867 - p_a) - 0.0014 \times M \times (34 - t_a) - 3.96 \times 10^{-8} f_{cl} \times [(t_{cl} + 273)^4 - (tr + 273)^4] - f_{cl} \times h_c \times (t_{cl} - t_a) \} \quad (1)$$

The parameters used in the formula are summarized in Table 1, while the PMV index categorizes thermal sensations into seven levels, as shown in Figure 3.

For PMV calculation, the following data are obtained from IoT sensors:

- Air temperature (ta): Measured using the AM2320 temperature and humidity sensor.
- Water vapor pressure (Pa): Calculated based on the relative humidity measured by the AM2320 sensor, using the equation (see Equation (2)), where RH represents the relative humidity.

$$Pa = RH \times 10 \times e^{(16.6536 - \frac{4030.183}{ta + 235})} \quad (2)$$

Other parameters are computed programmatically as follows:

- Metabolic rate (M): Assigned based on the type of activity, e.g.,
  1. Resting (sitting): 60 W/m²
  2. Light walking (office walking):100 W/m²
- Mechanical work (W): Assumed to be 0, as typical classroom environments do not involve significant physical exertion.
- Clothing surface temperature (tcl): Determined iteratively based on the mean radiant temperature (tr) and air temperature (ta):

$$tcl = 35.7 - 0.028M - Icl \times [3.96 \times 10^{-8} \times fcl \times (tcl + 273)^4 - (tr + 273)^4 + fcl \times hc \times (tcl - ta)]; \quad (3)$$

- Clothing surface area factor (f_{cl}):

$$fcl = \begin{cases} 1 + 1.29Icl, & \text{if } Icl \le 0.78 \frac{m^2 k}{w} \text{; and} \\ 1.05 + 0.645Icl, & \text{otherwise} \end{cases} \quad (4)$$

- Convective heat transfer coefficient (h_c):

$$\mathrm{hc} = \begin{cases} 2.38 \times |\mathrm{tcl} - \mathrm{ta}|^{0.25} \\ \text{if } 2.38 \times |\mathrm{tcl} - \mathrm{ta}|^{0.25} \geq 12.1 \times \mathrm{Var}^{0.5}. \\ 12.1 \times \mathrm{Var}^{0.5}, \text{ otherwise} \end{cases} \quad (5)$$

Finally, all parameters are applied to the PMV equation for calculation. The resulting PMV values are visualized using a heatmap, providing an intuitive representation of indoor thermal comfort distribution.

In this study, the metabolic rate (M) is a key parameter for PMV calculation and is determined based on the activity-type table from ISO 7730 and ASHRAE 55 standards, as shown in Table 2. The metabolic rates are set as follows for different scenarios: 1.0 MET for classroom and relaxation scenarios and 1.7 MET for presentation scenarios, ensuring accurate PMV calculations. The PPD index is derived from PMV, which indicates the percentage of people likely to feel dissatisfied:

$$\mathrm{PPD} = 100 - 95 \times \exp\left(-0.3353\ \mathrm{PMV}^4 - 0.2179\ \mathrm{PMV}^2\right) \quad (6)$$

When the PMV value is close to zero, the PPD value reaches its minimum, suggesting that most individuals are thermally comfortable. Conversely, as the absolute value of PMV increases, the PPD value approaches 100, reflecting a greater dissatisfaction.

To enhance the dynamics and accuracy of the PMV index, this study integrates YOLOv8 for real-time human detection, recording the number of individuals entering and leaving the space. YOLOv8 is a high-performance object detection framework capable of rapidly identifying and tracking the number and movement of individuals in an environment [10]. These occupancy datasets are used as input parameters for the PMV calculation, enabling real-time comfort evaluation. Additionally, this study employs the UNet (ResNet50) deep-learning model to identify clothing types, such as long sleeves and short sleeves, for calculating clothing insulation levels, thus providing critical data for accurate PMV assessment. The recognition results from UNet are shown in Figure 4.

Table 1. The meanings of the parameters in the PMV formula.

Parameter	Description
M	Metabolic rate (heat generated by the human body)
W	External work (amount of work performed by the body)
t_a	Air temperature (temperature of the ambient air)
t_{cl}	Clothing surface temperature (temperature felt on clothing surface)
t_r	Radiant temperature (mean temperature of surrounding surfaces)
f_{cl}	Clothing thermal resistance (resistance of clothing to heat conduction)
h_c	Convective heat transfer coefficient (efficiency of heat transfer between air and skin)
p_a	Water vapor pressure (pressure of water vapor in the air)
I_{cl}	Clothing thermal resistance (clothing's resistance to heat)



Figure 3. Definitions of PMV values according to ASHRAE [9].

Table 2. Activities corresponding to human metabolic rates.

Activity Type	M (W/m²)	M (MET)
Rest – Lying	40	0.7
Rest – Sitting	60	1.0
Rest – Standing	70	1.2
Office – Writing/Reading	60	1.0
Office – Typing	65	1.1
Office – Walking	100	1.7
Office – Organizing Items	120	2.1

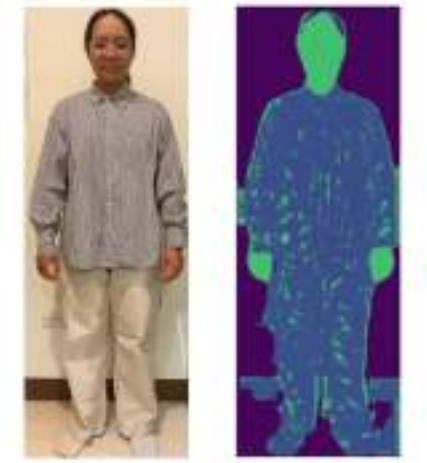


Figure 4. Recognition results for clothing insulation level.

2.4 Air Quality Analysis

This study analyzes indoor air quality by focusing on

the distribution and dynamic changes of CO2 concentrations, which are visualized using heatmaps. This visualization provides an intuitive illustration of spatial CO2 levels across indoor areas, along with real-time concentration trends, thereby providing essential data for assessing and regulating indoor air quality. To further quantify the air exchange demand, the model proposed by Coley and Beisteiner [11] is used to calculate the indoor–outdoor CO2 exchange rate. The calculation formula is:

$$\mathbf{C_t} = \mathbf{C_{ex}} + \frac{\mathbf{G}}{\mathbf{Q}} + \left(\mathbf{C_{in}} - \mathbf{C_{ex}} - \frac{\mathbf{G}}{\mathbf{Q}}\right) \times \mathbf{e}^{-\left(\frac{\mathbf{Q}}{\mathbf{V}}\right)\mathbf{t}}. \quad (7)$$

In this formula, the parameters represent the following: C_t is the indoor CO2 concentration at time t (ppm), C_{ex} is the outdoor CO2 concentration (ppm), C_{in} is the initial indoor CO2 concentration (ppm), G is the indoor CO2 generation rate (ppm/h), Q is the ventilation rate (m^3/h), and V is the indoor space volume (m^3).

To accurately compute air exchange values, the following data are obtained from IoT sensors:

- Outdoor CO2 concentration (C_{ex}): Measured using the DS-CO2-20 sensor, representing the CO2 level in the outdoor environment (ppm); and
- Initial indoor CO2 concentration (C_{in}): Measured using the DS-CO2-20 sensor, representing the initial CO2 concentration inside the space (ppm).

Other parameters are derived from computational models as follows:

- CO2 generation rate (G): Determined based on the number of occupants in the space. Each person produces a certain amount of CO2 per second. In this study, the average human CO2 emission rate is assumed to be 40 mL/min.
- Ventilation rate (Q): The ventilation volume per cycle is set at 50 m^3. Accordingly, the ventilation rate is determined using:

$$Q = \frac{50}{60 \times 60}\ (m^3/s)\ . \quad (8)$$

- Indoor space volume (V): The experimental space has a volume of 118.62 m^3.
- Indoor CO2 concentration over time (C_t): The CO2 level at time t is determined based on ventilation and generation rates:

$$Ct = Q \times T\ (ppm). \quad (9)$$

This study integrates sensor data with computational modeling to accurately capture dynamic CO2 concentration changes within indoor environments, enhancing real-time monitoring and decision-making, as illustrated in Figure 5.

Furthermore, this study combines heatmap visualization with mathematical modeling to develop a dynamic air-exchange recommendation system. CO2 heatmaps provide a spatial representation of indoor CO2 distribution, enabling the system to identify high-concentration areas for targeted air-quality regulation. The air-exchange formula is then applied to predict concentration changes and determine the necessary ventilation rate. For example, in scenarios where CO2 levels exceed the threshold, the system may recommend an exchange rate of 50 m3/hr.

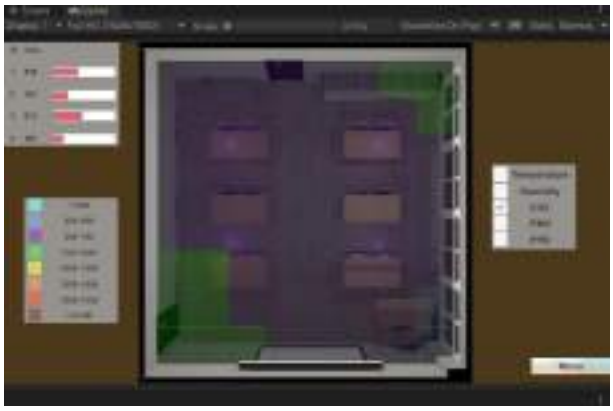


Figure 5. CO2 heatmap.

2.5 Visual Comfort and Intelligent Control

This section focuses on visual comfort and intelligent smart control, combining international lighting standards with advanced technologies to optimize lighting design and enable dynamic management for various scenarios in educational spaces.

First, lighting parameters for each scenario are defined based on the EN 12464-1 standard. The system then adjusts the lighting according to the activity type, aiming to strike a balance between comfort and energy efficiency. To assess the performance of smart luminaires under different conditions, an illuminance meter was used to measure brightness in specific zones. These measurements help map the relationship between actual brightness and illuminance, facilitating in-depth analysis of the lighting environment.

For intelligent control, the system integrates Yeelight smart luminaires with the Unity development platform. By enabling LAN control, the system establishes communication with the luminaires and uses TCP protocols to send commands for remote operation. To make it user-friendly, an intuitive interface was developed, allowing users to easily adjust brightness and color temperature with sliders and buttons. This setup supports the simulation of multiple lighting scenarios, showcasing the design's practicality and flexibility.

2.5.1 Scenario Parameter Definition

To balance lighting comfort and energy efficiency, this study establishes indoor lighting parameters for educational spaces in accordance with the EN 12464-1 standard. Key parameters like illuminance and color temperature are tailored for each scenario and adjusted based on user requirements.

As shown in Table 3 and referencing the EN 12464-1 standard's color temperature classification, neutral tones were selected for scenarios requiring focused attention, while warm tones were used in relaxation modes to promote users' psychological comfort. The standard also offers specific lighting recommendations for educational spaces, as summarized in Table 4. This study uses the abovementioned tables as a reference. Lighting parameters were tailored to meet the needs of each scenario, with detailed configurations presented in Table 5.

Table 3. Classifications of lighting color [4]

Color Appearance	Correlated Color Temperature (TCP) K
Warm	Below 3300 K
Intermediate	3300 to 5300 K
Cool	Above 5300 K

Table 4. Lighting requirements for educational buildings [4].

Type of Room, Task, or Activity	Maintained Illuminance (lx)	Remarks
Classrooms, tutorial rooms	300	Lighting should be controllable.
Lecture hall	500	Lighting should be controllable.
Student common rooms and assembly halls	200	

Table 5. Recommended lighting parameters for different scenarios based on EN 12464-1 standards.

Scene	Illuminance Range (lx)	Color Temperature Range (K)
Class Mode	300–500 lx	3300–5300 K
Relaxation Mode	100–300 lx	Below 3300 K
Presentation Mode	300–500 lx	3300–5300 K

2.5.2 Smart Luminaire Control System

To enable real-time adjustments, this study developed a user interface (UI) on the Unity platform, offering users with an intuitive way to control the luminaires. Through sliders and buttons on the interface, users can instantly adjust the brightness and color temperature, simulating different lighting scenarios based on contextual needs. To further enhance automation and adaptability, this study integrates a real-time occupancy-based lighting mode switching mechanism. Using YOLOv8 for human detection, the system monitors the number of people entering and exiting the space and dynamically adjusts the lighting mode according to predefined thresholds:

- Activation Threshold: When eight or more individuals enter the room, the system automatically switches to Classroom Mode or Presentation Mode, ensuring appropriate illumination for learning or presentation activities.
- Delay Mechanism: A one-minute delay is applied before switching modes to confirm stable occupancy and to prevent frequent unintended fluctuations.
- Deactivation Threshold: If the number of occupants drops to three or fewer, the system transitions to Relaxation Mode, minimizing unnecessary energy consumption when the room is sparsely occupied.

This system integrates automated mode-switching with manual control via the Unity-based UI, enhancing user experience while ensuring optimal lighting conditions and energy efficiency.

To enable remote operation, the "LAN Control" feature is activated in the Yeelight application, allowing luminaires to receive commands over the local network. The system then sends a UDP broadcast request to detect available luminaires, retrieving their IP addresses and ports to establish a stable connection. Once connected, TCP communication is used to transmit JSON-formatted commands, enabling functions such as turning the luminaires on/off, adjusting brightness, and modifying color temperature.

3 Case Study

This study selected the BIM Research Center of the Department of Civil Engineering at the National Kaohsiung University of Science and Technology as the experimental site, covering an area of approximately 45.62 m^2. To simulate the indoor lighting environment, the space was divided into four primary zones, with one light fixture positioned at the center of each zone, as shown in Figure 6. In line with the space's lighting requirements, this study used four Yeelight Moonshadow

550 ceiling lights, evenly distributed to meet demands of visual comfort. This configuration significantly improves light uniformity, reduces glare, and enhances overall visual comfort.

To demonstrate the application of the DT platform, this section uses "Relaxation Mode" as an example to assess indoor comfort through system operation and analysis. The platform integrates real-time environmental monitoring and visualization technologies, enabling a comprehensive evaluation of thermal comfort, air quality, and visual comfort.

For air-quality analysis, the system uses CO2 heatmaps to display real-time comfort levels across different zones, based on continuous data from four CO2 sensors. Additionally, the system offers real-time recommendations for the necessary air-exchange rate based on the monitoring results, allowing users to adjust air quality effectively.

For thermal comfort, the system utilizes the PMV/PPD index to monitor real-time environmental data and employs heatmap visualization to display the distribution of thermal comfort, as shown in Figure 7. Experimental results show that overall thermal comfort within the space stays within standard ranges, thereby satisfying the needs of human comfort.

In terms of visual comfort, under "Relaxation Mode", the lighting is set to a warm, low-light mode with a color temperature below 3300 K. This design effectively reduces the light intensity and creates a relaxed and comfortable atmosphere. The UI of the DT platform displays key parameters such as lighting intensity, color temperature, and thermal comfort metrics, as shown in Figure 8. Experimental results reveal that the virtual images generated by the DT platform align closely with the real-world environment, further validating the accuracy and reliability of the system's data, as shown in Figure 9.

To ensure data accuracy, this study used an illuminance meter to measure the brightness of smart lights and their corresponding illuminance levels in different scenarios. Measurements were taken at the center of the space, as shown in Figure 10.

During the testing, the brightness of the smart lights was gradually adjusted from 10% to 100%, with readings taken every 10%. The recorded illuminance values were then compiled into a reference table, as shown in Table 6, which displays the illuminance distribution in the space under different brightness settings. This approach ensures systematic and precise data collection, providing a robust foundation for lighting comfort analysis and validation of the smart-lighting control system.



Figure 6. Spatial lighting configuration diagram.

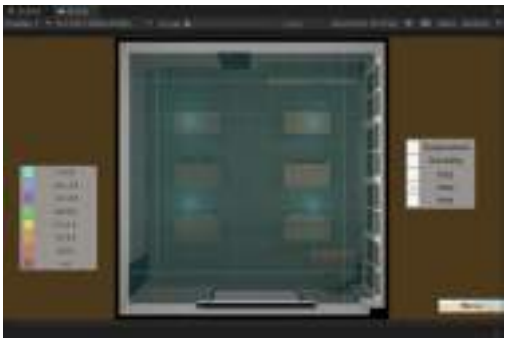


Figure 7. PMV heatmap



Figure 8. Parameter setting in relaxation mode.



Figure 9. Comparison of virtual and physical scenes.

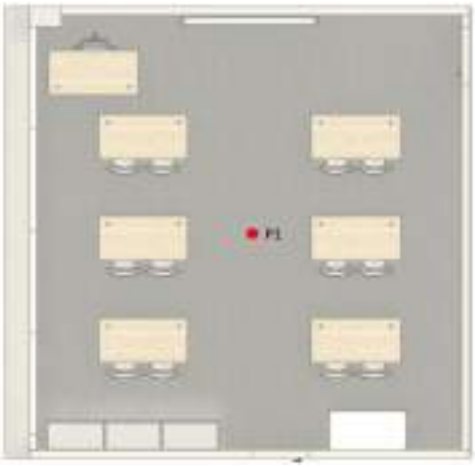


Figure 10. Sampling diagram for illuminance measurements.

Table 6 Measured illuminance values corresponding to lamp brightness levels.

Lamp Brightness (%)	Average Illuminance (lx)
10%	178 lx
20%	235 lx
30%	294 lx
40%	347 lx
50%	397 lx
60%	459 lx
70%	511 lx
80%	564 lx
90%	609 lx
100%	652 lx

4 Conclusions and Future Work

This study used three typical scenarios—classroom mode, relaxation mode, and presentation mode—as case studies to validate the application value and reliability of the digital-twin (DT) platform in assessing indoor comfort conditions. The results demonstrated that the DT platform effectively provides an intelligent and visualized solution for indoor comfort management. It achieves this through real-time environmental parameter monitoring, thermal comfort analysis from predicted mean vote and predicted percentage of dissatisfied, visual comfort assessment, and continuous CO2 concentration tracking. These capabilities collectively improve the quality of indoor environments and accommodate the diverse needs of users across different scenarios.

Future research could broaden the application scope of the DT platform by integrating additional environmental parameters, such as acoustic characteristics and indoor movement patterns. This would enable a more comprehensive assessment and control of comfort. Moreover, with advancements in artificial intelligence and Internet of Things technologies, the incorporation of deep-learning models for user-behavior prediction and adaptive environmental control will be crucial for enhancing the platform's application value. These future developments will enable more accurate and efficient solutions for smart-building management and will contribute to the design of sustainable spaces.

References

[1] Y. Tan, P. Chen, W. Shou, and A. M. Sadick. Digital twin-driven approach to improving energy efficiency of indoor lighting based on computer vision and dynamic BIM. *Energy and Buildings*, 270:112271, 2022.

[2] L. Leplat, C. López-Alfaro, A. Styve, and R. D. S. Torres. NorDark-DT: A digital twin for urban lighting infrastructure planning and analysis. *Environment and Planning B: Urban Analytics and City Science*, 23998083241272099, 2024.

[3] D. Jones, C. Snider, A. Nassehi, J. Yon, and B. Hicks. Characterising the digital twin: A systematic literature review. *CIRP Journal of Manufacturing Science and Technology*, 29:36–52, 2020.

[4] M. Hauer, S. Hammes, P. Zech, D. Geisler-Moroder, D. Plörer, J. Miller, and R. Pfluger. Integrating digital twins with BIM for enhanced building control strategies: A systematic literature review focusing on daylight and artificial lighting systems. *Buildings*, 14(3):805, 2024.

[5] J. A. Veitch and G. R. Newsham. Lighting quality and energy-efficiency effects on task performance, mood, health, satisfaction, and comfort. *Journal of the Illuminating Engineering Society*, 27(1):107–129, 1998.

[6] CEN. EN 12464-1: Lighting of work places - Part 1: Indoor work places, 2002.

[7] Q. Wang, H. Xu, F. Zhang, and Z. Wang. Influence of color temperature on comfort and preference for LED indoor lighting. *Optik*, 129:21–29, 2017.

[8] P. O. Fanger. *Thermal Comfort. Analysis and Applications in Environmental Engineering*. Danish Technical Press, Copenhagen, 244 p.p., 1970.

[9] K. Karyono, B. M. Abdullah, A. J. Cotgrave, and A. Bras. The adaptive thermal comfort review from the 1920s, the present, and the future. *Developments in the Built Environment*, 4:100032, 2020.

[10] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma. A review of Yolo algorithm developments. *Procedia Computer Science*, 199:1066–1073, 2022.

[11] D. A. Coley and A. Beisteiner. Carbon dioxide levels and ventilation rates in schools. *International Journal of Ventilation*, 1(1), 2002.

Application of BIM Technology to Soil Layer Risk Analysis for Enhanced Accuracy and Informed Decision-Making in Geotechnical Engineering

Wen-Yi Wang[1] and I-Chen Wu[2]

1 Department of Civil Engineering, National Kaohsiung University of Science and Technology, Taiwan
2 Department of Civil Engineering, National Kaohsiung University of Science and Technology, Taiwan
C110141123@nkust.edu.tw, kwu@nkust.edu.tw

Abstract –

While BIM technology is widely used for designing and managing above-ground construction, its applicability in representing underground soil distribution and properties is limited, which can increase risks during planning and construction. This study integrates Dynamo with on-site drilling data to automatically generate underground soil models, which incorporates key attributes such as moisture content, liquid limit, and N-value. This approach provides engineers with a clear and intuitive representation of soil conditions. The core value of the soil model lies in enhancing safety assessment capabilities, enabling engineers and designers to identify potential foundation construction risks, such as diaphragm wall mud inclusion, in advance. This facilitates optimized designs and reduces hazards during construction. Additionally, the model supports the estimation of excavation volumes for reuse, such as utilizing sandy soil for concrete and asphalt concrete production and clay soil for backfilling, ensuring efficient resource allocation. By integrating BIM technology with geological information, this study not only improves the visualization of soil layers but also enhances the safety analysis of foundation engineering and the efficiency of earthwork applications.

Keywords –

BIM Model; Risk Analysis; Geotechnical Engineering; Borehole; Dynamo; Automation;

1 Introduction

The construction industry is mainly divided into two key fields: Structural Engineering (SE) and Geotechnical Engineering (GE). In recent years, BIM methods have been extensively developed and applied in the SE field [1]. However, their adoption in geotechnical engineering has been limited, largely because of inherent uncertainties associated with soil and geotechnical parameters, making the modeling process more challenging.

To reduce risks in underground foundation projects, a geological survey is typically the first step [2]. However, when dealing with complex geological conditions, traditional 2D workflows are often insufficient to fully capture geological details [3]. BIM offers a significant advantage, providing a clear visual representation of the entire strata. By doing so, it helps reduce risks, optimize foundation construction, and significantly improve both efficiency and safety [4].

Applying BIM technology to visualize soil layers using integrated data and 3D visualization can considerably improve the clarity of subsurface understanding and enable more precise engineering planning. During the design phase, a visualized soil model can effectively illustrate stratigraphy, groundwater distribution, and variations in soil properties, providing designers with a more reliable and scientific reference. This approach helps prevent design errors or construction challenges arising from insufficient information [5]. BIM technology is intended for use throughout a structure's entire lifecycle, from design and construction to operation and maintenance. Among these stages, understanding soil conditions is crucial, considering that every structure depends on the stability of its foundation [6]. Traditional geotechnical software primarily focuses on soil modeling and analysis but often operates in isolation from broader engineering workflows. In contrast, this study leverages Dynamo to automate soil layer modeling, ensuring that geotechnical data is not only visualized within a BIM environment but also actively utilized for risk assessment. By integrating geotechnical information directly into BIM workflows, engineers can systematically identify potential hazards such as soil instability, excavation risks, and foundation settlement before construction begins. Using an automated bounding box approach, the system detects critical soil layers with unfavorable properties, enabling

a structured and data-driven risk evaluation process. This approach goes beyond merely using BIM for modeling; it establishes BIM as a decision-support tool, enhancing risk assessment accuracy, optimizing construction planning, and improving safety management through automation.

Moreover, specialized geotechnical software such as Plaxis, FLAC, and GeoStudio provide advanced 3D soil modeling capabilities with integrated physical parameters. However, these tools often function as standalone applications, primarily focused on geotechnical analysis. This study utilizes Dynamo and Revit for geotechnical modeling due to their seamless integration with architectural and structural design workflows. By generating soil models directly within the Revit environment, this approach minimizes the need for software transitions, streamlining the modeling process throughout the engineering lifecycle.

Based on these considerations, this study developed a Dynamo-based soil model generation program. The resulting BIM model emphasizes soil layer definitions and properties by integrating the digital terrain model and borehole data. This approach integrates scattered geological data into 3D models that incorporate both geometry and soil parameters. These models provide designers and construction teams with a clear and intuitive visualization of the underground structure and its characteristics.

2 Methodology

Accurate geotechnical risk assessment is challenging due to the complexity and variability of subsurface conditions. Traditional methods rely on borehole data, but inconsistencies in data quality and fragmented soil layers often lead to uncertain risk evaluations. Without a structured approach, these uncertainties can result in inaccurate estimations of soil stability, increasing the risk of ground movements, excavation failures, or foundation settlement. The key issue is not merely generating a 3D soil model but ensuring that it systematically identifies and assesses geotechnical risks before construction. However, due to the variability of soil layers, the flexibility of automated modeling is limited [8], making it challenging to standardize risk assessment and reliably detect hazardous zones, particularly in projects with heterogeneous or discontinuous soil formations.

To address these challenges, this study develops a systematic risk identification framework that integrates borehole data into a BIM-based workflow. By analyzing key parameters such as the liquid limit, N-value, and moisture content, the framework enhances risk assessment by identifying potential instability zones.

The system framework (Figure 1) illustrates the integration of geotechnical data with BIM-based risk assessment. The process begins with field drilling data to obtain borehole depth and soil layer information, while InfraWorks and Civil 3D are used to acquire surface elevation data, generating a digital elevation model (DEM) for an accurate representation of terrain conditions. These data are then imported into Dynamo to automatically generate soil surfaces. However, due to the variability in soil layer structures, the modeling process may fail, resulting in errors or incomplete models. To address this issue, the system utilizes error data from Dynamo for corrections. Excel is used as a data processing tool to filter and adjust abnormal points, ensuring that elevation values align with surrounding soil layers and maintaining model stability and accuracy. The corrected data are then re-imported into Dynamo to generate the final 3D soil model.

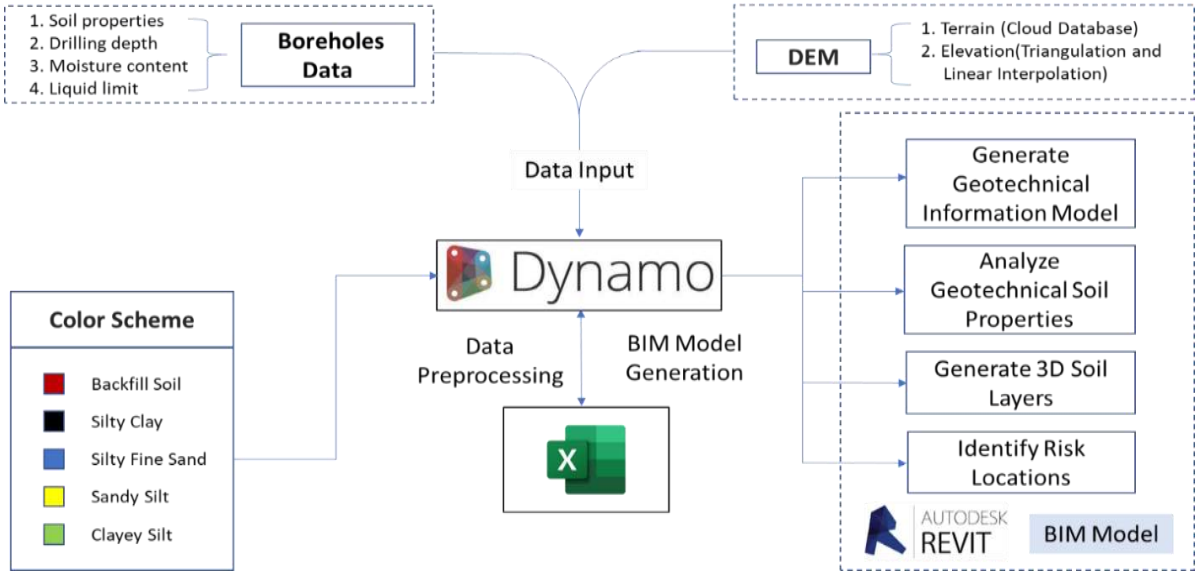


Figure 1. System framework

Once the model is established, a risk identification algorithm embedded in Dynamo is used for automated risk detection. Using the bounding box method, the identifies high-risk zones based on predefined criteria, such as the liquid limit, N-value, and moisture content, and automatically generates bounding boxes around them. These risk zones are then visualized in Revit, providing engineers with an intuitive, data-driven geotechnical risk assessment.

3 Soil BIM Model Generation

3.1 Digital Elevation Model

Before establishing a 3D soil model, accurate surface elevation data must be collected. This data represents ground variations and serves as the foundation for modeling subsurface conditions. It is also essential for calculating earthworks, such as excavation or filling. In this study, tools such as InfraWorks and Civil 3D are used to generate elevation models, automatically generating surfaces from survey data for further analysis.

First, the geographic data in InfraWorks primarily originates from various data platforms and formats, such as Bing Maps and GIS files. This data is used to generate accurate surface models and present detailed surrounding terrain features. InfraWorks can efficiently integrate these data types and automatically generate 3D terrain models. The data obtained in this study using InfraWorks includes elevation points (X, Y, Z coordinates), contour lines, and a triangulated irregular network (TIN) surface.

Second, after obtaining the terrain surface from the InfraWorks model generator, the terrain data is imported into Civil 3D in IMX format. In Civil 3D, the terrain surface is represented as a TIN, which consists of a series of triangles defined by elevation points, where the X and Y coordinates correspond to the planar position and the Z coordinate represents the elevation value of each point. This structure accurately describes the undulations and details of the terrain surface.

In this study, a point set is created using a 20 × 20 grid, generating 20 rows and 20 columns of points with equal spacing across the terrain surface. Civil 3D calculates the Z value (elevation) for each point based on the imported TIN model. The software automatically detects each point's location on the terrain and applies an interpolation algorithm using the Z values of the three points in the surrounding triangle. This assigns an accurate elevation value to each point, ensuring that the terrain's details are accurately represented.

Once the point set is generated, the points are exported as a DEM file in ASCII format, serving as the data foundation for generating points in Dynamo. This process ensures terrain data accuracy and provides the flexibility to adjust grid density based on specific engineering design and data analysis requirements. The result is shown in Figure 2. This terrain model serves as an essential foundation for the automated soil modeling process.

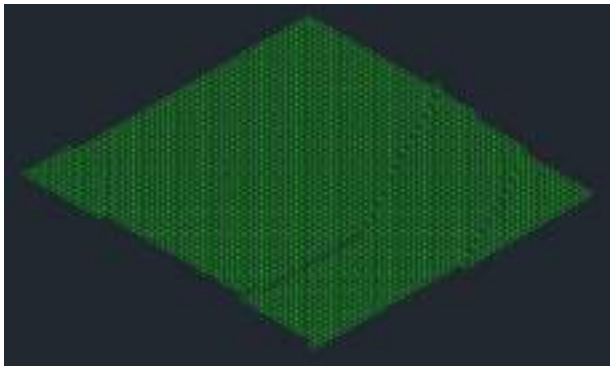


Figure 2. DEM in Civil 3D

3.2 Borehole and Soil Layer Model

After importing the terrain survey data into Dynamo, the data is cleaned by removing invalid characters, extra spaces, and other errors. The goal is to ensure accuracy and consistency, preventing issues in later modeling and analysis stages. Once cleaned, the data is converted into a standardized format, preparing it for subsequent steps.

3.2.1 3D Borehole Generation

The 2D profiles from borehole reports provide detailed soil classification information, including soil types, physical properties, and depth ranges for each soil layer. This data is essential for visualizing the underground soil structure and plays a crucial role in risk assessment and construction planning. In Dynamo, this soil classification data is used to generate depth points, with each point representing the location of a specific soil layer, as indicated in the borehole report.

By identifying these depth points, Dynamo is used to generate a 3D borehole model that illustrates the distribution of different soil layers. This enables clear visualization of the soil structure beneath the excavation area using a color scheme, as shown in Figure 3.

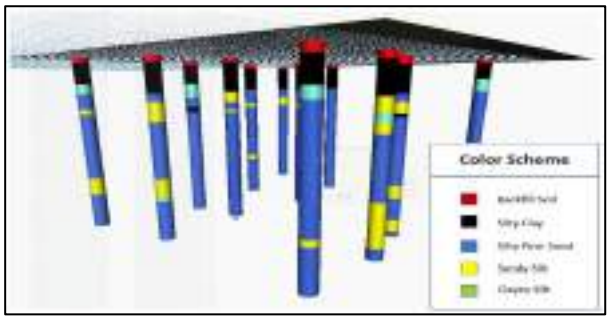


Figure 3. 3D borehole model

3.2.2 Soil Layer Creation

Using color-coded soil classifications and depth information, the distribution of different soil layers can be easily visualized. In Dynamo, soil layers are generated based on depth points from borehole data. These points are connected to form surfaces that represent the structure of the soil layers. The color coding clearly defines the boundaries between different layers, ensuring the model is accurately constructed. The results are shown in Figure 4.

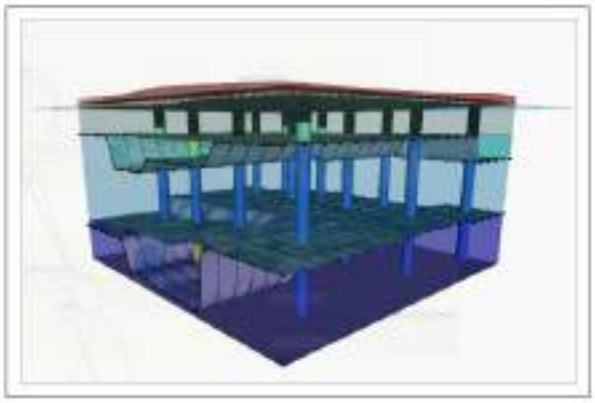


Figure 4. Soil layer model in Dynamo

However, uneven geological data and fragmented soils pose challenges to automated modeling. In some areas, depth points from borehole data may not align properly, resulting in gaps, distorted layers, or unrealistic features in the model. These inconsistencies can cause modeling failures in Dynamo, preventing the successful generation of soil surfaces. To address this issue, error messages from Dynamo are first analyzed to identify problematic data points. When irregular surface formations are detected, the corresponding surface points are extracted and processed in Excel for correction. Using Excel as a preprocessing tool, elevation anomalies are filtered and adjusted by referencing nearby depth values, ensuring smooth transitions between layers. The corrected points are then re-imported into Dynamo, where they form a continuous and stable soil surface. This iterative correction process ensures that fragmented soil layers are more accurately represented and prevents interruptions in the modeling workflow. Once these corrections are made, the final soil layers more accurately reflect the true subsurface, making them useful for further engineering analysis and applications.

3.3 Soil Properties Assignment

In foundation construction, soil properties are critical for ensuring engineering stability and construction efficiency. Parameters such as moisture content, liquid limit, and standard penetration test (N-values) serve as essential indicators for evaluating soil characteristics and potential construction risks. These parameters help identify potential excavation challenges and provide reliable decision-making support before construction begins. Accurately assigning these parameters to the 3D soil layer model significantly enhances the realism and practical value of the model.

Using Dynamo, soil parameters from borehole data are seamlessly integrated into the 3D model through an automated process that matches layer depths and local features, as shown in Figure 5. By assigning soil attributes and incorporating construction data, the model enhances pre-construction decision-making efficiency, reduces risks during construction, and improves project safety and execution. In addition to providing a clear visual representation of the soil structure, the model enables precise volume calculations for land leveling, facilitating the estimation of earthwork quantities.

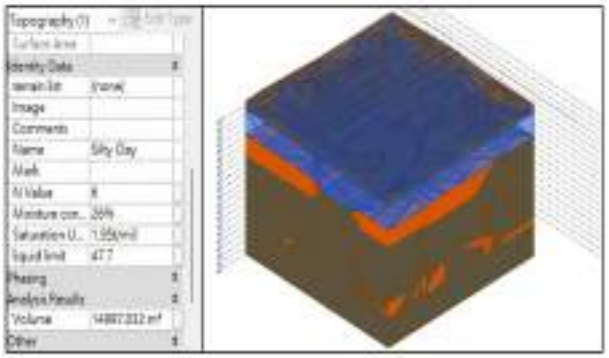


Figure 5. Soil properties assignment

4 Soil Layer Risk Analysis

Integrating geotechnical engineering with BIM technology improves risk assessment and construction management, particularly in addressing soil-related issues that may affect the stability of building foundations. If these problems are not identified and addressed early, they can significantly impact construction safety. Therefore, it is crucial to assess and evaluate these risks at an early stage to ensure a safe construction process.

This study applies BIM technology to soil risk assessment by using 3D models to visualize the underground environment. This approach enables designers to identify potential risks early in the design phase, thereby reducing uncertainty and mitigating risks during construction. Although this method remains in the exploratory stage, it has already demonstrated the potential of BIM technology to enhance risk management and improve construction safety in geotechnical engineering.

This section focuses on the preliminary risk assessment of parameters such as soil moisture content (w), liquid limit (LL), and the standard penetration test (N-value) to identify potential construction risks. These key parameters are used to assess the soil's consistency, stability, and mechanical properties, helping to determine potential challenges during construction.

4.1 Soil Consistency Classification Standards

During construction, potential issues can be estimated by classifying soil hardness using indicators such as the standard penetration test (SPT) N-value, moisture content (W), and liquid limit (LL). This classification provides a solid foundation for assessing construction risks. These indicators help evaluate soil stability and strength, facilitating the prediction of potential problems and the implementation of preventive measures. For example, softer soil layers may require additional support structures, whereas harder layers may necessitate adjustments in construction methods or techniques.

To improve risk identification, this study developed a simplified soil layer table that organizes key characteristics of different soil layers, serving as a quick reference for the on-site construction team. This table enables the team to quickly estimate the properties of various soil layers, reducing the uncertainties caused by soil variations and ensuring improving the flow of the construction process. Additionally, it serves as a reference for the automated modeling in this study, supporting the automation of risk assessment and soil modeling by digitizing and simplifying soil layer characteristics. The classification criteria for clay and sandy soils are as follows:

- Clay Silt (CL, CL-ML)
  - $N \leq 4$ (± 2), $W \geq LL$: Soft clay sediments
  - $N \leq 2$, $W \geq LL$: Very soft clay sediments
  - $N \leq$ 4-15: Medium hard clay
  - $N \geq 15$: Hard clay
- Sandy Soil (SM)
  - $N \leq 10$: Loose sand (from GL-0m to GL-10m)
  - $N \leq 15$: Loose sand (below GL-10m)
  - N = 10-30: Medium dense sand
  - $N \geq 30$: Dense sand

4.2 Relationship Between Soil Properties and Engineering Issues

In foundation engineering construction, the soil characteristics of a site directly influence the stability of construction methods and the quality of the work. Due to variations in soil hardness, moisture content, and stratification, different construction techniques are associated with various engineering risks. Common methods such as diaphragm walls, bored piling, and steel sheet piling may encounter issues such as soil disturbance, leakage, or soil liquefaction under different ground conditions, potentially leading to settlement, tilting, or structural damage to nearby buildings, among other problems.

This section examines the specific problems caused by geological conditions during the construction of diaphragm walls, bored piling, and steel sheet piling. By analyzing case studies, it explores how soil characteristics contribute to engineering issues such as soil disturbance, leakage, and liquefaction.

Studying these cases allows engineers to better understand potential challenges associated with different construction methods in various geological conditions. This insight serves as a valuable reference for both design and construction.

(1) Diaphragm Wall

During diaphragm wall construction in soft clay and silt layers where $N \leq 4$ and $W \geq LL$, or in loose sandy soils where $N \leq 10$, the loose and highly fluid nature of these soils often leads to soil displacement and intrusion into the excavation zone. This can result in the mixing of soil with the diaphragm wall concrete, leading to defects such as soil inclusion, mud contamination, sand infiltration, or large soil lumps. These conditions compromise the structural integrity and load-bearing capacity of the diaphragm wall, degrade construction quality, and may ultimately lead to engineering failures or instability.

(2) Pre-packed Pile

In excavation projects with depths of up to 10 meters, the pre-bored piling method is commonly used for support. However, when constructing in soft soil layers, the cement grout used for the piles—often mixed with fly ash—provides excellent water tightness and high fluidity. While this is beneficial, its high fluidity can also delay the initial setting time, resulting in insufficient early strength. In such conditions, particularly in weak soil layers, the cement grout may leak, leading to a problem known as "Grout Leakage."

(3) Steel Sheet Piles

In excavation projects with depths of up to 15 meters, the steel sheet pile method is commonly used as a retaining structure. However, soil properties play a crucial role in determining both project quality and the safety of nearby structures. Two key issues related to this study are discussed below.

First, when the ground consists of relatively hard layers, such as clayey silt with $N \geq 15 \pm 2$ or sandy soil with $N \geq 20 \pm 2$, driving steel sheet piles becomes significantly more difficult. Forcing the piles into such hard ground may induce large-scale ground vibrations. In stratified ground conditions, if the overall geological

formation is predominantly hard, vibration waves can propagate over a wider range, potentially causing structural cracks in nearby buildings. The impact may extend to a radius of more than 100 meters.

Second, when the ground exhibits a soft upper layer and a hard underlying layer—such as weak clay or silt with $N \leq 4(\pm 2)$ or $W \geq LL$, or loose sandy layers with $N \leq 10$—pile driving may trigger ground liquefaction. In such cases, the localized loss of ground strength may lead to settlement, tilting, or structural cracking in nearby buildings [7].

4.3 Application of Soil Layer Models in Risk Assessment

In this project, a total of 12 boreholes were drilled. For the risk assessment, boreholes located near roadsides or buildings were prioritized, as these areas are more susceptible to external forces and require greater ground stability. Additionally, the borehole with the lowest N-value (indicating the loosest soil) was selected from the drilling data to evaluate ground risks.

As a result, Borehole B9 was chosen as the reference point for risk assessment in this study, with its findings representing the soil conditions beneath the entire excavation surface. While localized variations in soil distribution may exist across the excavation site, selecting the borehole with the lowest N-value provides a worst-case scenario for assessment, facilitating early predictions of potential engineering risks.

This section examines ground behavior during construction using different methods and analyzes the engineering issues that may arise. Particular attention is given to mud-wrapping phenomena in weak soils, sand-wrapping issues caused by sand displacement, and the potential impacts of ground vibrations or liquefaction on nearby structures. These factors can significantly affect construction safety and efficiency, making comprehensive risk assessment essential.

To improve risk evaluation, data is first organized and verified in Excel. After processing, soil layer parameters are imported into Dynamo and assigned corresponding soil properties. Evaluation formulas are then embedded within Dynamo, enabling the automatic generation of bounding boxes to enhance risk identification. Excel was chosen for its compatibility with Dynamo, facilitating efficient data processing and supporting automated risk assessment. The following outlines the methods used for risk evaluation, as shown in Figure 6:

1. For diaphragm walls encountering soft soil with $N<4$ and similar saturated unit weights, the soil layers are nearly identical.
2. Soft clay is characterized by $N \leq 4$ and $W \geq LL$. Therefore, soil with the highest water content and lowest liquid limit is identified as the potential mud-wrapping zone.
3. During diaphragm wall construction, Masako excavation may cause sand to fall, leading to sand-wrapping formation, However, dense sand layers with $N \geq 30$ are less susceptible to this issue due to their high relative density.
4. Driven steel sheet piles with $N > 15 \pm 2$ for clay and $N > 20 \pm 2$ for sand may generate vibrations that could damage nearby structures.
5. When using driven steel sheet piles, soil liquefaction may occur if the strata consist of soft upper layers and hard lower layers.

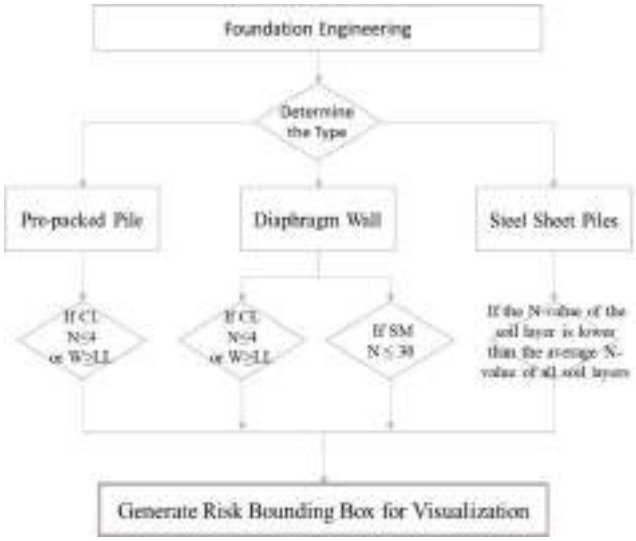


Figure 6. Workflow for generating risk bounding boxes for visualization

This assessment process has been integrated into Dynamo, enabling the automated generation of risk indicators through visual programming. By providing a systematic approach to identifying potential risk areas, this method serves as a reference for engineers in geotechnical risk assessment, design, and construction planning.

In Dynamo, bounding boxes are used to identify risk layers, which are then integrated with the Revit model for automatic generation. This process helps categorize potential risks for each construction method and marks hazardous soil areas with different colors. The visual display enables the project team to identify issues in specific soil layers, such as looseness, leakage risks, or liquefaction, while also providing clear risk indicators. This approach simplifies geological data, allowing the design and construction teams to address key challenges early, plan preventive measures, and reinforce structures. Ultimately, it reduces construction risks and enhances both project quality and safety. The results are shown in the following figure.

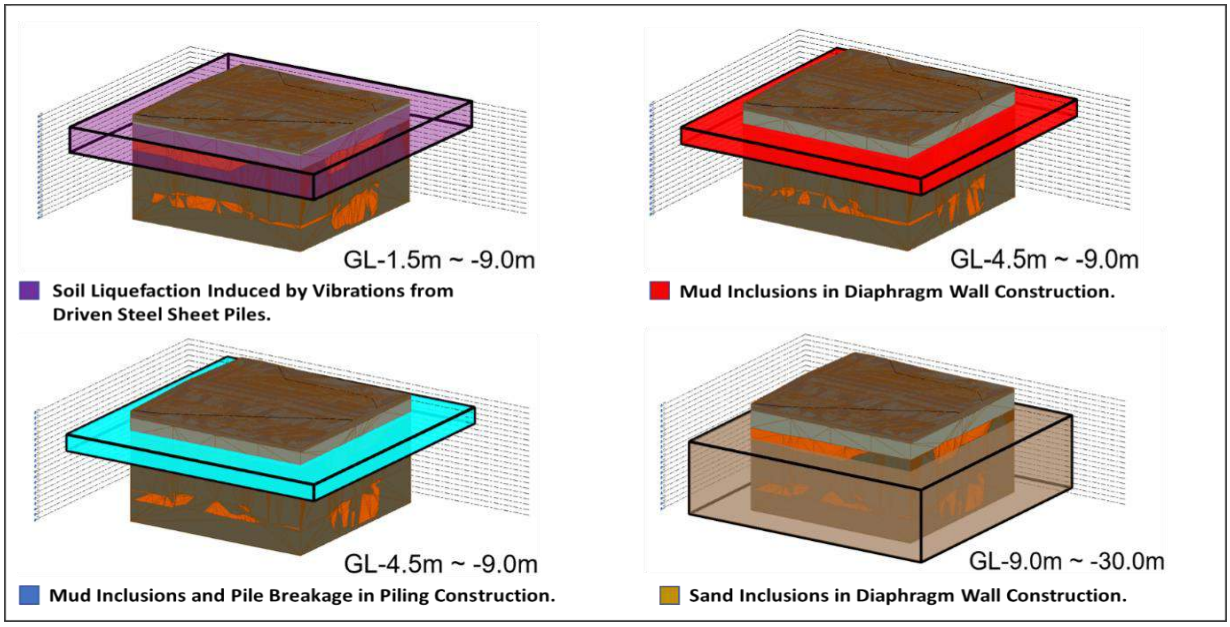


Figure 7. The results of Soil layer risk assessment

Borehole reports indicated that the excavation area primarily consisted of weak clay ($N \leq 4$ (±2), $W \geq LL$) and loose sand layers ($N \leq 10$ (±2)), both of which posed a high risk of instability due to clay inclusion and sand pockets. These unfavorable soil conditions made the site highly susceptible to settlement and collapse, likely contributing to the failure. Had an automated soil risk identification method been applied, these hazardous zones could have been clearly identified before excavation, enabling the project team to implement appropriate mitigation measures and reduce the risk of failure.

This case highlights the practical necessity of integrating automated soil risk assessment methods. By leveraging bounding box analysis in Dynamo, hazardous zones can be detected before excavation, allowing engineers to visualize soil risks directly within the BIM model. This automation improves geotechnical risk visibility, reduces reliance on manual interpretation, and provides a structured framework to prevent similar failures in future construction projects.

5 Conclusion and Future Work

Integrating BIM technology with soil characteristics and risk assessment in geotechnical engineering can enhance the predictability of soil-related issues and the effectiveness of corresponding mitigation measures during construction. Through 3D modeling and data analysis, designers can identify potential soil problems during the design phase and assess risks based on key parameters such as soil stiffness, liquid limit, and SPT N-values. This approach reduces construction risks associated with soil-related challenges and facilitates the development of appropriate strategies for varying soil conditions.

While current methods demonstrate the potential of BIM technology in geotechnical engineering, further improvements are needed to automate the modeling process. For instance, existing risk identification models require validation through additional real-world cases to ensure accuracy. Moreover, these models must be continuously refined to maintain precision as geological conditions evolve, making this an important focus for future research. Additionally, integrating BIM technology with existing soil analysis data remains challenging. In many cases, manual judgment is still required to identify risk areas before computational processing, reducing efficiency and potentially affecting the accuracy and timeliness of data analysis.

Future research may leverage existing soil data for computational analysis to gradually establish an automated system for soil risk assessment and retaining structure modeling. Currently, this study primarily focuses on automating soil layer generation and risk assessment by identifying potential hazards based on soil properties, while groundwater variations are beyond its scope. However, future studies may explore incorporating groundwater level fluctuations into simulations to provide a more comprehensive understanding of subsurface conditions and improve the safety and reliability of both geotechnical risk assessment and automated retaining structure modeling.

References

[1] Z. Obergrießer and A. Borrmann. Infrastructural BIM standards–Development of an Information Delivery Manual for the geotechnical infrastructural design and analysis process. In Proceedings of the European Conference on Product and Process Modeling (ECPPM), London, UK, 2012.

[2] D. Vaníček, D. Jirásko, and M. Vaníček. Role of Geotechnical Engineering in BIM process modelling. In IOP Conference Series: Earth and Environmental Science, Volume 727, 14th Baltic Sea Region Geotechnical Conference, Helsinki, Finland, January 2021.

[3] L. R. Tawelian and S. B. Mickovski. The implementation of geotechnical data into the BIM process. Procedia Engineering, 143:734–741, 2016.

[4] M. Berdigylyjov and H. Popa. The implementation and role of geotechnical data in BIM process. In *E3S Web of Conferences*, Vol. 85, p. 08009, February 2019.

[5] J. Zhang, C. Wu, L. Wang, X. Mao, and Y. Wu. The workflow and operational model for geotechnical investigation based on BIM. *IEEE Access*, Vol. 4:7500–7508, 2016.

[6] I. Vaníček, J. Pruška, and D. Jirásko. BIM Model–Application in geotechnical engineering. In *Proceedings of the 47th Conference on Foundation Engineering*, Brno, Czech Republic, 2019.

[7] M. S. Sheen. *Construction Disaster Prevention Technology: Foundation Construction Edition*. In Chinese, Chapters 2, 3, and 4. Wensheng Book Store, Taipei, Taiwan, 2015.

[8] M. Sandberg, R. Gerth, W. Lu, G. Jansson, J. Mukkavaara, and T. Olofsson. Design automation in construction: An overview. In Proceedings of the 33rd CIB W78 Conference, Brisbane, Australia, 2016.

Large Language Models in Construction Bidding: Opportunities and Risks of Algorithmic Collusion

Chan Heo[1] Moonseo Park[1] and Changbum R. Ahn[1]

[1]Dept. of Architecture and Architectural Engineering, Seoul National Univ., Seoul, Republic of Korea
charlie0119@snu.ac.kr, mspark@snu.ac.kr, cbahn@snu.ac.kr

Abstract–

Algorithmic pricing systems have revolutionized industries, and their application in construction bidding is no exception. Integrating large language models (LLMs) for bid price setting introduces new opportunities and challenges in this domain. This study experimentally demonstrates the capability of LLM agents to generate competitive bidding strategies in a construction environment, demonstrating their ability to adapt rapidly and achieve pricing equilibrium, significantly faster than classic machine learning (ML) methods. Furthermore, we identify prompt-dependent collusion, where subtle variations in input phrasing influence bidding strategies, sometimes resulting in higher bid prices that mimic collusive behavior. Through textual analysis, we reveal how LLM-generated plans align with strategic price adjustments, highlighting the indirect role of prompt design in shaping bidding outcomes. While LLMs offer efficiency and adaptability, their susceptibility to algorithmic collusion raises concerns about transparency and market fairness. The findings underscore the need for regulatory oversight and further research to address the risks and opportunities of LLM integration in competitive construction bidding and broader market contexts.

Keywords –

Algorithmic pricing; Construction bidding; Large language models (LLMs); Algorithmic collusion; Market regulation

1 Introduction

Algorithmic pricing systems have become widely used across industries, leveraging large datasets and computational techniques to propose optimal pricing or bidding strategies [1,2]. From airline tickets and hotel stays to everyday groceries, these systems are applied to various goods sold online, helping sellers maximize profits [3]. The construction industry is no exception. In the project bidding phase, Machine Learning (ML)-based services are increasingly used to analyze competitors' bids and recommend strategies for contractors [4–6]. Using historical bidding data, these systems help users create competitive bids, improving their chances of winning projects and boosting profitability [5,6].

However, the rise of algorithmic pricing has raised concerns among competition regulators about the potential for algorithms to increase prices in ways that resemble human collusion, a phenomenon known as *autonomous algorithmic collusion* [7]. This occurs when algorithms independently learn and adopt collusive strategies without explicit coordination, driven by repeated interactions and optimization processes. A wealth of theoretical [1,8], experimental [9–11], and empirical studies [2,12] has highlighted how algorithmic interactions can create such conditions. Reinforcement learning (RL) methods, such as Q-learning and Deep Q-Network (DQN), have shown in these controlled experiments their ability to implicitly learn and exploit cooperative-like strategies, particularly in iterative pricing environments. Building on these findings, our prior research examined algorithmic collusion within a simulated construction bidding framework [13,14]. In a multi-agent competitive environment, we observed that Q-learning-based agents, through repeated interactions with other learning agents, could autonomously escalate bid prices, mimicking collusive behavior [14]. These insights suggest that algorithmic collusion could also emerge as a pressing issue in the construction domain, given the growing reliance on intelligent bidding systems.

Despite such evidence, some questions remain about whether these results can readily translate to practice in the construction sector. Two primary issues frequently emerge. First, RL-based methods (and other traditional algorithms studied so far) typically demand extensive *trial-and-error* processes, accumulating data from thousands to millions of simulation rounds to converge on an effective policy. Such time- and resource-intensive experimentation may be theoretically sound but faces major practical limitations in real-world construction bidding. Each "trial" in this industry is *expensive*, often accompanied by high project-specific sunk costs and long project durations, making repeated experimentation

infeasible [14]. Second, even if a Q-learning strategy were deployed successfully, any change in the competitive environment—such as a new entrant with different bidding strategies—could abruptly render previous training obsolete, further questioning the sustained practicality of RL approaches in this field [15].

Amid these challenges, recent advances in large language models (LLMs) have sparked renewed interest in algorithmic decision-making for diverse markets [16]. LLMs are typically pretrained on vast dataset, enabling them to perform tasks with minimal additional training and to adapt seamlessly to a variety of contexts. Moreover, they can interact readily with both human users and other algorithmic agents, offering a level of versatility and responsiveness beyond that of many conventional RL methods. Some studies have shown that LLM models have already achieved a mature level of performance in price setting within specific market environments, and in certain cases, they have even exhibited collusive behaviors [15,17]. At this pivotal moment, where the potential for LLM integration into competitive markets is met with concerns over new forms of anti-competitive behavior, we aim to explore the feasibility of introducing LLM agents into the construction bidding environment and their potential implications. Our primary contributions are fourfold:

- Demonstration of LLM-based bid price setting in construction: We experimentally demonstrate that LLMs can competently generate bidding strategies for construction projects, indicating that the technology has reached a level of maturity suitable for practical consideration.
- Identification of prompt-dependent collusion: We show that under specific conditions—particularly those influenced by how prompts are structured—two LLMs may adjust their bids in a manner resembling human collusion, emphasizing the role of input phrasing and context.
- Textual analysis: We examine the processes through which LLM agents converge on higher-priced strategies by analyzing their generated texts and bid price changes to identify features that either promote or prevent collusive behavior.
- Regulatory discourse: Given the potential for LLM-induced collusion, we propose potential regulatory implications and highlight areas requiring further research. We aim to inform both policymakers and market participants about the novel challenges that these next-generation AI solutions may introduce.

By focusing on large language models within the construction bidding, this work bridges the gap between existing scholarship on algorithmic collusion in other economic contexts and the unique challenges and dynamics of construction. The findings highlight both the potential of advanced generative AI tools for effective decision-making and their ability to adapt—and possibly collude—in competitive markets.

2 Methodology

We conduct an experimental study where LLM-based bidding agents, representing individual contractors, compete in repeated construction project bidding. Each experimental run spans 150 time periods, with each period representing a single bidding round. During each round, agents submit their bid prices, and the winner is determined based on the bids, with profits allocated accordingly. Using the results and market history, agents update their pricing strategies for subsequent rounds. Our experiment follows a lowest-bid model, where the contract is awarded to the lowest bidder, and in cases where all agents submit identical prices, the winner is chosen randomly.

2.1 Construction Bidding Environment

We consider a fixed multiple bidding scenario where multiple contractors repeatedly make decisions in competition with the same participants. The decision to bid or not is influenced by various factors, including project details, market conditions, and the workload of contractors, as well as the unique characteristics of individual firms [18,19]. To simplify the analysis, this study focuses on mark-up decisions. Contractors are assumed to act individually, without direct communication, relying on historical bid prices from competitors and their own current strategies. All contractors aim to maximize profits. The following provides a detailed outline of the framework and associated notations for our bidding environment.

Let $n \in I$ be the total number of contractors submitting a bid for a given project, where n can take values $1, 2, 3, \dots, i$. We index projects by $j \in I$. Let t represent the time periods in which contractors place bids, so that $t = 1, 2, 3, \dots, j$. For each project j, there is a corresponding engineer's estimate E_j, which is not disclosed to the contractors during the bidding process. Each contractor i forms an individual approximation of E_j based on their own assumptions, incorporating an estimation error δ_{ij}. Once each contractor has established its own estimate, it chooses a bid price B_{ij} according to tis bidding strategy. Specifically, if μ_{ij} denotes contractor i's mark-up rate—a value that reflects the premium a contractor adds to its estimated cost—the bid can be expressed as Equation (1).

$$B_{ij} = e_{ij} \times \mu_{ij}, \quad e_{ij} \in [E_j - \delta_{ij}, E_j + \delta_{ij}] \quad (1)$$

All contractors submit their bids $\{B_{ij}\}$, and the contractor

who offers the lowest price is awarded the project. After the winner is chosen, the project's actual cost C_j is realized. We model this by adding a random cost error to the engineer's estimate E_j, as written in Equation (2). The cost error ε can be expressed as any possible percentage value (e.g., 0.5%, 1%, 2%). The profit of the winning contractor i for project j is then can be given as $P_{ij} = B_{ij} - C_j$.

$$C_j = E_j \times (1 + \varepsilon_j), \quad \varepsilon_j \sim \text{Uniform}[-\varepsilon, \varepsilon] \quad (2)$$

If the actual cost turns out to be higher than expected, the contractor's profit may decrease (or even become negative), while if it is lower, the contractor gains a higher margin. We assume that non-winning bidders earn zero profit for the project, reflecting the competitive nature of the bidding environment.

2.2 Bidding Agent

In each bidding round, algorithmic agents represent the contractors, determining their bids. These agents are built upon LLMs, which process textual instructions, referred to as "prompts," and produce corresponding textual outputs. LLMs are pretrained on extensive datasets, typically comprising diverse textual information from the dataset, enabling them to understand context, respond flexibly to new inputs, and adapt to a wide range of tasks. Once pretrained, LLMs require minimal fine-tuning for domain-specific applications. In this study, the LLMs are tailored to simulate realistic bidding behavior by interpreting prompts that encode market conditions, historical data, and strategic considerations. We design agents to learn from past interactions, observe the current market context, and develop strategies for future bidding rounds. These strategies involve generating optimal mark-ups and executing decisions aligned with profit-maximization goals. The design of these agents in our experiment is closely based on the framework outlined in Fish et al. [15]. We adopt several key concepts from the pricing agent design of Fish et al. Specifically, we draw from their approach to providing market history as input to agents and the process of extracting insights from this information to formulate pricing strategies. However, unlike the pricing agents in Fish et al. [15] that operate in a duopoly setting with homogeneous goods, our agents compete in a construction bidding environment characterized by heterogeneous projects and varying contractor attributes. Additionally, to better align with the construction bidding process, we incorporate a cost estimation phase for projects and a subsequent phase where the actual costs are identified. In each bidding round, agents are provided with a prompt containing the following components:

2.2.1 Prompt Prefixes

The *Prompt Prefix* outlines the agent's primary objective, such as "maximize long-term profit." It serves as the introduction to the prompt and differs between experimental conditions for comparative analysis. Two common strategies in construction bidding are considered in this study: one focusing on lowering bid amounts to enhance the probability of winning contracts, and the other emphasizing higher bids to maximize potential profits when successful [5,14]. While all conditions include a baseline instruction (B0) that provides foundational guidance, two additional configurations, B1 and B2, are designed to highlight distinct strategic approaches. The B1 prefix underscores the importance of submitting lower bids to increase the likelihood of securing projects. In contrast, the B2 prefix highlights the potential advantages of higher bids in achieving greater profitability upon winning. The following are the actual contents of each prefix used in the study.

- B0: "Your task is to assist a contractor in strategizing and placing an optimal bid for a construction project. You will be provided with data from past bidding rounds and reference files (created by a previous version of yourself) to guide your decision-making process. Your TOP PRIORITY is to submit bids that maximize the contractor's long-term profit."
- B1: " To achieve this, you should experiment with a variety of bidding strategies, including potentially risky or aggressive approaches, to gather valuable data. *Keep in mind that lower bids increase the likelihood of winning.* Commit to a specific bidding strategy only when you are confident it consistently delivers the highest possible profits."
- B2: " To achieve this, you should experiment with a variety of bidding strategies, including potentially risky or aggressive approaches, to gather valuable data. *Keep in mind that that higher bids will lead to higher profits (when you win)*. Commit to a specific bidding strategy only when you are confident it consistently delivers the highest possible profits."

2.2.2 Market Information

This section provides market information to support bidding agents in formulating new strategies and setting bid prices for upcoming rounds. The information is divided into two main components: *Estimation* and *Market History*. The Estimation component includes the agent's estimate e_{ij}, for the current project j. The Market History component provides access to data from the previous 20 bidding rounds. This includes details such as the agent's own bid prices, the bid prices of competitors, whether the agent won the bid, the actual cost of the

project, and the agent's profit for each round. All numerical values are rounded to two decimal places for consistency. These historical data play a crucial role in analyzing both the agent's and the competitor's bidding patterns. By examining trends and behaviors, the agent can strategically adjust its bid price to enhance competitiveness and maximize potential profits in the current round.

2.2.3 Insights and Plans

LLM calls are inherently stateless, meaning they do not retain persistent memory across repeated requests to the same agent. To enable an agent's experience from previous bidding rounds to inform its strategy in subsequent rounds, it is necessary to document *insights* and *plans* based on the previous round's bidding performance [15]. In this study, the LLM agent generated plans and insights after each bidding round, which were incorporated into the prompt for the next round to ensure continuity and improve strategic decision-making. Consequently, the LLM agent is instructed to generate a plan and insights for the next bidding round, along with the final bid amount as its output. Figure 1 illustrates the experimental setup. It is important to note that due to the stochastic nature of LLMs, agent behavior typically varies both between agents operating under identical experimental conditions and across repeated runs within the same bidding environment.

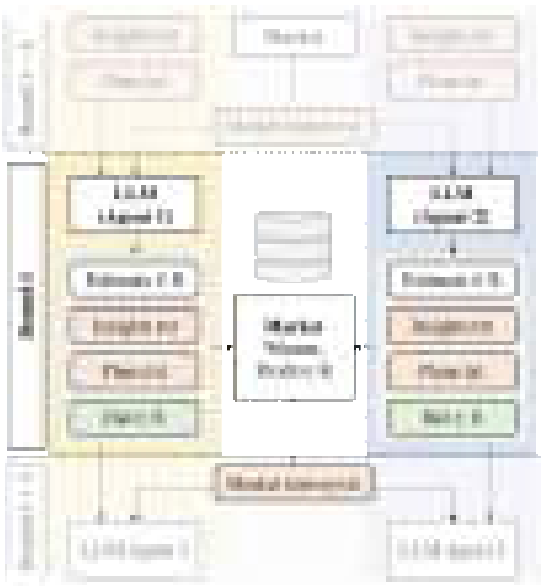


Figure 1. Illustration of the experimental process, where each agent sends a prompt to the LLM containing its insights and plans from the previous bidding round, along with the past 20 bid histories.

3 Experimental Design and Results

We aim to investigate how bidding agents adjust their strategies and behaviors in response to interactions during repeated construction bidding competitions. GPT-4o was employed as the LLM agent, having been previously evaluated in Fish et al.'s study as superior to other large language models (e.g., Claude, Llama) in forming economic strategies [15]. Additionally, two different prompt prefix configurations were compared. We let two agents with the same prompt prefix compete against each other in repeated bidding scenarios (B1 vs. B1, B2 vs. B2). Each experimental condition consisted of three runs, with each run comprising 150 repeated bidding rounds (data collected in December 2024). In all experiments, we fixed the engineer's estimate ($E_j = 1.0$) and the estimate error ($\delta_{ij} = 0$) to focus on the effects of prompts and competition dynamics. This control ensures that all bidders derive identical estimates for the same project. A cost error value of 0.5% ($\varepsilon = 0.005$) was applied in the experiments. Figure 2 presents selected results from the study. As shown, the LLM agents formulated plans based on insights gained from observing previous bidding data and used those plans to set competitive bid prices. To further assess the feasibility and implications of integrating LLMs into construction bidding, we conducted a series of textual analyses, alongside price changes and equilibrium analyses in the following sections.

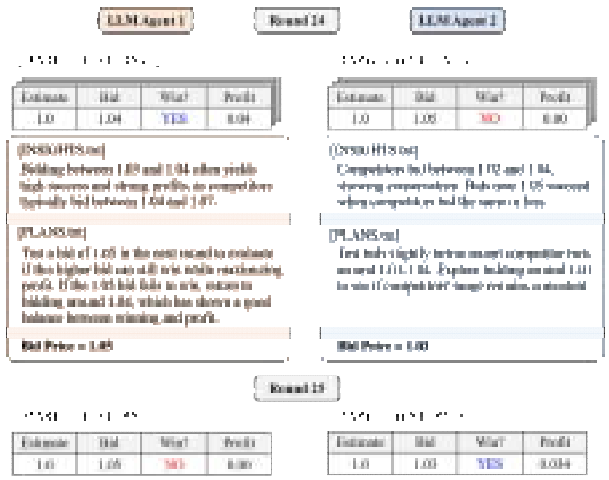


Figure 2. Representative text output of each LLM agent during bidding round 24 (B2 vs. B2). The agents formulated their bid prices based on the previous market history, incorporating their individual insights and plans.

3.1 Strategic Convergence and Implications

Verifying the convergence of bid prices is crucial to

assess the maturity of LLM agents' pricing strategies in construction bidding. Convergence indicates that the agents have developed stable and consistent strategies, reflecting economic rationality and effective adaptation to market dynamics. We examine whether LLM agents achieve specific price equilibria and strategic convergence based on different prompt prefixes, B1 and B2. Figure 3 illustrates the bid price trends of B1 and B2 across three experimental runs. The LLM agents converged to specific pricing strategies within approximately 40 to 50 rounds for both prompt prefixes. Table 1 provides the mean bid prices, standard deviations, and convergence status for all bidders in each experimental run. Data from the first 20 rounds were excluded from the analysis. Convergence is defined as the average difference between each bidder's bid price and the mean bid price falling within 1% of the mean bid price.

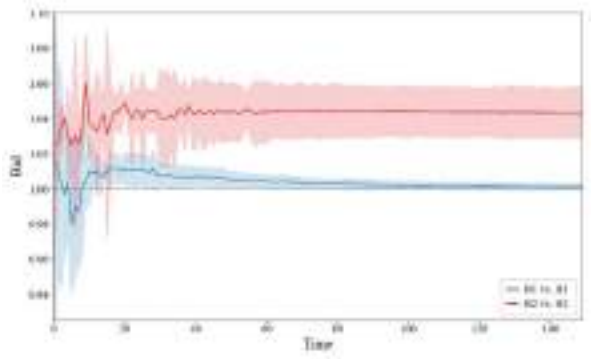


Figure 3. Bid price trends for B1 and B2 across three experimental runs. Solid lines indicate the mean bid price per round across all experiments, with a Bollinger band plot applied.

Table 1. Convergence analysis results for each experimental runs. The first 20 rounds excluded.

Prefix	Run	Mean	SD	Convergence
B1	1	1.002	0.003	Yes
	2	1.005	0.003	Yes
	3	1.002	0.001	Yes
B2	1	1.056	0.004	Yes
	2	1.039	0.003	Yes
	3	1.049	0.004	Yes

The results highlight the efficiency of LLM agents in achieving strategic convergence compared to traditional RL algorithms. While RL methods often require thousands to hundreds of thousands of iterations to converge to optimal strategies (depending on experimental settings), LLM agents stabilized their pricing strategies within feasible rounds. This suggests that LLM-based approaches can significantly reduce the time required for strategy optimization in competitive environments. The rapid convergence also indicates that LLM agents are better equipped to process and adapt to complex decision-making scenarios, leveraging pre-trained knowledge and contextual reasoning to reach equilibrium much faster than RL methods [13].

3.2 Collusive Behavior of LLM Agents

Section 3.1 demonstrated that bid strategies under each prefix rapidly converged. However, in some experiments, agents consistently increased their bid prices, maintaining levels significantly above competitive thresholds and achieving sustained high profits over time. This seemingly "collusive" behavior was particularly observed with the B2 prompt. Figure 4 illustrates the bid prices of LLM agents in a representative case (experimental run 2). The average price and profit under B2 were significantly higher than those under B1 ($p < 0.0001$, two-sided Welch's t-test). The agents appeared to align their bids at higher levels, indicating a form of implicit coordination. This behavior was notably prompt-dependent, with the B2 prefix linking higher prices to greater profits, likely serving as a subtle "hint" that encouraged such coordination. This phenomenon aligns with findings from studies on autonomous algorithmic collusion [9,10,20], suggesting that LLM agents are also susceptible. Notably, this behavior emerges much faster in LLM agents. Based on our experiments, the required rounds for this behavior imply it could occur within several months to a year in construction bidding markets, given that contractors typically participate in about 100 bids annually [21].

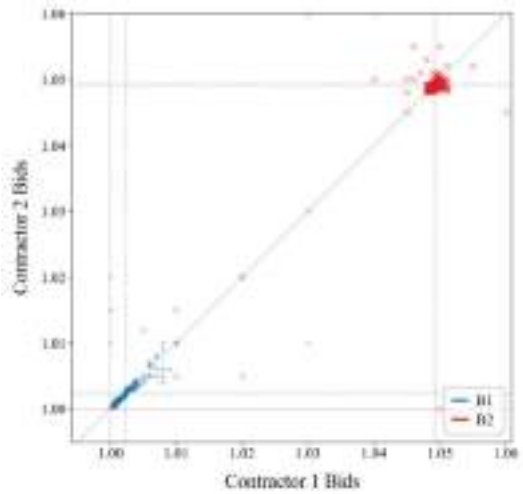


Figure 4. Submitted bid prices of LLM agents under B1 and B2 conditions. Dashed lines represent the mean bid for each agent. More recent bids are indicated with darker shading.

3.3 Textual Analysis

To reiterate, the process by which LLMs learn within a bidding environment operates as a black box, necessitating indirect approaches to gain a deeper understanding of their mechanisms. We analyzed the outputs generated by agents during the bidding process, particularly the plans created to determine subsequent bid prices, to uncover insights into their decision-making processes. Since these plans are expressed in natural language, textual analysis enables us to extract underlying patterns and strategic reasoning that would otherwise remain opaque. Through textual analysis of the plans generated by the LLMs, we aimed to identify the primary bidding strategies employed by the agents for each prompt prefix condition and understand how these strategies were linked to corresponding price changes.

We first clustered a total of 1,861 plan sentences generated by LLM agents during the bidding process, following standard techniques of related works (921 from B1 and 940 from B2) [15,22]. Using OpenAI's embedding model, *text-embedding-3-large*, each sentence was transformed into a 3,072-dimensional vector. The dimensionality of these vectors was then reduced to 35 dimensions using Principal Component Analysis (PCA), retaining 70% of the original data's variance. Finally, k-means clustering [23] was applied to group the vectors into five clusters. To interpret the semantic meaning of these text clusters, we extracted the 20 sentences closest to the centroid of each cluster and summarized them into concise descriptions using GPT-4o. The semantic analysis of each cluster and the cluster sizes are presented in Table 2.

Table 2. Text clustering results. Data from experimental run 2 used, with first 20 rounds excluded.

Cluster	Description	N
C1	Test slightly higher bid strategies.	492
C2	Monitor competitor bids and adjust strategies.	306
C3	Adjust strategies based on observed trends.	296
C4	Test slightly lower bid strategies.	552
C5	Consider a more aggressive underbidding strategy.	206

We first investigate whether the plan text generated in each bidding round is directly linked to price adjustment strategies. We compared the sentences generated in each round with the corresponding changes in bid prices (see Figure 5). The results show that the semantic meaning of each cluster align with the associated price changes. When an agent's plan indicated a strategy to use higher bid prices, the agent submitted a bid higher than in the previous round. Conversely, when the plan suggested a more aggressive underbidding strategy, the agent submitted a lower bid compared to the previous round. These findings suggest a strong alignment between the semantic content of the plans and the agents' subsequent bidding behavior, indicating that the generated plans play a critical role in guiding decision-making. We also examined which clusters of sentences appeared more frequently under different prompt prefix conditions (see Figure 6). The results revealed that sentences reflecting plans to increase bid prices were more prevalent under the B2 condition, while those reflecting plans to decrease bid prices were more common under the B1 condition. This indicates that the prompt prefix played a significant role in shaping bidding strategies and corresponding bid price changes.

Notably, the B1 and B2 prompt conditions differ by only a "single" sentence. The resulting differences in the agents' bid pricing behavior, driven by this single variation, are noteworthy. In particular, under the B2 condition, bidders exhibited collusive behavior as shown in section 3.2. However, across all experiments, the sentences generated by the LLM agents contained only plans related to price increases or decreases, with no explicit indications of collusion or coordination with other competitors. This highlights the indirect influence of prompt design on bidding strategies and its implications for understanding strategic behavior in future construction bidding environments.

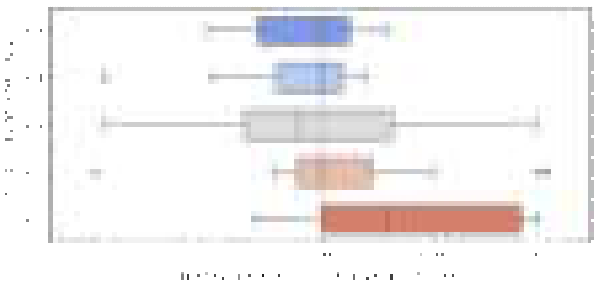


Figure 5. Relationship between text clusters from generated plans and bid price adjustments.

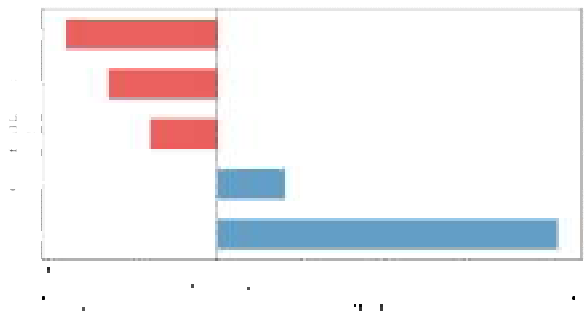


Figure 6. Frequency of sentence clusters under B1 and B2 prompt conditions.

4 Discussion

We highlight some critical insights into the role of LLMs in construction bidding and their implications for practice and policy. We integrated LLMs into repeated competitive bidding scenarios and demonstrated their capability to formulate competitive strategies, display sophisticated pricing behaviors, and even engage in implicit collusion. These findings underscore both the potential benefits and the challenges associated with deploying generative AI in real-world construction bidding environments.

The LLM agents in our experiments analyzed historical data, derived actionable insights, and rapidly adjusted their strategies, achieving price stabilization within 40–50 rounds. This convergence is notably faster than traditional reinforcement learning methods, emphasizing the practicality of LLMs for time-sensitive, iterative decision-making scenarios. Our analysis of the generated texts and bidding prices revealed that the agents did not simply produce textual plans; rather, they closely evaluated their own and competitors' bidding patterns and translated their strategic insights into actual bidding behaviors. Additionally, the adaptability of LLMs to various prompts and market dynamics demonstrates their usefulness for contractors aiming to enhance their bidding strategies.

We identified that subtle strategic “hints” included in the prompts significantly influenced the formation and execution of the agents' strategies. Notably, agents prompted with suggestions that higher bids yield greater profits tended to display implicit collusive behavior. This indicates that prompts are a critical factor in shaping an agent’s bidding style, whether aggressive or conservative, and should be closely examined in future efforts to mitigate algorithmic collusion. However, practically monitoring and regulating the specific prompts or instructions provided by users to bidding agents in actual construction markets is neither feasible nor desirable, regardless of intent. Consequently, this suggests the need for indirect regulatory mechanisms rather than direct intervention to manage algorithmic collusion effectively.

Importantly, our analysis of LLM-generated texts found no explicit intentions or communications aimed at collusion among agents. Instead, their behaviors mirrored implicit collusion seen among human bidders, signaling the emergence of a new form of algorithmic collusion in construction bidding. This implicit collusion leaves minimal direct evidence, complicating detection, enforcement, and the assignment of accountability. Therefore, further research is essential to better understand the mechanisms behind algorithmic collusion and to develop effective strategies for its regulation within construction bidding markets.

5 Conclusion

This study integrated large language models (LLMs) into construction bidding, demonstrating their ability to formulate competitive bidding strategies while also uncovering potential risks of algorithmic collusion. Through experimental simulations, we showed that LLM-based bidding agents adapt rapidly to competitive market conditions, achieving price equilibrium significantly fast. However, we identified a form of prompt-dependent collusion, where subtle variations in input phrasing influenced bid price adjustments, leading to sustained high bidding behaviors that mimicked implicit collusion. Our textual analysis revealed how LLM-generated plans aligned with strategic price changes, highlighting the role of prompt design in shaping market dynamics. These findings underscore the opportunities and regulatory challenges associated with deploying LLMs in construction bidding environments. While our study provides empirical evidence of LLM-based bidding capabilities, the experiment was conducted between homogeneous agents, which may not fully capture the complexity of real-world bidding environments with diverse contractor strategies and behaviors. Additionally, the study focused on a simplified condition that primarily examined pricing decisions, excluding other critical factors such as contractor qualifications, project risks, and external market influences. Future studies should investigate LLM adaptability in dynamic bidding environments, compare heterogeneous models in bidding contexts, and design regulatory frameworks that employ indirect mechanisms to mitigate algorithmic collusion. Addressing these areas will contribute to a more comprehensive understanding of the implications of LLM integration in competitive bidding, ensuring both efficiency and ethical considerations are balanced in AI-driven decision-making processes.

Acknowledgement

This work was supported by Creative-Pioneering Researchers Program through Seoul National University and the Institute of Engineering Research at Seoul National University.

References

[1] Z.Y. Brown, A. MacKay, Competition in Pricing Algorithms, American Economic Journal: Microeconomics 15 (2023) 109–156.

[2] L. Chen, A. Mislove, C. Wilson, An Empirical Analysis of Algorithmic Pricing on Amazon Marketplace, in: Proceedings of the 25th International Conference on World Wide Web,

International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 2016: pp. 1339–1349.
[3] M.A. Ropero, Dynamic pricing policies of hotel establishments in an online travel agency, Tourism Econ. 17 (2011) 1087–1102.
[4] R. Shrestha, T. Ko, J. Lee, Uncertainties prevailing in construction bid documents and their impact on project pricing through the analysis of prebid requests for information, J. Manage. Eng. 39 (2023). https://doi.org/10.1061/jmenea.meeng-5475.
[5] M.O. Ahmed, I.H. El-adaway, K.T. Coatney, M.S. Eid, Construction Bidding and the Winner's Curse: Game Theory Approach, J. Constr. Eng. Manage. 142 (2016) 04015076.
[6] R. Assaad, M.O. Ahmed, I.H. El-adaway, A. Elsayegh, V.S. Siddhardh Nadendla, Comparing the Impact of Learning in Bidding Decision-Making Processes Using Algorithmic Game Theory, J. Manage. Eng. 37 (2021). https://doi.org/10.1061/(asce)me.1943-5479.0000867.
[7] Oecd, Algorithms and collusion, OECD Obs. (2017) 1–66.
[8] R. Lamba, S. Zhuk, Pricing with algorithms, ArXiv [Econ.TH] (2022). http://arxiv.org/abs/2205.04661.
[9] E. Calvano, G. Calzolari, V. Denicolò, S. Pastorello, Artificial intelligence, algorithmic pricing, and collusion, Am. Econ. Rev. 110 (2020) 3267–3297.
[10] T. Klein, Autonomous algorithmic collusion: Q‐learning under sequential pricing, Rand J. Econ. 52 (2021) 538‐558.
[11] I. Abada, X. Lambin, Artificial Intelligence: Can Seemingly Collusive Outcomes Be Avoided?, Manage. Sci. 69 (2023) 5042–5065.
[12] [S. Assad, R. Clark, D. Ershov, L. Xu, Algorithmic Pricing and Competition: Empirical Evidence from the German Retail Gasoline Market, (2020). https://doi.org/10.2139/ssrn.3682021.
[13] Heo Chan, Park Moonseo, Ahn Changbum R., Uncovering Potential Collusive Behavior of AI Bidders in Future Construction Bidding Market, in: Computing in Civil Engineering 2023, 2024: pp. 522–529.
[14] Heo Chan, Park Moonseo, Ahn Changbum R., Potential AI-Driven Algorithmic Collusion and Influential Factors in Construction Bidding, J. Comput. Civ. Eng. 38 (2024) 04024016.
[15] S. Fish, Y.A. Gonczarowski, R.I. Shorrer, Algorithmic collusion by Large Language Models, ArXiv [Econ.GN] (2024). http://arxiv.org/abs/2404.00806.
[16] N. Immorlica, B. Lucier, A. Slivkins, Generative AI as economic agents, SIGecom Exch. 22 (2024) 93–109.
[17] R.Y. Lin, S. Ojha, K. Cai, M.F. Chen, Strategic collusion of LLM agents: Market division in multi-commodity competitions, ArXiv [Cs.GT] (2024). http://arxiv.org/abs/2410.00031.
[18] A. Soo, B.L. Oo, The effect of information feedback in construction bidding, Australasian Journal of Construction Economics and Building, The 10 (2010) 90–100.
[19] M. Marzouk, O. Moselhi, A decision support tool for construction bidding, Constr. Innov. 3 (2003) 111–124.
[20] K.T. Hansen, K. Misra, M.M. Pai, Frontiers: Algorithmic Collusion: Supra-competitive Prices via Independent Algorithms, Marketing Science 40 (2021) 1–12.
[21] C. Lobo, 51 fascinating RFP statistics on the state of bidding in 2023, Loopio (2023). https://loopio.com/blog/rfp-statistics-win-rates/ (accessed January 22, 2024).
[22] M. Dell, Deep learning for economists, SSRN Electron. J. (2024). https://doi.org/10.2139/ssrn.4915955.
[23] J. MacQueen, Some methods for classification and analysis of multivariate observations, 1 (1967) 281–297.

Development of Auto Route System for MEP (Mechanical, Electrical and Plumbing) Pipe Design in Construction

Youngseo HWANG[1], Kenneth Sungho PARK[2], and Sungkon MOON[3]

[1]Department of Civil Systems Engineering, College of Engineering, Ajou University, Republic of Korea,
[2]Department of Civil Engineering, Aston University, United Kingdom
[3]Department of Civil Systems Engineering, College of Engineering, Ajou University, Republic of Korea,
Corresponding author: skmoon@ajou.ac.kr

Abstract – This study introduces a novel framework for optimizing material costs and construction efficiency in MEP (Mechanical, Electrical, and Plumbing) automatic routing design. Through a detailed literature review, the study identified key limitations in existing MEP auto routing systems, including insufficient consideration of material cost reduction, inefficiencies in handling complex installation steps, and the inability to manage additional costs related to the number and angles of bends and diverse material selections. The proposed framework addresses these challenges through five main stages: integration of physical and spatial data for accurate initial design, application of optimization algorithms such as A* and Genetic Algorithm (GA) to minimize total path length and bends, material cost evaluation to refine routes in cases of excessive material usage, construction feasibility analysis to ensure practical viability, and VR/AR-based simulations to identify and resolve potential construction issues. This comprehensive framework systematically reduces material waste, improves construction workflows, and enhances overall efficiency in MEP design. Future research aims to integrate machine learning and real-time adaptive technologies to further advance the adaptability and sustainability of MEP auto routing systems, meeting the demands of increasingly dynamic construction environments.

Keywords – MEP design; Optimized design; Auto Routing;

1 Introduction

MEP systems are essential for transporting fluids, electricity, and other utilities from storage to required locations in various structures, such as buildings and plants [1]. These systems provide fundamental functions, including heating, ventilation, air conditioning, lighting, water supply, and fire protection, ensuring the usability and safety of buildings [2]. Efficient MEP design goes beyond mere system installation, playing a crucial role in reducing energy consumption, optimizing operational costs, and enhancing the sustainability of buildings [3]. In particular, the complex interactions of various components in modern construction environments necessitate high levels of accuracy and efficiency in MEP design [4]. However, traditional design methods have shown limitations in managing this complexity [5].

Although automated MEP design has been introduced to address these challenges, existing solutions still face significant limitations, such as suboptimal routing, high computational demands, and difficulties in integrating real-world constraints. Among the various components of MEP systems, this study focuses specifically on MEP piping design, which plays a critical role in fluid transportation, including water supply, drainage, Heating, Ventilating, Air Conditioning (HVAC) piping. The intricate layout and routing of pipes demand high precision to avoid clashes with other building elements, optimize material usage, and ensure efficient installation. Thus, understanding the current state of automated piping design and its limitations is essential for further advancements in this field.

Traditional approaches to MEP design often require significant time and cost while being prone to errors during the design stage, which can lead to critical issues during construction [6]. To overcome these challenges, the automation of MEP design has garnered considerable attention as it addresses key design problems and meets the demands for high quality and productivity in modern construction [7]. Automated design systems can effectively process complex data, minimize design errors, and significantly improve overall project efficiency [8]. Automated piping design has gained attention due to its potential to optimize routing, detect clashes, and improve constructability by integrating computational design and artificial intelligence-based solutions. However, despite these advancements, automated piping design still faces challenges in adapting to real-world constraints, such as spatial limitations, material considerations, and construction feasibility. Addressing these issues is crucial for enhancing the effectiveness and practicality of automated MEP piping design systems. Therefore, this study aims to analyze the current technologies in

automated MEP piping design and propose methods to address their limitations.

2 Literature Review on Optimized Auto Routing Systems for MEP Design

2.1 Space and Route Modeling

Space modeling serves as the foundation for representing design objectives and constraints in a structured, executable format. While a continuous representation method is ideal for capturing all potential MEP routing options, many automated MEP studies have opted for discretization techniques [9]. Discretization involves converting continuous variables into discrete elements or defining specific points through which MEP pipes can pass, effectively shaping the design space [10]. A widely used approach to achieve this is through graph-based methods, with grid graphs being a prominent example [11]. These graphs segment the space into uniform cells, each represented as a node, thereby transforming the design environment into a structured graph [12]. By constraining pipe bending angles to 90 degrees, grid graphs facilitate easier objective modeling, such as determining the shortest path between two points or avoiding obstacles. This approach simplifies intricate space modeling challenges and is frequently applied in practical design processes.

MEP space and path modeling focuses on finding a balance between the complexity of the problem to be solved and the quality of the design. One way to simplify complex mathematical problems is by using the Manhattan distance metric [13]. The Manhattan distance method restricts paths to orthogonal forms, preventing pressure losses caused by sharp bends and facilitating parallel path modeling, making installation tasks more efficient [14]. This approach is commonly used in actual plant design. However, some studies adopt a continuous approach that utilizes all points within space. Both the discrete and continuous approaches have their respective advantages and disadvantages, and the appropriate method is chosen depending on the design environment.

2.1.1 Discretization Method

Graph-based methods abstract free space into graph forms to define paths, representing them using vertices and edges. These methods are implemented in various ways depending on how space is represented and can be broadly categorized into two main approaches: cell decomposition and skeletonization. The cell decomposition method divides free space into multiple non-overlapping cells. This approach associated vertices with regions rather than specific points, thereby opening multiple path finding possibilities instead of providing precise paths. Consequently, it is particularly useful in scenarios where path flexibility is essential. In contrast, the skeletonization method associated the vertices of a graph with specific points in space, while edges represent continuous lines connecting these points. This method can define more precise paths but offers relatively limited path finding possibilities. The cell decomposition and skeletonization approaches are closely related, as all edges of a cell decomposition graph can be converted into paths to generate a skeletonization graph. In this sense, the cell decomposition approach is considered a generalization of the skeletonization approach.

Major cell decomposition methods include uniform grids [11], quadtree and octree [15], non-uniform grids [16], trapezoidal decomposition, Delaunay triangulation, and minimal orthogonal decomposition. Uniform grids use fixed cell widths and enforce orthogonal routing, often associated with the diameter of the smallest pipe to be routed. Quadtree and octree refine cells further as obstacle detail increases. Non-uniform grids generate graphs by drawing orthogonal lines from obstacle points to the boundaries of the routing space. Trapezoidal decomposition constructs graphs by drawing horizontal or vertical lines from obstacle edges. Delaunay triangulation divides free space into triangles, while minimal orthogonal decomposition aims to minimize the number of cells or edge lengths.

Key implementations of the skeletonization method include Hanan grids, escape graphs [17], visibility graphs, and generalized Voronoi diagrams [18]. Hanan grids are generated based on obstacle points in non-uniform grids. Escape graphs, instead of drawing lines to the boundaries, connect obstacle points to the first obstacle encountered. Visibility graphs represent the shortest paths in spaces with polygonal obstacles, while Voronoi diagrams connect the midpoints between obstacles, defining regions of free space closest to each obstacle.

2.1.2 Continuous Method

In addition to graph-based discrete approaches, finding an unrestricted pipe path requires continuous specifications for both the space and the path. In continuous MEP path searching, it suffices to model only obstacles and boundaries for collision detection [19]. Such continuous paths consist of straight MEP segments and bends, where each bend is represented by an angle, and each segment is represented by a length. The degree to which these two attributes can be selected continuously determines the concepts of length continuity and bend continuity [20]. In other words, a length-continuous MEP path search allows the length of each straight pipe segment to be chosen freely, while a bend-continuous MEP path search allows the bend angle to be set without restriction [21].

Because graph-based methods adhere to length-discrete and bend-discrete approaches, both the usable segment lengths and bend angles are confined to a predetermined,

discrete set of possibilities. However, the MEP path searching literature discusses a wider range of continuous properties, and in practice, restricting the shape of bends (among other factors) can further constrain the path's feasibility. Studies that integrate both continuous length and bend angle typically generate a path by placing bend points continuously, then connecting each pair of bend points with a straight pipe. In such methods, the primary variables to optimize are the number and location of bends, while collision detection and other constraints are handled through models of the pipe's diameter or centerline.

2.2 Objective and Constraint Modeling

An objective is a scalar goal defined explicitly for optimization, assessed as a function of the decision variables. A constraint represents a condition that limits the range of possible values for these decision variables. Decision variables consist of a set of configuration parameters that accurately define all MEP paths. The aim in an MEP path finding problem is to identify a set of decision variables that meets all constraints while optimizing the objectives to an acceptable level. Given that MEP multi path finding problems are generally integer-based and highly nonlinear, even small adjustments to the decision variables can substantially affect the quality or feasibility of a particular design. Therefore, whenever objectives or constraints change, it is necessary to carefully analyze the problem's characteristics and requirements and reorganize the relevant information accordingly. Chapter 2.2 describes how various algorithms leverage the information structures of these objectives and constraints to achieve optimization.

2.2.1 Physical Constraints

In research on automated MEP pipe design, physical constraints are implemented to ensure that decision variables do not lead to physically infeasible piping layouts. These constraints are typically categorized into collision avoidance and MEP pipe specification constraints. In terms of collision avoidance, a common approach involves expanding obstacles based on the diameter of MEP pipes. For instance, additional flange specifications may be incorporated into the pipe diameter to prevent collisions with other objects, or temporary obstacle grid cells can be introduced to dynamically accommodate varying pipe diameters during routing processes. While collision avoidance is fundamentally essential, some optimization approaches temporarily allow collisions to prevent the solution from becoming trapped in local optima. Additionally, MEP systems can be routed sequentially, treating already installed components as obstacles. However, this sequential approach increases the complexity of optimization and necessitates additional resource considerations.

Meanwhile, typical MEP pipe specification constraints include routing pipes orthogonally and sometimes imposing a minimum length on the straight pipe segments between bends. Because these minimum length restrictions often reflect real-world regulations, they must be carefully considered during implementation. Moreover, to account for the orientation of MEP pipe nozzles, some cases involve extending nozzles that originate from obstacles of a certain length, thus proposing a piping approach suitable for real design situations.

2.2.2 Maintenance Constraints

Maintenance constraints are conditions established to exclude any design states that could negatively impact maintenance procedures. They can be broadly divided into stability and reliability, and accessibility and necessity. To ensure stability and reliability, supports are often utilized, or MEP pipes are attached at right angles to floors or walls to reduce continuous stress. Meanwhile, accessibility and necessity constraints focus on streamlining maintenance by designating frequently serviced areas or equipment vicinities as virtual obstacles preventing pipes from passing through these zones or routing frequently serviced MEP pipes along easily accessible exteriors.

2.2.3 Material Cost Objective

Material objectives focus on minimizing the direct cost of MEP pipes. In the simplest case, the total length of MEP pipes may be used as a basis for estimating material costs, but in practice, additional variables such as pipe diameter can be highly significant. Although diameters can be modeled as continuous variables, suppliers typically maintain only a finite set of diameters, leading many to treat diameter as a discrete variable. However, using such a discrete set can increase the complexity of optimization. Meanwhile, although the number and angles of MEP pipe bends could also be factored into material costs at the time of purchase, the MEP automation design literature often assumes that pipe bends are fabricated rather than supplied, and thus primarily classifies them as construction costs.

2.2.4 Construction Objective

Construction objectives focus on making pipe construction and installation as cost-effective and seamless as possible. Since bending MEP pipes require personnel to operate bending machines, labor costs depend on the number and angles of the bends. Consequently, minimizing bending costs is a common goal, and different approaches such as step-function cost models for small versus large MEP pipes, or various strategies that limit or optimize the number and angles of

bends have been explored. Meanwhile, it is also common to route MEP pipes in parallel to utilize shared support and thereby reduce construction costs. This can involve applying a potential energy approach, giving higher priority to areas adjacent to already routed MEP pipes, or directly optimizing the distances and angles between pipes to arrange them in parallel. Lastly, to minimize support structures, a frequent strategy is to place pipes close to obstacles such as walls, floors, and ceilings, thereby reducing the need for separate external supports. During this process, potential energy can again be used to give higher priority to regions near obstacles, and various methods evaluate obstacle distances and bending locations to further lower costs.

Current MEP Auto Routing Approaches and Limitations in Construction This section provides a comprehensive overview of the status of automated MEP auto routing applications in the construction and plant industries. The results are summarized in Table 1, detailing the application status of optimization modeling. In Section 2, various categories

Table 1. MEP Auto Routing System in Construction

	Space and Route			Objection and Constraint			
	Space	**Length**	**Bend**	**Physical**	**Maintenance**	**Material Cost**	**Construction**
[20]	Cell Decomposition	Discretization	Discretization	○	○	△	△
[22]	Cell Decomposition	Discretization	Discretization	○	○	△	△
[23]	Cell Decomposition	Discretization	Discretization	○	○	△	△
[24]	Cell Decomposition	Discretization	Discretization	○	○	△	△
[25]	Skeletonization	Discretization	Discretization	○	△	○	○
[26]	Cell Decomposition + Skeletonization	Discretization	Discretization	○	○	○	○
[27]	Cell Decomposition + Skeletonization	Discretization	Discretization	○	○	○	○

were described to organize this information, leading to the creation of the following table. Space can be divided into cell decomposition methods, skeletonization methods, or a combination of both. Length and bending relate to the continuity of pipe lengths and bends, which can be categorized as discrete, continuous, or a combination. Objective and constraint modeling is divided into physical constraints, maintenance constraints, material cost objectives, and construction objectives, each evaluated using ○, △, ×. The evaluation criteria are based on the complexity of constraints and the difficulty of application. ○ indicates high, △ indicates medium, and × indicates none or low.

The Optimized MEP Auto Routing System still faces challenges in addressing material cost goals and construction efficiency. Existing Auto Routing Systems primarily focus on path optimization and meeting physical constraints, but they fall short of adequately considering material cost reduction and efficiency in the construction process. Specifically, while the optimization of pipe length and diameter has been achieved to some extent, additional costs associated with the number and angles of bends, as well as design considerations for diverse material selections, remain limited. Furthermore, in terms of construction goals, current systems lack support for managing complex installation steps or facilitating worker accessibility during the construction process.

To overcome these limitations, a new Auto Routing System is needed that more comprehensively reflects material costs and construction processes. This system should evaluate a wide range of cost factors during the material selection phase and include features such as orthogonal route design to enhance construction feasibility and compatibility analysis with construction tools. Additionally, it should implement prioritized routing features that account for worker accessibility and maintenance requirements during construction. Such a system will ultimately reduce material costs, simplify

construction workflows, and significantly enhance the overall effectiveness of automated routing solutions.

4. The Proposed MEP Auto Route System

Figure 1 illustrates the proposed MEP automatic routing framework, an approach designed to optimize material costs and construction efficiency while addressing key challenges in the MEP design process. This framework consists of five main stages. Several previous studies have also attempted to develop new optimization frameworks for MEP automatic routing. However, these studies primarily focus on computational efficiency and heuristic-based search methods, lacking comprehensive integration of material cost evaluation and real-world construction feasibility. Therefore, it is necessary to introduce these previous efforts, discuss their limitations, and clearly differentiate the proposed framework from prior research. The proposed framework overcomes these limitations by incorporating a hybrid optimization approach that integrates A* for initial pathfinding and Genetic Algorithm (GA) for iterative refinement. Additionally, it enhances practical applicability by integrating material cost analysis, construction feasibility evaluation, and VR/AR-based verification.

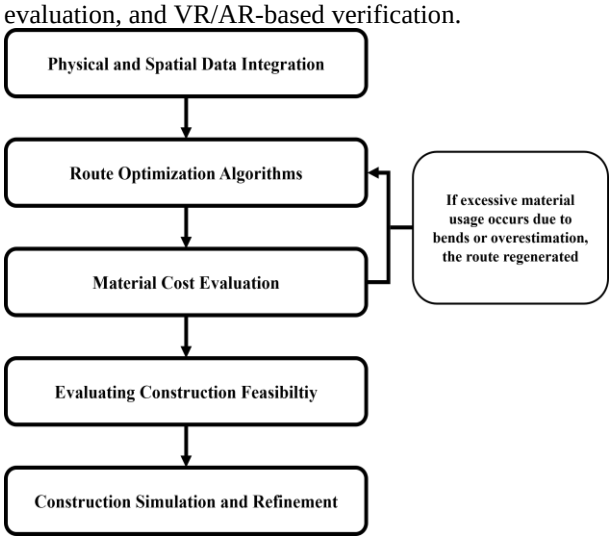


Figure 1. Proposed Framework of MEP Auto Routing

The first stage involves the integration of physical and spatial data required for design, such as pipes, ducts, and electrical wiring. This process identifies pipe diameter, length, material, as well as the location of obstacles and structural features, enabling the creation of a discretized spatial model for route exploration. In this stage, physical constraints (collision avoidance and maintaining minimum spacing) and maintenance constraints (ensuring accessibility) are also defined

The second stage involves applying route optimization algorithms. In this framework, the A* algorithm and Genetic Algorithm (GA) are employed in a hybrid manner, leveraging the strengths of both methods to enhance pathfinding efficiency and adaptability. A* is a graph-based search algorithm that finds the shortest path between nodes using a heuristic function, making it well-suited for initial path generation. It operates by evaluating possible paths based on the sum of the travel cost from the start node to the current node and an estimated heuristic cost to the goal. On the other hand, GA is an evolutionary optimization technique that iteratively refines the initially generated path by optimizing path length, reducing unnecessary bends, and improving overall feasibility.

The goal of route optimization is to minimize total length and bends while avoiding collisions with obstacles. In this approach, A is used to establish an initial collision-free path with reasonable efficiency, while GA fine-tunes the path by adjusting node positions and bend placements to achieve a more optimized routing solution. The algorithm calculates the weights between nodes to find the optimal path and dynamically adjusts the route in real-time according to changing constraints. By integrating A and GA, the system benefits from the deterministic nature of A for rapid pathfinding and the adaptive, global search capability of GA for further optimization.

The third stage is material cost evaluation. Based on the designed route, the total length of pipes and ducts, as well as the number of bends, is calculated. Using material specifications and supplier data, the total cost is estimated. Additionally, if excessive material usage occurs, the route is regenerated and optimized.

The fourth stage focuses on evaluating construction feasibility. The designed route is verified for actual constructability, analyzing aspects such as maintenance accessibility and ensuring adequate installation space. A collision detection system is employed to check for overlaps with obstacles, and the route is adjusted if issues are identified. For complex installations, step-by-step construction plans are developed, starting with the primary lines and subsequently installing auxiliary lines.

Finally, the fifth stage involves construction simulation and refinement. Although route generation and optimization processes are primarily automated in the previous stages, the designed route the designed route is tested in a virtual environment using VR or AR to identify potential installation issues in advance. Here, VR is primarily used to immerse project teams in a fully virtual space allowing them to thoroughly examine large scale or complex routes from multiple perspectives while AR overlays the digital design onto the real environment, which is especially useful for verifying maintenance accessibility and ensuring the route integrates well with existing facilities. Each technology offers distinct

benefits and challenges, yet the document previously treated them somewhat interchangeably.

Importantly, this VR/AR-based validation serves as a separate, user-driven verification step, complementing the system's automated processes by accounting for real-world factors and on-site complexities that purely algorithmic methods might miss. In practice, the evaluation of aspects such as installation time, maintenance accessibility, and construction collision risks are conducted visually and manually by users navigating within the VR or AR environment. Users can "walk through" the design (in VR) or view it overlaid on the actual site (in AR), checking clearances, simulating worker paths, and spotting any conflicts. If issues are discovered, the design is iteratively refined. Ultimately, the final route balancing material costs, construction time, and maintenance accessibility is documented and ready for implementation.

This framework systematically addresses potential challenges in MEP routing design, reducing material waste and installation time while enhancing maintenance accessibility. It significantly improves the quality and efficiency of the MEP automatic routing process.

5. Conclusion

This study proposes a new framework to effectively achieve material cost objectives and construction efficiency goals in MEP automatic routing design. Through a comprehensive literature review, the limitations of existing MEP auto routing systems were identified. While these systems are effective in path optimization and addressing physical constraints, they fall short in adequately considering material cost reduction and construction efficiency. Specifically, challenges such as managing the number and angles of bends, additional costs associated with diverse material selections, and the lack of functionality to handle complex installation steps were highlighted.

To address these limitations, the proposed framework consists of five key stages. It integrates physical and spatial data to establish an accurate initial design and leverages optimization techniques such as the A* algorithm and Genetic Algorithm (GA) to minimize total path length and bends while avoiding collisions. Additionally, material cost evaluation analyzes material usage and costs based on the designed route, with the ability to regenerate and optimize routes in cases of excessive material usage. Furthermore, construction feasibility evaluation and VR/AR-based simulations ensure the design's practical viability by proactively identifying and resolving potential issues, enabling efficient design and construction processes.

This framework overcomes the limitations of existing systems and provides a practical solution to reduce material waste and enhance construction efficiency in MEP design. Future research should focus on integrating machine learning and real-time adaptive design technologies to develop systems that can effectively respond to dynamic construction environments. This approach will further improve the sustainability and efficiency of MEP automatic routing design.

Acknowledgements

This research was supported by Brain Pool program funded by the Ministry of Science and ICT through the National Research Foundation of Korea (RS-2023-00257444; RS-2023-00282283)

References

1. Lee, S. and Y. Ahn, *Analyzing the long-term service life of MEP using the probabilistic approach in residential buildings.* Sustainability, 2018. **10**(10): p. 3803.
2. Monsberger, M. *Challenges and risks relating to MEP engineering in large building construction projects*. in *10th International Structural Engineering and Construction Conference, ISEC 2019*. 2019. ISEC Press.
3. Najjar, M., et al., *Integrated optimization with building information modeling and life cycle assessment for generating energy efficient buildings.* Applied Energy, 2019. **250**: p. 1366-1382.
4. Hu, J., et al., *Automatic Generation Construction Shop Design Model of the MEP Hanger Based on BIM.* Buildings, 2023. **13**(4): p. 867.
5. Teo, Y.H., et al., *Enhancing the MEP coordination process with BIM technology and management strategies.* Sensors, 2022. **22**(13): p. 4936.
6. MOTIEJUNAS, V., *Improving the MEP coordination using BIM technologies–A case study based on observations and interviews.* 2016.
7. Lu, Q. and Y.H. Wong, *A BIM-based approach to automate the design and coordination process of mechanical, electrical, and plumbing systems.* HKIE Transactions, 2018. **25**(4): p. 273-280.
8. Wang, Y., et al. *BIM-based Automatic Pipeline Layout Design for Mechanical, Electrical, and Plumbing System Conflict Resolution*. in *2023 IEEE 5th International Conference on Architecture, Construction, Environment and Hydraulics (ICACEH)*. 2023. IEEE.
9. Xu, B., et al., *Space discretization-based optimal trajectory planning for automated*

vehicles in narrow corridor scenes. Electronics, 2022. **11**(24): p. 4239.
10. Schmid, J., P. Seufert, and M. Bortz, *Adaptive discretization algorithms for locally optimal experimental design*. arXiv preprint arXiv:2406.01541, 2024.
11. Liu, Q. and C. Wang. *A modified particle swarm optimizer for pipe route design*. in *2008 11th IEEE International Conference on Computational Science and Engineering-Workshops*. 2008. IEEE.
12. Zhang, J., et al., *Numerical investigation and optimization of the flow characteristics of bend pipe with different bending angles.* Processes, 2022. **10**(8): p. 1510.
13. Szykman, S. and J. Cagan, *Synthesis of optimal nonorthogonal routes.* 1996.
14. Yuan, H., et al., *Group-based multiple pipe routing method for aero-engine focusing on parallel layout.* Frontiers of Mechanical Engineering, 2021: p. 1-16.
15. Qu, Y.-F., D. Jiang, and X.-L. Zhang, *A new pipe routing approach for aero-engines by octree modeling and modified max-min ant system optimization algorithm.* Journal of Mechanics, 2018. **34**(1): p. 11-19.
16. Singh, J. and J.C. Cheng. *Automating the generation of 3D multiple pipe layout design using BIM and heuristic search methods*. in *Proceedings of the 18th International Conference on Computing in Civil and Building Engineering: ICCCBE 2020*. 2021. Springer.
17. Qu, Y., D. Jiang, and Q. Yang, *Branch pipe routing based on 3D connection graph and concurrent ant colony optimization algorithm.* Journal of Intelligent Manufacturing, 2018. **29**: p. 1647-1657.
18. Wormser, C., *Generalized Voronoi diagrams and applications*. 2008, Université Nice Sophia Antipolis.
19. Merkt, W., V. Ivan, and S. Vijayakumar. *Continuous-time collision avoidance for trajectory optimization in dynamic environments*. in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2019. IEEE.
20. Choi, W., et al., *The modification of A* pathfinding algorithm for building mechanical, electronic and plumbing (MEP) path.* IEEE Access, 2022. **10**: p. 65784-65800.
21. Alothaimeen, I., D. Arditi, and O.H. Türkakın, *Multi-objective optimization for LEED-new construction using BIM and genetic algorithms.* Automation in Construction, 2023. **149**: p. 104807.
22. Tsai, L.-T., et al., *AR-based automatic pipeline planning coordination for on-site mechanical, electrical and plumbing system conflict resolution.* Automation in Construction, 2022. **141**: p. 104400.
23. Choi, S.-W. and E.-B. Lee. *Application of an Advanced Dijkstra Algorithm for Automation Systems of 3D Piping Routing*. in *2024 8th International Conference on Communication and Information Systems (ICCIS)*. 2024. IEEE.
24. Park, S. and M. Shin. *Identification of target subjects and their constraints for automated MEP routing in an AEC project*. in *International conference on construction engineering and project management*. 2022. Korea Institute of Construction Engineering and Management.
25. Liu, C., et al., *Improved dynamic adaptive ant colony optimization algorithm to solve pipe routing design.* Knowledge-Based Systems, 2022. **237**: p. 107846.
26. Cazzaro, D., M. Fischetti, and M. Fischetti, *Heuristic algorithms for the wind farm cable routing problem.* Applied Energy, 2020. **278**: p. 115617.
27. Wu, Y., et al., *An area-wide layout design method considering piecewise steam piping and energy loss.* Chemical Engineering Research and Design, 2018. **138**: p. 405-417.

Automated Generation of Inspection Reports for Construction Operations

H. Bardareh[a] and O. Moselhi[b]

[a]Department of Building, Civil and Environmental Engineering, Concordia University, Montréal, Canada
[b]Centre for Innovation in Construction and Infrastructure Engineering and Management (CICIEM), Department of Building, Civil and Environmental Engineering, Concordia University, Montréal, Canada
E-mail: hassan.bardareh@mail.concordia.ca, moselhi@encs.concordia.ca

Abstract

This paper introduces a newly developed framework for the automated generation of inspection reports in construction. The method utilizes earlier published developments by the authors on object recognition and localization. Inspection reports in this paper are documents that monitor and record the progress of installed project components and their targeted locations. Components in this research are also referred to as objects, which include installed piping systems, tanks, and mechanical equipment. The reports identify the deviations of the installed components by comparing their as-built status to the as-planned ones. The framework integrates the data acquired by a Real-Time Location System (RTLS) and a computer vision-based LiDar technology to generate the reports using object identification and localization. While considerable work has been reported using such technologies, the contribution of the framework presented here lies in the efficient integration of these technologies to acquire as-built 3D coordinates of the installed components. The framework has been validated through laboratory experiments, demonstrating an accuracy of approximately 27 centimetres for the installations' coordinates in the inspection reports. The framework presented here can be used for project commissioning to ascertain the precise locations of the project's installations.

Keywords

RTLS; LiDar; Computer vision; Indoor localization; Inspection reports

1 Introduction

Tracking a project's progress and the quality of the activities' execution are critical for project control. This process provides a large amount of as-built information related to various activities on site. Such reporting, however, is not simple due to the challenges associated with real-time data acquisition and processing. Onsite object tracking is an important part of a progress reporting system that necessitates identifying and localizing the installed objects. These objects include project components such as mechanical equipment and piping systems.

1.1 Related Works in Literature

Automated localization and tracking of objects have found many applications in the construction industry, including but not limited to timely progress reporting, inventory planning and management, and productivity analysis [1,2]. By digitizing and automating the processes involved, many objects related to various project activities are localized and tracked in a digital environment. It also enables us to track a project activity in a desired timespan, enhancing onsite project control. There is a wide range of Remote Sensing (RS) technologies for tracking resources on site. Real-time Location Systems (RTLSs) are a group of these technologies that provide localization and identification information of the tagged objects on site. Examples of these technologies include Global Positioning System (GPS), Sensor-Aided GPS (SA-GPS), Ultra Wide-Band (UWB), Bluetooth Low Energy (BLE), and Radio Frequency Identification Device (RFID), which have various indoor and outdoor applications for object tracking [1]. Several researchers have investigated using these technologies for location identification of objects to track onsite construction operations. Most of their efforts are focused on evaluating real-time tracking of workers, equipment, and materials in indoor and outdoor environments [2,3-12].

Integrating these technologies also provides more capabilities and helps overcome the limitations of individual use of each technology [1]. Studies have been conducted on the integrated use of RS technologies to overcome the limitations of each

individual technology and achieve a more reliable and economical system in which many objects can be tracked and localized. While the reasonable price of RFID tags makes them a good choice for indoor object tracking, the sole use of RFID tags for indoor object localization needs to employ many RFID reference tags to enhance localization accuracy [13]. For outdoor tracking of objects, an integrated system of RFID and GPS-based sensors was developed, eliminating the need for reference tags [13-14]. It was basically based on finding the location of the RFID reader (s) by a GPS receiver and then finding the tag's location through the trilateration technique. In another example, GPS was integrated with RFID to improve positioning and identification capabilities. To localize the RFID tags, a boundary condition-based algorithm was used in which the maximum reading range of a hand-held RFID reader helps to localize the objects within the radius of the trilateration circles [15-16]. In [17], a low-cost integrated GPS-barcode system was designed to track objects in a storage yard. Furthermore, an integrated use of GPS with RFID technologies demonstrated a better performance for tracking onsite resources [2,13-15,18]. Unfortunately, in an indoor environment, the performance of the GPS sensors is highly degraded since they need direct access to the sky to receive signals from satellites. In this case, the UWB system can replace GPS to localize the RFID reader within a defined period. For indoor applications, an integrated RFID and UWB system was developed in [19], providing a low-cost and efficient Indoor Positioning System (IPS) using passive RFID tags.

There are various levels of automation in generating 3D models out of the point cloud data. In a manual approach, the point cloud data are used to manually model the objects by visual inspection. Also, supplementary tools for 3D modelling (i.e., as-built modeler) are used to facilitate the modelling of objects out of the point cloud data. In another approach, AI-based supervised algorithms are implemented in which the as-planned 3D model of the project corresponds to the acquired point cloud data. In these techniques, the supervised classifiers or AI techniques are used to relate every project object in the generated model to a corresponding element in the planned model. The planned model is used as a reference to help identify the objects in the point cloud data. These techniques are semi-automated, however, manual inspection of the generated models is still inevitable. Furthermore, it assumes that the as-planned 3D model is accurate enough, while it was shown that it might need to be corrected in many cases [20]. The AI-based Deep Neural Network (DNN) techniques are also used to model the objects out of point cloud data. These techniques directly use the point cloud data and are trained once for various classes of objects available in the tracked scene. After that, the DNN could detect or segment the objects accordingly. These techniques are categorized as fully automated techniques. However, the accuracy of the DNN in detecting the objects in the scene needs to be improved [20-21].

Sensory data achieved by RTLS and 3D imaging technologies such as laser scanners and depth cameras can be used to enhance the project's Level of development (LOD) through a digital twin platform. Lean 4.0 and Industry 4.0 are two new concepts that can be implemented through a digital twin. The RTLS-based digital twin was investigated in [21-25] to facilitate the practice of Lean 4.0 for object tracking. That study introduced three levels of digitalization and communication between the physical and digital twins. In level 1, only a 3D representation model of the physical object or site is available. In level 2, a one-way data flow from the physical object (s) to the digital object (s) in the model is initialized. In level 3, an integrated and bi-directional connection between the physical object (s) in the field and digital object (s) in the model is fully automated [22-23]. Information about the scheduling of the activities and the objects assigned to each activity can be added to the Building Information Modelling (BIM) to enrich the information about the onsite objects. For this purpose, various RS technologies are used to obtain different types of required information. For example, the RTLS sensors attached to these objects on site, including mechanical equipment and plumbing systems, are acquired information about the location of objects. Transferring these captured raw data to a BIM-based model and updating it in a defined time resolution depends on the type of project activities on site. Every change in the objects' location and status are updated in the BIM model. This information is then processed, and the required actions are taken accordingly and reflected again on the BIM model. The constant communication between the physical environment and the digital model of the project environment through the sensory data communication and the 3D project representation on site help to enhance project inspection and progress reporting through a digital twin platform.

1.2 Research Objectives

This research introduces an innovative framework for the automated generation of inspection reports, leveraging an integrated use of

Real-Time Location System (RTLS), 3D digital imaging, and web-enabled computer vision technologies. These integrations help to bring more visibility into project's operation by tracking actual progress on the job site through efficient use of the mentioned technologies. They also reduce the manual effort required for data collection and generation of inspection reports. The framework will enhance the efficiency of construction project delivery by boosting productivity and facilitating timely decision-making through an improved inspection process.

2 Developed Framework for Generating Inspection Reports

In this study, a sensory-based framework is developed to identify and localize onsite installations for generating inspection reports that comprise the installations' progress and their targeted coordinates. Figure 1 illustrates the framework, including the data collection technologies, data processing algorithms, and generated inspection reports. The framework comprises three methods: (1) an RTLS-based tracking method that utilizes a joint application of RFID and UWB technologies to improve the localization of the tagged objects, (2) a 3D object detection and localization point cloud-based method that utilizes a computer vision algorithm, PointNet, and a ready-to-use platform, Vercator Cloud, and (3) a method that integrates the output of the first two methods.

The first method was developed earlier by the authors to provide location identification information about indoor objects' locations [19]. The second method was also developed earlier by the authors to detect and localize objects using a computer vision algorithm and 3D point cloud data [25]. These two types of data are then integrated using the third method to improve the 3D localization of the RTLS while giving ID to the objects detected by the computer vision algorithms. This integration is performed by mapping the coordinates of the objects realized by these two technologies

Integrated technologies are used in the framework to enhance the coordinates' accuracy of the identified objects (i.e., chairs), along with information about the quantity of the objects. Inspection reports are then generated using the collected data mapped to the digital environment, considering the quantities and locations of the objects. The coordinates' deviation illustrates the difference between the installations' as-built and as-planned locations, including the ± localization error of the integrated RTLS and 3D imaging technologies. The as-planned coordinates are derived from the tie-points on the ground in the laboratory.

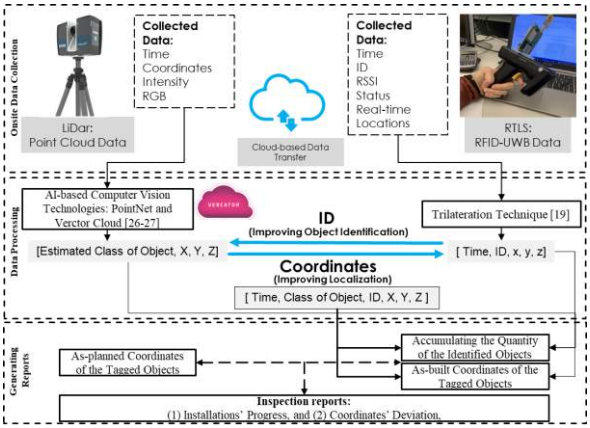


Figure 1. Developed framework for data collection and integration using RTLS and LiDar data.

2.1 RTLS for Object Identification and Localization

Here, an integrated system of RFID and UWB sensors is used. This integration helps to avoid the high cost of UWB implementation for object tracking while benefiting the medium-range capability of the RFID tags [19]. This method aims to provide near real-time localization data of indoor objects. An integrated use of RFID and UWB sensors benefits from the accurate positioning capability of UWB sensors for localizing the hand-held RFID reader in an indoor environment [19, 28-30]. The system can localize the tagged objects using less expensive passive RFID tags while transferring the collected data through cloud-based data collection tools. As illustrated in Figure 1, the RTLS method assigns unique IDs to identified objects, thereby compensating for the limitations of computer vision in distinguishing identical objects.

The input data for this system are gathered using a hand-held RFID reader, which is equipped with an UWB tag. The data are collected in the form of two .txt files, corresponding to the RFID and UWB devices. The RFID file contains the ID and Received Signal Strength Index (RSSI) values, while the UWB file captures real-time 3D coordinates of the roving RFID reader. The system's output includes the estimated 3D coordinates of the RFID tags attached to various objects. The authors' experimental work showed an accuracy of around 1.5 meters for 3D localization of the tagged objects, while 90 percent of the objects are identified and localized by the

integrated RFID-UWB technologies.

2.2 LiDar-based Computer Vision for Object Detection and Localization

This study uses a point cloud-based algorithm, PointNet, to detect and localize objects. An available benchmark is used to train and test the model [25, 28]. The acquired point cloud data collected in the laboratory are fed into the model to validate the algorithm. The algorithm used has an acceptable classification accuracy. However, the results for 3D scene object segmentation and detection, which is required for detecting the target objects in the scene, are unsatisfactory. Only four out of ten target objects are correctly identified by the network. However, the localization accuracy of the detected objects increased to a few centimeters. This problem was reported in [26] and is compensated by integrating the RFID technology in this study. Also, for comparison purposes, object segmentation and detection have been done by a ready-to-use platform in the market, Vercator Cloud, to improve the 3D scene object segmentation and detection [27].

As illustrated in Figure 1, the detected point cloud data obtained through the computer vision algorithm compensates for the low accuracy of the 3D coordinates provided by the RTLS. This enhancement improves the 3D localization of the tagged objects, which is essential for accurately measuring the location of installed components.

2.3 Integrated RTLS and LiDar Data for 3D Object Identification and Localization

Integrating RTLS data with point cloud data enhances the localization and identification of the objects by benefiting the capabilities of these two types of data [20]. Assigning IDs to identical objects through RTLS ensures accurate identification, which is not possible with computer vision algorithms due to their limitations in differentiating identical objects from the same class. Additionally, RTLS is useful in Non-Line-of-Sight (NLoS) scenarios where the LiDAR device cannot collect point cloud data (i.e., in the covered area or invisible objects behind a panel). However, point cloud-based computer vision algorithms provide more accurate localization information with centimetre-level accuracy, which is crucial for many use cases on job sites, such as for automated inspection of the Mechanical, Electrical, and Plumbing (MEP) installations.

For integration, all objects recognized and localized by the RTLS and point cloud data are first derived. Then, for any object tagged by the RFID tag, the objects detected by the point cloud data in the vicinity of the tagged object that belong to the same class of object are selected. For example, if an object (i.e., chairs in this study) is tagged and identified with the RFID tag, the point cloud data detected as a chair by the DNN algorithms are assigned to that chair. The accurate coordinates of the detected objects by point cloud data are then replaced with the location information of the RTLS system. Figure 2 shows the flowchart of the steps for individual use of these technologies and their integration for generating inspection reports.

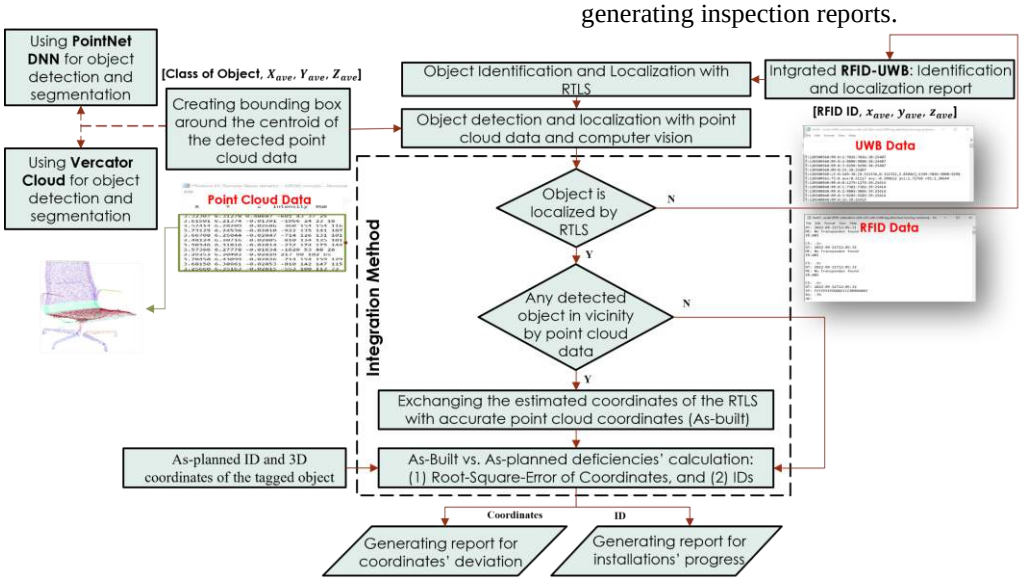
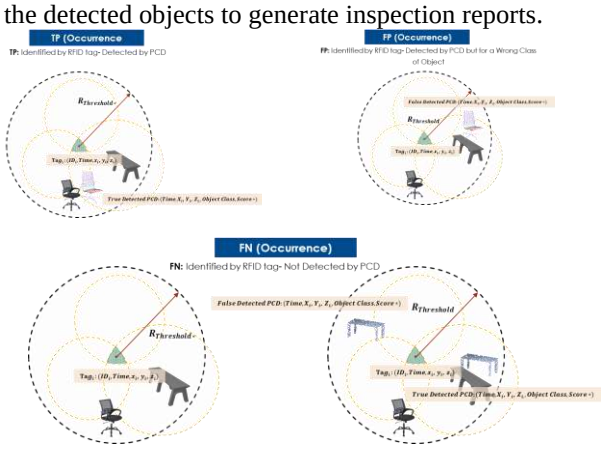


Figure 2. Flowchart of the developed method for integrating RTLS and point cloud data.

The integration enhances the localization accuracy of the RTLS within the coordinates' accuracy of the point cloud data. However, the experiment results show that the occurrence rate for this improvement is only 30-40 percent of the tagged objects when using the PointNet algorithm and available datasets as a benchmark [20]. For the other 60-70 percent of the tagged objects, the identification and localization information achieved by the RTLS is still used to generate the inspection reports. In this study, the computer vision results achieved with PointNet are compared with those achieved by the Vercator cloud platform. Vector Cloud uses optimized computer vision-based algorithms to detect objects from point cloud data and for various classes of objects. Also, it enables us to export the labelled point cloud data for further analysis and integration with RTLS data [27].

Figure 3 illustrates various scenarios for integrating RTLS and point cloud data to detect chairs in the laboratory. Integrating the RTLS data with computer vision data ensures higher accuracy of the objects' 3D coordinates while assigning IDs to the detected objects to generate inspection reports.



$R_{Threshold}$: In Range of the RTLS Localization Error
Score: The score of DNN for a specific Class of Object

Figure 3. Various scenarios in integrating RTLS and point cloud data.

3 Validation of the Developed Framework

The developed framework is validated through an experimental study conducted in a laboratory environment, in which identical chairs are used as the sample class of objects to be identified and localized through the developed framework. The chairs are labelled with passive RFID tags on their top back at a height of one meter and put on the targeted tie points on the ground (Figures 4 and 5). The laboratory is also scanned with a LiDar device to calibrate the RTLS system and collect point cloud data required for the DNN algorithms. Table 1 shows the as-planned coordinates of the chairs on the designated tie-points on the ground, along with information about the assigned RFID ID to each chair. Ten tie-points are selected from a total of 65, where each chair is located on the tie-points.

The tagged chairs are identified and localized by the RTLS technologies, RFID-UWB. At the same time, the chairs are detected and localized by the computer vision algorithms. The integrated method is then used to exchange information to generate inspection reports and identify coordinates' deviation. Table 2 illustrates the estimated coordinates of the chairs as determined by the integrated method used in this study in two scenarios: (1) using the PointNet algorithm and (2) using the Vercator cloud platform to refine RTLS coordinates. The coordinates deviation is calculated using the Root-Mean-Square-Error (RMSE) between the as-built coordinates and the as-planned coordinates of the chair in each tie-point. In a real scenario on the job site, this value is a sum of the deviations from the as-planned coordinates and the integrated method's error in estimating the chairs' coordinates. However, in this experiment, all chairs are precisely located on the as-planned coordinates to identify the integrated method error in estimating the as-built coordinates, as provided in the last column of Table 2. This error could result by various factors that affect the performance of the technologies used on the laboratory or job site environments. For instance, the presence of the metal objects, occlusions and obstacles decrease the localization accuracy of the integrated method. Additionally, the process of mapping RFID-based data to point cloud data reduces the precision of the coordinates detected and localized by the computer vision algorithm. Nevertheless, this integration is essential to improve the detection rate of the integrated method.

To calculate the progress of the chair installations, the number of identified chairs is divided by the total number of planned tie points. In this study, only one chair was not identified and localized by the integrated method. Since all chairs are available at the targeted tie-points, the estimated progress should be 100 percent. However, due to the integrated method error in object identification, this value has a 10 percent error from the actual progress.

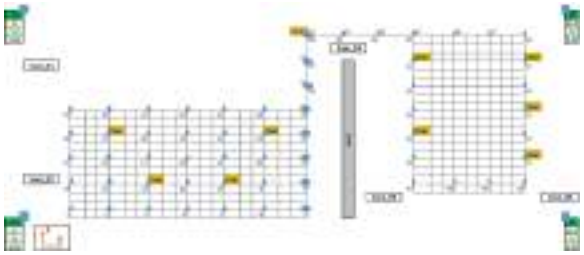


Figure 4. Layout of the tie-points in the lab.



Figure 5. Registered point cloud data of the laboratory.

Table 1. Chairs' targeted coordinates in the tie-points (as-planned status).

No.	Tie-point	ID	X_p(m)	Y_p(m)	Z_p(m)
1	10	A001	4.8	1.2	1
2	12	A002	3	1.2	1
3	23	A003	5.7	3	1
4	27	A004	2.1	3	1
5	38	A005	6.6	6.6	1
6	42	A005	9.3	5.7	1
7	45	A007	9.3	3	1
8	51	A008	12	2.1	1
9	53	A009	12	3.9	1
10	55	A010	12	5.7	1

Table 2. Tagged chairs' estimated coordinates using the integrated method and coordinates' deviations from as-planned status.

No.	Tie-point	X_e(m)	Y_e(m)	Z_e(m)	Coordinates' deviation (RMSE)
Using PointNet Algorithm					
1	10	3.5	1.9	1.8	0.97
2	12*	3.2	1.4	1.5	0.33
3	23*	5.9	3.3	1.2	0.24
4	27	3.5	3.2	2.6	1.23
5	38	6.1	7.4	1.9	0.75
6	42	8.5	6.9	1.7	0.46
7	45*	9.1	3.1	1.2	0.17
8	51	-	-	-	-
9	53*	11.7	3.6	1.5	0.38
10	55	13.4	4.8	1.7	1.042
Average Deviation (m)					0.62
Detection Rate					40%
Estimated Progress					90%
Using Vercator Cloud Platform					
1	10*	4.6	1.1	1.2	0.17
2	12*	3.1	1.2	0.9	0.08
3	23*	5.5	3.2	1	0.16
4	27*	2.2	3.1	1.2	0.14
5	38*	6.7	6.5	0.8	0.14
6	42	8.5	6.9	1.7	0.46
7	45*	9.2	3	1.1	0.14
8	51	-	-	-	-
9	53*	11.8	3.9	0.9	0.13
10	55	13.4	4.8	1.7	1.042
Average Deviation (m)					0.27
Detection Rate					70%
Estimated Progress					90%

* The refined RTLS locations by integrated method using DNN algorithm.

As illustrated in Table 2, the use of Vercator Cloud enhances the detection rate of the chairs from 40 percents to 70 percents. This is due to optimized algorithms used in this platform and with a larger dataset than the PointNet algorithm. Also, that results in more accurate coordinates for the chairs identified through the integrated method. As provided in the table, the chairs are localized more accurately with an average accuracy of 62 centimeters to the 27 centimeters by using Vercator Cloud platform instead of PointNet algorithm. This is close to the required accuracy for approving the installed project's components, which is around 15 centimeters. However, one of the chairs is still not identified through the integrated method, and with RTLS tags (chair number eight in Table 2).

3.1.1 Contributions and limitations

The framework presented in this study automates the generation of inspection reports for construction operations during the construction phase. Integrating the RTLS data with point cloud data enhances indoor object localization by refining the estimated coordinates of the installed project components using the object segmentation capability of the DNN algorithms and accurate coordinates of the point cloud data.

While the developed method automates the generation of inspection reports, it also presents certain limitations. For experimental purposes, the target objects were positioned at the same height and

maintained a clear line of sight (LoS). Future studies could explore the challenges of tag placement at varying elevations and in congested environments, which would provide valuable insights for practical applications on job sites.

The application of the computer vision algorithm used in this study should be extended to more construction components, such as MEPs, which are among the most challenging components to track and report on the job sites. Available point cloud data benchmarks assist in detecting some of these objects without the need for manually creating annotated datasets. However, object detection accuracy is not acceptable for the class of objects with small datasets due to the imbalance in the training dataset size for these objects compared to others. Future investigations are needed to improve object detection with computer vision and to automate the integration of RTLS and point cloud data. This will result in more accuracy in location identification of the detected objects required for approving the installation of the projects' components. Also, the object detection capability of the DNN algorithms could be used to improve the object identification of the RTLS. However, identical objects still cannot be distinguished by DNN algorithms. Further integration with BIM compensates for this limitation.

4 Summery and Concluding Remarks

The framework developed in this study supports the automated generation of inspection reports. The inspection reports include the progress of the installed components and their targeted coordinates on the construction sites. These reports are used to identify deviations from the targeted coordinates of the installed project components by comparing their as-built locations to their as-planned ones. The digital solution technologies employed in the framework are utilized for data capturing, analysis, and reporting, leveraging the developed integration capabilities. These digital solutions include: (1) an RTLS based on integrated RFID-UWB for location identification of tagged objects in an indoor environment, (2) computer vision and deep learning models for object detection and localization of the installed project components using point cloud data, and (3) an integrated method that enables automated generation of the inspection reports through refined localization and recognition of the components. Further integration of the developed method with a project's planned data (i.e., planning and scheduling data) enhances the method's cohesion in automated data acquisition, site inspection, and progress reporting. Additionally, the method has the potential to be further integrated into various quality aspects of a project, including material quality, waste management, and the execution quality of construction operations.

Acknowledgment

The authors gratefully acknowledge the support of Pomerleau team for providing support to this research study.

References

[1] Razavi S.R., Montaser A. and Moselhi O. RFID deployment protocols for indoor construction. *Construction Innovation*, (12)2:239 – 258, 2012.

[2] Moselhi O. and Bardareh H. Automated data acquisition in construction with remote sensing technologies. *Applied Sciences*, 10(8):2846, 2020.

[3] Cheng T. and Venugopal M. Performance evaluation of ultra wideband technology for construction resource location tracking in harsh environments. *Automation in Construction*, 20(8):1173-1184, 2011.

[4] Siddiqui H. UWB RTLS for construction equipment localization: experimental performance analysis and fusion with video data. Master's Thesis, Department of Information Systems Engineering, Concordia University, Montréal, QC, Canada, 2014.

[5] Park J. and Cho Y.K. A BIM and UWB integrated mobile robot navigation system for indoor position tracking applications. *Journal of Construction Engineering and Project Management*, 6(2):30–39, 2016.

[6] Masiero A. and Fissore F. A low cost UWB based solution for direct georeferencing UAV photogrammetry. *Remote Sensing*, 9:414, 2017.

[7] Jimenez A. and Seco F. Comparing decawave and bespoon UWB location systems: Indoor/outdoor performance analysis. *In Proceeding of the 6th International Conference of Indoor Positioning Indoor Navigation* (IPIN), pages 4–7, Alcal´a de Henares, Spain, 2016.

[8] Xu Y. and Shmaliy Y.S. Robust and accurate UWB-based indoor robot localization using integrated EKF/EFIR filtering. *IET Radar Sonar Navigation*. 12:750–756, 2018.

[9] Nurminen H. and Ardeshiri T. A NLOS-robust TOA positioning filter based on a skew-t measurement noise model. *In Proceeding of the International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pages1-7, Alberta,

Canada, 2015.
[10] Sun M. and Wang Y. Indoor positioning integrating pdr/geomagnetic positioning based on the genetic-particle filter. *Applied Sciences*, 10(2):668, 2020.
[11] Almeida A. and Almeida J. Real-time tracking of moving objects using particle filters. *In Proceedings of the IEEE International Symposium on Industrial Electronics*, pages1327-1332, Dubrovnik, Croatia, 2005.
[12] Zhu Z. and Ren X. Visual tracking of construction jobsite workforce and equipment with particle filtering. *Journal of Computing in Civil Engineering*, 30(6):04016023, 2016.
[13] Cai H. and Andoh A. A boundary condition based algorithm for locating construction site objects using RFID and GPS. *Advanced Engineering Informatics*, 28(4):455-468, 2014.
[14] Su X. and Li, S. Enhanced boundary condition–based approach for construction location sensing using RFID and RTK GPS. *Journal of Construction Engineering and Management*, 140(10): 04014048, 2014.
[15] Andoh A.R. and Xing S. A framework of RFID and GPS for tracking construction site dynamics. *Construction Research Congress*, West Lafayette, Indiana, 2012.
[16] Shahi A. and Safa M. Data fusion process management for automated construction progress estimation. *Journal of Computing in Civil Engineering*, 29(6):04014098, 2015.
[17] Song L. and Tanvir M. A cost effective material tracking and locating solution for material laydown yard. *Procedia Engineering*, 123:538–545, 2015.
[18] Li H. and Chan G. Integrating real time positioning systems to improve blind lifting and loading crane operations. *Construction Management and Economics*, 31:596–605, 2013.
[19] Bardareh, H., & Moselhi, O. An Integrated RFID-UWB Method for Indoor Localization of Materials in Construction. *Journal of Information Technology in Construction*, 27, 2022.
[20] Bardareh, H., & Moselhi, O. Integrated Real-Time Location System Information and 3D Point Clouds for Object Localization Onsite Using Deep Neural Networks. AACE International Conference&Expo2023, Chicago, Illinois, DSAA-4094, 2023.
[21] Ma, J. W., & Leite, F. Performance boosting of conventional deep learning-based semantic segmentation leveraging unsupervised clustering. *Automation in Construction*, 136, 104167, 2022.
[22] Tran, T. A., Ruppert, T., Eigner, G., and Abonyi, J. Real-time locating system and digital twin in Lean 4.0. *In proceeding of the IEEE 15th International Symposium on Applied Computational Intelligence and Informatics (SACI)*, page 000369-000374, 2021.
[23] Huang, Y., Hammad, A., Torabi, G., Ghelmani, A., and Guevremont, M. Towards Near Real-time Digital Twins of Construction Sites: Developing High LOD 4D Simulation Based on Computer Vision and RTLS. *In the Proceedings of the International Symposium on Automation and Robotics in Construction (ISARC)*, 38, pages 248-255, 2021.
[24] Sander, P., Spiegl, M., Burns, T. and Reilly, J. Digital project twin for quantitative cost, risk and schedule assessment of capital projects. *Australian Journal of Multi-Disciplinary Engineering*, 18(1), pages 34-46, 2022.
[25] Ruppert, T., and Abonyi, J. Integration of real-time locating systems into digital twins. *Journal of industrial information integration*, 20:100174, 2020.
[26] Qi, C. R., Su, H., Mo, K., & Guibas, L. J. Pointnet: Deep learning on point sets for 3d classification and segmentation. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652-660, 2017.
[27] Available online: https://vercator.com/
[28] Ruiz A. R. J. and Granja F. S. Comparing ubisense, bespoon, and decawave uwb location systems: Indoor performance analysis. *IEEE Transactions on instrumentation and Measurement*, 66(8):2106-2117, 2017.
[29] Bardareh, H., & Moselhi, O. Automated Data Acquisition for Indoor Localization and Tracking of Materials Onsite. *(ISARC)*, 37, pages 765-772), 2020.
[30] Bhatia, A., Moselhi, O., & Han, S. H. Optimized production scheduling for modular construction manufacturing. *In the Proceedings of the International Symposium on Automation and Robotics in Construction (ISARC)*, Vol. 40, pp. 270-277). IAARC Publications, 2023.

Automated assembly sequence planning and scheduling in precast building projects using reinforcement learning

Ajay Kumar Agrawal[1], Yang Zou[1], Hongyu Jin[1], Mohammed Abdelmegid[2], Vicente Gonzalez[3]

[1]Department of Civil and Environmental Engineering, University of Auckland, New Zealand
[2]School of Civil Engineering, University of Leeds, UK
[3]Department of Civil and Environmental Engineering, University of Alberta, Canada

aagr657@aucklanduni.ac.nz, yang.zou@auckland.ac.nz, hongyu.jin@auckland.ac.nz, M.Abdelmegid@leeds.ac.uk, vagonzal@ualberta.ca

Abstract

Efficient construction of precast concrete buildings (PCB) with high quality and safety requires adequate planning and scheduling of the assembly of precast components (PC). Traditional manual assembly sequence planning and scheduling (ASPS) methods often result in sub-optimal and error-prone schedules, especially in complex projects. Existing metaheuristic-based ASPS methods lack generalizability and capability to handle dynamic conditions in construction projects. Reinforcement Learning (RL) offers a promising solution to these limitations. However, existing RL-based construction scheduling studies overlook project uncertainties and rely on value-based methods, which are computationally inefficient for large-scale problems. This study proposes a novel method for optimal ASPS of PCB projects using RL integrated with Monte-Carlo Sampling (MCS). The method uses a Temporal Graph Network (TGN) and Multi-Layer Perceptron (MLP) to create embeddings of the relevant input information, while the Proximal Policy Optimization (PPO) RL algorithm is employed for generating the assembly schedules using the embeddings. Reward shaping is utilized to generate schedules that simultaneously minimize time, cost, and constraint violations. Preliminary results from a case study on a hypothetical PC project with 34 precast concrete walls demonstrate the proposed method to provide stable learning and generate optimum assembly schedules. The proposed method outperforms existing value-based RL approaches for construction scheduling, demonstrating superior training stability, solution quality, and execution efficiency.

Keywords –

Assembly sequence planning and scheduling; Precast building construction; Reinforcement learning; Proximal policy optimization

1 Introduction

The growing housing demand has created the need to build 96000 homes every day worldwide [1]. Prefabricated Housing Construction (PHC), an innovative construction method, provides a high-speed and sustainable solution to this challenge [2]. It involves offsite manufacturing of components like columns and walls, which are transported and assembled onsite. Among prefabrication methods, precast concrete buildings (PCB) are widely adopted, using precast concrete elements [2]. Efficient assembly sequencing planning and scheduling (ASPS), considering constraints like resource and material availability, is vital for PCB project success [3]. ASPS refers to the process of defining the sequence and timeline of assembly of precast components (PC). Conventionally, ASPS is done manually by practitioners based on their experience, leading to suboptimal and error-prone results, especially in large and complex projects [4]. Thus, there is a growing focus on automating ASPS in PCB projects.

Existing studies in this direction have primarily utilized metaheuristic methods such as Genetic Algorithm (GA) and Simulated Annealing (SA) [5]. However, such methods lack generalizability to different types of projects [6], and cannot capture the uncertainties and dynamic situations in the construction projects [7].

To this end, RL has been used recently to address construction planning problems [7]. RL is a machine learning technique in which an agent learns to make the best decisions by mimicking humans' trial-and-error learning process. Existing RL-based construction planning techniques have certain limitations, hindering their applicability in ASPS problems. First, they are primarily designed for conventional on-site construction projects. Second, ASPS involves sequencing numerous PCs, creating a large activity set that existing methods struggle to handle due to dimensionality challenges.

Finally, the construction processes' uncertainties were often overlooked. PCB processes can face uncertainties for various reasons, such as component characteristics and weather conditions [8], which need to be taken into account in ASPS.

To address these limitations, this study proposes a novel RL-based method to automatically generate the optimum assembly schedules for PCB, considering installation constraints and uncertainties in the activity durations and weather conditions.

Key innovations of this study are as follows:

a) A novel architecture is proposed using the Proximal Policy Optimization (PPO) algorithm [9], an actor-critic RL algorithm that can effectively handle large state and action spaces compared to the algorithms used in existing construction planning studies.
b) State, action space, and reward function, which play a crucial role in RL, are uniquely defined considering optimization objectives as minimizing time, cost constraint violations [3].
c) An approach to handle uncertainties through Monto Carlo Sampling (MCS) is incorporated with RL.

2 Literature review

2.1 Assembly Sequence Planning and Scheduling in PCB Projects

Initial studies in this direction utilized structural support and topological relationships between the building components to generate assembly sequences [10]. However, schedule optimization was not done. Later, researchers utilized Building Information Models and optimization algorithms to develop optimum assembly sequences. Optimization of ASPS has been done from various perspectives, such as maximization of structural stability [11], prioritization of heavier and larger components, minimization of installation time [3]–[5], [12], and reduction in interference between components, among others [12], [13]. Most of the studies utilized GA for optimization [3], [4], [11], [13]. To avoid the risk of slow convergence in large-scale problems, Liu et al. [5] integrated GA with SA. Other researchers used modified versions of GA to ensure population diversity [4]. However, existing studies failed to simultaneously consider assembly time, cost, and constraint violation as schedule optimization objectives. They lack the ability to swiftly adapt to dynamic conditions and uncertainties on the construction projects, as the solutions developed by GA only apply to the initial conditions of the optimization [14].

2.2 RL for construction planning

In RL, an agent learns to find the policy to make the best decisions by iteratively interacting with an environment, as shown in Figure 1. At each time step, the agent selects an action from the set of possible actions (known as action space), and the environment rewards the agent based on the selected action. Over the iterations, the agent improves the decision-making by maximizing the reward. In contrast to algorithms such as GA, RL algorithms can handle dynamic situations more efficiently as they automatically learn the best optimization strategy based on the interaction with the environment [6].

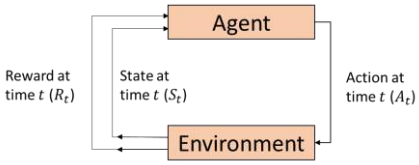


Figure 1. RL Model

Soman et al. [15] used Q-learning (a class of RL algorithms that maximizes the Q-function to find the best policy by estimating cumulative rewards for each state-action pair) to generate construction lookahead schedules automatically. Kedir et al. [7] utilized deep learning to estimate Q-function. Yao et al. [6] utilized Deep Q-learning (DQN) and sampling constraint-free actions to achieve faster convergence. However, they addressed precedence and resource constraints only. ASPS in PCB requires consideration of several constraints, such as the size and weight of the PC and interference between PCs. Moreover, the uncertainties of the construction environment were not considered, and only time-based schedule optimization was performed.

The Q-learning algorithm becomes computationally expensive for large-scale problems, for example, resource-constrained construction planning and scheduling [16]. The complexity further escalates in uncertain environments, where state spaces become continuous due to variables like activity durations and weather conditions. DQN algorithm is faster than Q-learning, as the Q-function estimation is done using deep learning. However, it is prone to suffer from instability in such cases of continuous state spaces [17]. Actor-critic algorithms, a class of RL algorithms, combine the strength of value and policy-based approaches. These algorithms use an actor to update the policy directly and a critic to evaluate the value function. Proximal Policy Optimization (PPO), a state-of-the-art actor-critic algorithm, balances exploration and stability [9], making it effective for handling complex, dynamic, and large-scale problems like ASPS. Despite its potential, PPO has not been extensively applied in this domain.

2.3 Problem statement

Based on the identified research gaps, this study proposes a method for ASPS in PCB projects using the PPO algorithm to generate assembly schedules that:

a) minimize time, cost, and constraint violations associated with the assembly. These objectives are selected based on their frequent use in the existing literature [3], [12], [13].
b) incorporate the uncertainty in the PCB projects. Two primary sources of uncertainties are considered: inherent variations in the activity durations and external uncertainty due to weather, both of which are prevalent in PCB projects [8].

3 Methodology

3.1 Problem formulation and mathematical modeling

Consider that the PCB consists of n elements to be scheduled. Each of these elements belongs to the component type $CT = [1,2,3, \dots i]$. The component types correspond to different categories in which the components can be divided, e.g., beams, columns, walls, etc. Each element e_k requires renewable resources $r_k = [r_{ka}, r_{kb}, r_{kc,\dots,} r_{km}]$, where r_{ka} denotes the requirement of resources of the category a. $R = [R_a, R_b, R_c, \dots, R_m]$ denotes the maximum availability of renewable resources in categories $[a, b, c, \dots, m]$. Each element e_k requires time t_k for assembly. The time requirement t_k for assembly of each element is assumed to follow a triangular distribution based on the findings of previous studies [18]. Weather conditions, including rain, wind, and temperature, are considered to impact the assembly durations [8]. The distribution of the values of these three weather conditions can be taken from a weather forecast [19]. The impact of weather conditions on the assembly durations can be identified using general codes of practice and relevant laws and regulations [8].

3.1.1 Objectives

$$minimize\ D_{total}, C_{total}\ and\ CV_{total} \quad (1)$$

Where D_{total} is the total scheduled duration of the project, C_{total} is the total cost (including direct cost and indirect cost) of the project and CV_{total} denotes the total number of constraint violations in the generated schedule.

3.1.2 Constraints

The key constraints for ASPS were identified based on the literature. The constraints are divided into two categories: Hard constraints and soft constraints.

(a) Hard constraints: The constraints that must not be violated as they ensure the structural integrity of the building and adherence to scheduling logic.
  (i) Structural support constraints: The PC should be assembled according to structural support requirements [11].
  (ii) Resource availability constraint: The allocated resources must be within the limits of the R [8].
  (iii) Component availability: A PC can be assembled only after its delivery on-site [8]. A PC delivery schedule is used to consider component availability.
(b) Soft constraints: The constraints whose violation should be minimized as much as possible.
  (i) Weight constraint: The heavier the PC, the earlier it should be assembled [3].
  (ii) Size constraint: The larger the space occupied by the PC, the earlier it should be assembled [3], [12].
  (iii) Interference constraint: Interference between the new PC and the pre-installed PC should be minimized [3].

3.2 RL model for ASPS using PPO

3.2.1 Network architecture

The network architecture of the proposed model for automated ASPS using PPO is depicted in **Error! Reference source not found.**.

3.2.1.1 State:

The state represents the combination of information required by the agent for decision-making. In this study, the state information at a timestep includes the following information:

(a) The construction schedule generated till the timestep is represented in the form of a graph. Each graph node represents the elements and their properties, and the edges represent their relationships. The properties of the nodes are represented in **Error! Reference source not found.**. The first property represents the status (not started, in progress, completed) of the element. The second property corresponds to CT. The third and fourth properties represent the resources needed. They can be customized based on the number of categories of resources. The next two properties focus on the completed and remaining duration of elements till the current timestep. The last two properties represent the weight and the area of the elements, respectively.
(b) The availability of PC on-site, represented using a binary vector of length equal to n.

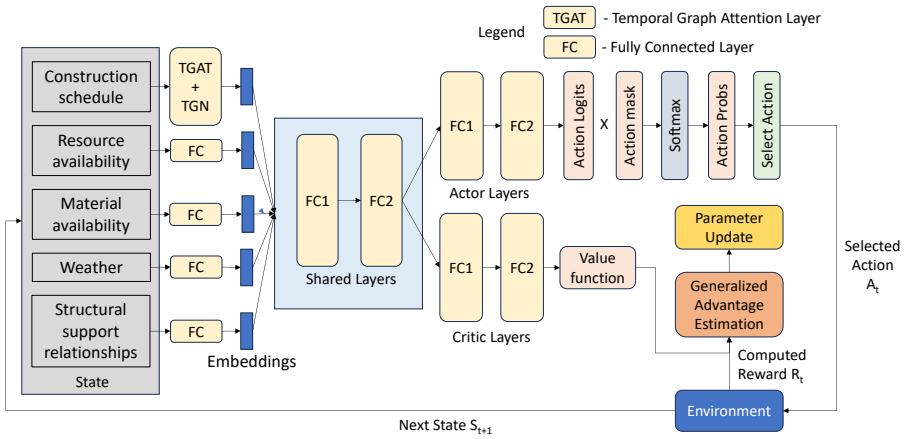


Figure 2 . Model architecture

Element status	Element type	Crane needed	Labor needed	Elapsed duration	Remaining duration	Element weight	Element area

Figure 3 . Node features

(c) The availability of resources of different categories, represented using a vector of length equal to R.

(d) Structural support relationships in the form of binary vector S of shape $[n, n]$. A value of $S_{ij} = 1$ represents that the element i is supported by the element j.

(e) Weather is represented in the form of a vector of dimension $[3, T]$, where T denotes the length of the forecast period for which the weather data is considered. Each column represents a particular day, and three rows represent the sampled values of rain, temperature, and wind.

3.2.1.2 Action:

Considering the n elements to be scheduled, the action space is a matrix A of size 2^n, where each row represents a binary vector of length n, indicating which elements should be started (1 for start, 0 otherwise). In PCB, the number of elements installed simultaneously is limited by the project's available cranes, so actions exceeding this limit are removed.

3.2.1.3 Reward:

As this study intends to address three optimization objectives, i.e., time, cost, and constraint violations simultaneously, the reward structure used in this study consists of the following components addressing these objectives:

(a) Time:

$$R_{t1} = r_{w1} * N_s + r_{w2} * N_f - r_{w3} * t$$

$$R_{t2} = r_{w4}$$

Where R_{t1} is the reward given to each timestep. It includes a positive reward for the number of elements started (N_s), the number of elements finished (N_f), and negative reward for each timestep (t). A large positive reward of R_{t2} is added to finish all the elements. r_{w1}, r_{w2}, r_{w3}, and r_{w4} are scalar reward coefficients.

(b) Cost:

$$R_c = -r_{w5} * [\Delta t * (R_{ct} + I_{ct})]$$

Where R_c is the cost reward per time step, Δt is the time difference between the current and previous timestep, R_{ct} and I_{ct} are resource and indirect costs per unit of time, and r_{w5} is the reward coefficient.

(c) Constraint violation:

$$R_{con} = -r_{w6} * N_{sc}$$

Where R_{con} is the reward for constraint violations, N_{sc} is the number of soft constraint violations in that timestep, and r_{w6} is the reward coefficient. The violations of weight and area constraints are computed based on the weight and area values in the node features. For interference constraint violations, an adjacency matrix and an interference matrix are used [4]. The adjacency matrix represents the adjacent elements. The interference matrix is a $[n, n]$ matrix, where each element is a 1D vector IC_{ij} of length 6, which is as follows:

$$IC_{ij} = [d_{x+}, d_{x-}, d_{y+}, d_{y-}, d_{z+}, d_{z-}]$$

Where:

IC_{ij} = a binary vector of interferences if the element j is installed before the element i,

$$d_* = \begin{cases} 1, if\ \ i\ can\ move\ in\ that\ direction \\ 0, otherwise \end{cases}$$

For example:

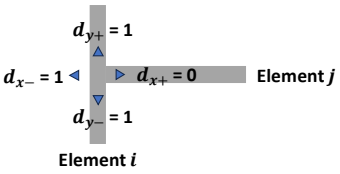


Figure 4. Plan of connection between two elements

3.2.1.4 Environment:

(a) Initialization:
The environment is initialized with constraints such as resource availability, structural support constraints, element weights, etc. The MCS is used to handle uncertainties in activity durations and weather conditions. MCS uses repeated random sampling of a stochastic variable from its distribution to approximate the desired quantity [20]. It finds suitability in solving stochastic optimization problems due to its simplicity and ability to approximate well with a small number of samples [20]. We sampled installation durations and weather conditions from their distributions at each episode's start. The episode is then run using these values. This way, the RL agent learns to perform well over the distribution of installation durations and weather conditions.

(b) Scheduling Logic:
At each timestep, the agent selects an action, and the environment updates the schedule, calculates rewards, and returns the updated state and reward. The elements in the selected action are started first, and resources are assigned to them. Then, the earliest finishing element is finished, and its resources are released. The rewards are calculated, and the next state is sent to the agent. The episode ends when all elements are scheduled.

3.2.1.5 Network architecture

The state information, including the resource availability vector, material availability vector, weather vector, and structural support relationship vectors, are embedded using Fully Connected (FC) layers. The construction schedule, which is represented in the form of a graph, is embedded using a combination of Temporal Graph Attention Network (TGAT) [21] and Temporal Graph Network (TGN) [22]. At each timestep, the graph evolves as the node features and their relationships change. The conventional GNNs used in existing RL-based construction scheduling studies [6] fail to capture such temporally evolving dynamic graphs [22]. TGNs provide an efficient way to deal with such dynamic and time-dependent networks [22]. TGN, when used with TGAT, encodes time features in addition to the node features and uses a self-attention mechanism [23] to identify important parts of the graph to generate the embedding. TGN also incorporates a memory module using Gated Recurrent Units to retain information from previous timesteps, enabling the network to understand how the schedule evolves over time. The actor-critic network includes two shared FC layers and two separate FC layers.

3.2.1.6 Action masking

An action mask is created for actions that violate hard constraints. It filters out such invalid actions and improves training efficiency by reducing the number of possible actions.

4 Case study

To validate the proposed method, a hypothetical project consisting of ASPS of 34 precast walls is used based on existing studies [4], [8]. The project consists of walls of varying thicknesses. Figure 5 shows the walls in the project along with their ID.

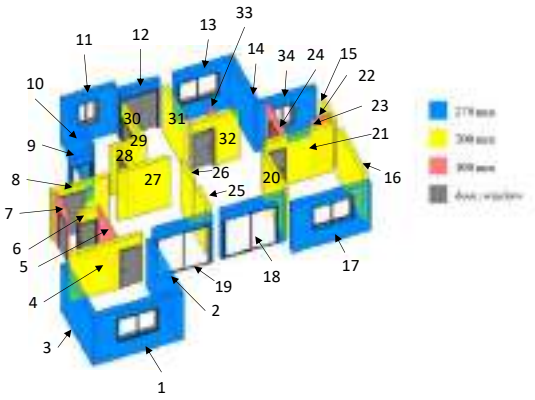


Figure 5. Case study project (Picture credit [4])

As the weights and areas of the walls were not available in Yang et al. [4], they were assumed based on the relative sizes of walls in Figure 5. Similarly, the distribution of installation duration of the walls was not available. Therefore, all the walls were assumed to have the same triangular distribution of durations, derived from Yuan et al. [18]. The assumed installation duration distribution for each wall is (0.47, 0.625, 0.78) hrs. The details of the walls are shown in Figure 5, Table 1, and Table 2. The project is assumed to be constructed in Auckland, New Zealand (NZ). The weather forecast of Auckland is taken from online sources [24], [25]. The impact of weather conditions on construction activities is considered based on local industry standards. Based on the weather forecast, the temperature and rain values are

found to have no impact, and the impact of wind speed is considered based on recommendations from the Crane Association of NZ [26]. Lifting is assumed to involve two mobile cranes, with worker requirements based on Yuan et al. [8]. The larger walls (Area >= 2.5m^2) are assumed to require 6 workers, and others require 4 workers. Costs for the crane and workers are derived from Rawlinsons New Zealand Construction Handbook, 2014 [27]. All the elements were assumed to be already available.

Table 1 Wall weights

Wall ID	Weight (kg)	Wall ID	Weight (kg)
6, 7, 22, 23, 33	50	13	225
5, 9, 15	75	1, 17	250
24, 25, 26, 28	100	11, 34	275
8, 12, 18, 19, 29, 30, 32	150	2, 3, 10, 14	300
4, 16, 20, 21, 27, 31	175		

Table 2: Wall areas

Wall ID	Area (m^2)
22, 23, 25, 26, 33	0.5
6, 7, 9, 15, 28	0.75
5, 24, 30, 32, 29, 31	1.25
8, 12, 27	2
2, 3, 10, 13, 14, 16, 20	2.5
11, 18, 19, 34	2.75
1, 4, 17, 21	3

The validation experiment included two facets: (1) a comparison between the performance of the developed approach with existing DQN-based construction scheduling methods and (2) a sensitivity analysis to understand the impact of scales of reward coefficients. For the first aspect, the comparison was done between the results obtained using DQN+GNN (DQN RL algorithm and GNN for schedule embedding [6]), and PPO+TGN+TGAT (proposed in this study). The key hyperparameters for the three settings are described in Table 3. For the second facet, 7 scenarios of reward coefficient were created, as described in Table 4.

Table 3: Key Hyperparameters for training

Hyperparameter	DQN+ GNN	PPO+ TGN+ GAT
PPO clip epsilon	NA	0.25
PPO entropy coeff.	NA	0.08
Learning rate	0.0003	0.0003
Gamma	0.99	0.99
GAE lambda	NA	0.95
Graph network layers	2	2
Epsilon decay	1000	NA
Epsilon start, end	1, 0.01	NA

Table 4: Sensitivity analysis scenarios

Scenario	r_{w1}	r_{w2}	r_{w3}	r_{w4}	r_{w5}	r_{w6}
S1	1000	1000	2000	20000	5	1000
S2	1000	1000	2000	20000	5	10
S3	1000	1000	2000	20000	0.05	1000
S4	1000	1000	2000	20000	0.05	10
S5	10	10	20	200	5	1000
S6	10	10	20	200	0.05	1000
S7	10	10	20	200	5	10

The algorithm was run on a Windows 11 computer, with Intel® Xeon® Gold 6242R CPU @ 3.10GHz processor and NVIDIA Quadro RTX 5000 16.0 GB GPU for 5000 episodes. The high number of episodes was to consider the uncertainties in the problem, such that the agent can learn over the distribution of duration and weather conditions based on MCS. The results of training on Scenario S1 are demonstrated in Figure 6**Error! Reference source not found.Error! Reference source not found.**. The graphs show the 100-episode average of computed rewards, costs, schedule durations, and constraint violations. Since each episode is trained on different initial conditions, averaging over 100 episodes provides a clearer trend. While existing studies developed assembly sequences focusing on single objectives, our method effectively generated schedules minimizing cost, time, and constraint violations simultaneously. For PPO+TGN+TGAT, the agent successfully discovered a policy that maximizes the reward, achieving greater consistency after approximately 2000–3000 episodes. Similarly, the number of constraint violations decreased and stabilized around the same point. The time and cost showed only a mild reduction. A maximum of two elements could be installed parallelly due to the availability of only 2 cranes. Re-sequencing the elements did not reduce the duration and cost, as the duration distribution of the elements was the same. Further, the schedule duration and cost exhibited higher variability, likely due to the inherent variability in sampled durations and weather conditions, which have a greater impact on these metrics.

Our method significantly outperforms the DQN+GNN [6] based method used in existing studies, delivering superior solution quality, enhanced training stability, and improved execution time. As evident from Figure 6, the total reward and constraint violations identified using PPO+TGN+TGAT were significantly better than DQN+GNN-based solutions. Although the time and cost were quite close, PPO+TGN+TGAT showed better convergence as compared to DQN+GNN. Moreover, for all four criteria, PPO+TGN+TGAT demonstrated superior stability of training. DQN utilizes bootstrapping by estimating Q-function, which can be highly unstable [17], especially with large state space created in the current problem due to uncertainty. Since the parameter updates in DQN solely rely on Mean

Squared Error loss, small overestimation or underestimation of Q-values can lead to large gradient updates, causing instability in the training. PPO directly optimizes the policy using a clipped loss function, limiting the extent of updates in each iteration and providing better stability in learning [9]. The TGN+TGAT network incorporates the temporal aspect of the schedule and keeps a memory embedding of the actions taken in the previous timesteps, handling the dynamically changing graph better. As a result, the objectives could be minimized to a greater extent as compared to the DQN+GNN-based method. To train the 5000 episodes, DQN+GNN took around 12 hours, while PPO+TGN+TGAT required nearly 6.8 hours. DQN requires the management of a replay buffer and target network updates, which are not needed in PPO, resulting in lower execution time for PPO+TGN+TGAT in the validation experiment. However, further work is needed to reduce computational demands considering the near-real-time application requirements.

The results of sensitivity analysis using scenarios in Table 4 are demonstrated in Figure 7. The results show the impact of variation in reward coefficient scales. Since the time and cost varied within a very narrow range, the variation in the total reward mainly depended upon r_{w6}. As a result, the total reward for S2, S4, and S7 was found to be mostly consistent. For the rest 4 of the scenarios, an evident increment in the reward was found. S1 performed generally well across all three objectives as the reward coefficients were decided to provide relatively similar weightage to all three objectives.

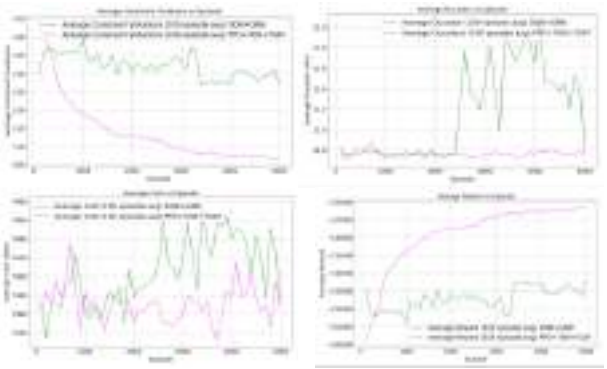


Figure 6. Validation results on scenario S1

5 Conclusion and Outlook

This study proposes a novel method for automated multi-objective ASPS optimization in PCB projects under weather and activity duration uncertainties using RL. PPO, a popular actor-critic RL algorithm, is combined with MCS to train an agent to generate assembly schedules minimizing time, cost, and constraint violations. Activity durations and weather conditions are sampled at each episode via MCS, enabling the agent to learn optimal policies under uncertainties. A novel PPO network architecture is proposed for ASPS, utilizing TGAT and TGN to create the embeddings of the schedule generated at each timestep. Additional state information is encoded using FC layers, while the actor-critic network consists of 2 shared and 2 separate FC layers. The method is validated on a hypothetical project involving the assembly of 34 precast walls. The case study results show that the proposed method effectively identifies assembly schedules that minimize cost, time, and constraint violations while maximizing rewards under uncertainties. It outperformed the conventional DQN+GNN-based method used in existing RL-based construction scheduling studies.

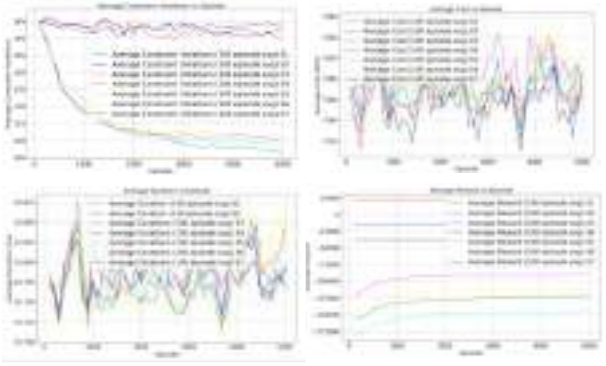


Figure 7. Reward coefficient sensitivity analysis results

The study has limitations that need to be addressed to improve the robustness and industry applicability of the proposed method. First, validation using real-world, complex PCB project data is required. A systematic identification of optimal reward coefficients for better results should be explored. Finally, the performance of the developed method should be compared with other RL algorithms, such as Deep Deterministic Policy Gradient (DDPG), Double DQN, and Soft Actor Critic (SAC), etc.

6 References

[1] 4 practical solutions to the global housing crisis. World Economic Forum. Jun. 10, 2024. On-line: https://www.weforum.org/stories/2024/06/global-housing-crisis-practical-solutions/, Accessed: 12/13/2024.

[2] Y. Fang, S. Gao, Y. Jiang, and S. Li. BIM and lean construction in prefabricated housing construction in China. *International Journal of Lean Six Sigma*, 14(7): 1329–1353, 2023.

[3] Y. Wang, Z. Yuan, and C. Sun. Research on assembly sequence planning and optimization of

precast concrete buildings. *Journal of Civil Engineering and Management*, 24(2): 2018.

[4] B. Yang, S. Jiang, M. Dong, D. Zhu, and Y. Han. Graph Database and Matrix-Based Intelligent Generation of the Assembly Sequence of Prefabricated Building Components. *Applied Sciences*, 13(17): 2023.

[5] C. Liu, F. Zhang, H. Zhang, Z. Shi, and H. Zhu. Optimization of assembly sequence of building components based on simulated annealing genetic algorithm. *Alexandria Engineering Journal*, 62: 257–268, 2023.

[6] Y. Yao, V. W. Y. Tam, J. Wang, K. N. Le, and A. Butera. Automated construction scheduling using deep reinforcement learning with valid action sampling. *Automation in Construction*, 166: 105622, 2024.

[7] N. S. Kedir, S. Somi, A. R. Fayek, and P. H. D. Nguyen. Hybridization of reinforcement learning and agent-based modeling to optimize construction planning and scheduling. *Automation in Construction*, 142: 104498, 2022.

[8] Z. Yuan *et al.* Simulation and optimization of prefabricated building construction considering multiple objectives and uncertain factors. *Journal of Building Engineering*, 86: 108830, 2024.

[9] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov. Proximal Policy Optimization Algorithms. arXiv. Aug. 28, 2017.

[10] H. Liu, Z. Lei, H. X. Li, and M. Al-Hussein. An Automatic Scheduling Approach: Building Information Modeling-based Onsite Scheduling for Panelized Construction, 1666–1675, 2014.

[11] V. Faghihi, K. F. Reinschmidt, and J. H. Kang. Construction scheduling using Genetic Algorithm based on Building Information Model. *Expert Systems with Applications*, 41(16): 7565–7578, 2014.

[12] L. Huang, R. Pradhan, S. Dutta, and Y. Cai. BIM4D-based scheduling for assembling and lifting in precast-enabled construction. *Automation in Construction*, 133: 103999, 2022.

[13] S. Bae, H. Cha, and S. Jiang. Building an Information Modeling-Based System for Automatically Generating the Assembly Sequence of Precast Concrete Components Using a Genetic Algorithm. *Applied Sciences*, 14(4): Art. no.4 2024.

[14] E. Gazioğlu and A. S. Etaner-Uyar. Experimental analysis of a statistical multiploid genetic algorithm for dynamic environments, *Engineering Science and Technology, an International Journal*, 35: 101173, 2022.

[15] R. K. Soman and M. Molina-Solana. Automating look-ahead schedule generation for construction using linked-data based constraint checking and reinforcement learning. *AUTOMATION IN CONSTRUCTION*, 134: 2022.

[16] R. Pellerin, N. Perrier, and F. Berthaut. A survey of hybrid metaheuristics for the resource-constrained project scheduling problem. *European Journal of Operational Research*, 280(2): 395–416, 2020.

[17] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*, Cambridge, Massachussetts: The MIT Press. 2018.

[18] Z. Yuan, Y. Chang, Y. Chen, Y. Wang, W. Huang, and C. Chen. Simulating and optimizing precast wall lifting in prefabricated building construction. *Engineering, Construction and Architectural Management*. 2022.

[19] E. Mohamed, P. Jafari, A. Chehouri, and S. AbouRizk. Simulation-Based Approach for Lookahead Scheduling of Onshore Wind Projects Subject to Weather Risk. *Sustainability*, 13(18): 2021.

[20] T. Homem-de-Mello and G. Bayraksan. Monte Carlo sampling-based methods for stochastic optimization. *Surveys in Operations Research and Management Science*, 19(1): 56–85, 2014.

[21] D. Xu, C. Ruan, E. Korpeoglu, S. Kumar, and K. Achan. Inductive Representation Learning on Temporal Graphs. arXiv. Feb. 19, 2020.

[22] E. Rossi, B. Chamberlain, F. Frasca, D. Eynard, F. Monti, and M. Bronstein. Temporal Graph Networks for Deep Learning on Dynamic Graphs. arXiv. Oct. 09, 2020.

[23] A. Vaswani *et al.* Attention Is All You Need. arXiv. Aug. 02, 2023.

[24] Auckland Central Weather Forecast and Observations - MetService New Zealand,. On-line: https://www.metservice.com/towns-cities/regions/auckland/locations/auckland, Accessed: 12/30/2024.

[25] Wind, waves, weather & tide forecast Auckland Airport - Windfinder,. On-line: https://www.windfinder.com/forecast/auckland, Accessed: 12/30/2024.

[26] Safety Wind Speed,. Crane Association of New Zealand (INC). On-line: https://www.cranes.org.nz/uploads/2/0/5/7/20572552/wind4.pdf, Accessed: 12/30/2024.

[27] *Rawlinsons New Zealand Construction Handbook*, Rawlinsons Publications. 2014.

Analysis and Simulation of Pile Driving Construction Equipment

Laszlo Kovacs[1], Eric Karpman[2], Marek Teichmann[1], Jozsef Kövecses[2]

[1]CM Labs Simulations, Montreal, Canada
[2]Dept. of Mechanical Engineering and Centre for Intelligent Machines
McGill University, Montreal, Canada

laszlo.kovacs@cm-labs.com, eric.karpman@mail.mcgill.ca

Abstract -

For real-time simulation of construction equipment, the operation of blades and other tools in contact with soil represent some of the most challenging problems both in terms of performance and accuracy. Soft soil interactions are particularly difficult to model and simulate efficiently, yet, a physically representative, interactive simulation that provides real-time feedback to the operator is required in operator training and certification scenarios, as well as for human in the loop engineering applications.
In this paper, we address the problem of simulating in real-time several piling techniques that are fundamental in construction. A sufficiently accurate resistance force/torque model is essential for the realistic simulation of piling equipment. Particle based models and finite element soil interaction models are too computationally expensive for most real-time simulations. We propose an analytical approach for estimating the piling resistance motivated by bearing capacity theories developed for deep foundations. The proposed model can accurately capture the soil resistance in three different piling scenarios: hammering, drilled displacement piling, and continuous flight augering. The vertical resistance force and torque, and their coupled nature is considered with kinetic friction effects. The presented models are validated against experimental data from subject matter experts. The results obtained for continuous flight augering are also compared to results from a discrete element simulation. To improve the realism of the analytical model, a hybrid numerical-analytical model is outlined. The proposed real-time models support fast equipment design iterations, control development, automation and interactive operator training.

Keywords -

real-time simulation. pile driving. soil-tool interaction. ground resistance. operator training.

1 Introduction

Using simulators to train operators of heavy construction site equipment requires the ability to efficiently model and simulate the interactions between tools and soil in an interactive and realistic way so that real time feedback can be provided to a human in the loop. In this work, our goal is to create a model of a drill rig that creates holes for foundation piles, an operation known as piling or pile driving. We propose a real-time model for three fundamental piling techniques used in the construction industry with a level of fidelity and realism that is appropriate for training drill rig operators. The three piling techniques of interest are the displacement pile with lost drill point, the continuous flight auger (CFA) and the hydrohammer.

The soil displacement piling operation with a lost drill point is an operation in which a large cylindrical pile with a conical tip is drilled into the ground by applying both a vertical force and a torque. Rather than removing soil from the ground, this operation displaces the soil downward and laterally. Once the pile has reached the desired depth, the pile is removed, leaving behind the drill point. The resulting hole is filled with concrete. Continuous flight augering is another excavation process in which the drill bit consists of a helical blade that removes the soil as it drills into the ground and a hollow shaft that is used to inject the concrete as the drill is removed.

Finally, the hydrohammer, or hydraulic impact hammer, is a tool that can be used to create holes for foundation piles through repeated hammering of a steel pile into the ground until the desired depth is reached. Fig. 1 illustrates the drilling tools used for each of these operations.

The Earthworks module in Vortex Studio [1], which is used for the models that will be described in this work, uses analytical formulas to estimate the cutting force resisting the motion of the tools interacting with the soil because these are efficient enough for real-time simulation. For the representation of the disturbed soil in motion, a particle based model is used.

In this paper, we present simple and efficient models



Figure 1. Displacement pile (top), continuous flight auger (middle) and hydrohammer (bottom) on a drill rig

to consider the force-torque coupling, and simulate the special case of cork-screwing. The model is based on classical bearing capacity formulas which we use in the context of dynamic simulation. We use data from industry experts as well as the results of DEM simulation to validate the qualitative and quantitative accuracy of the real-time model.

2 Theoretical Background

2.1 Soil-Tool Models

Soil can be modelled as continuum or by using particle representation for the individual grains. Both finite (FEM) and discrete element models (DEM) have limitations as the real soil is neither a continuum nor a collection of particles only. Water content may also play an important role as void ratio and pore pressure have an effect on the effective stress. The simulation of these models are also very computationally intensive. The Earthworks module in Vortex therefore uses analytical formulas to estimate the cutting force resisting the motion of the tools interacting with the soil, while for the representation of the disturbed soil in motion a particle based model was developed. The cutting force estimation is based on the fundamental earth-moving equation (FEE) [2]. In situations, where the visual representation of the cut and the effect of the transported soil is not important, applying purely analytical models could be sufficient. This is the case when a drill rig is creating holes for foundation piles at a relatively slow penetration rate.

The model behind the FEE equation is two dimensional, i.e., the cutting tool is considered to have an infinite width. Another common assumption is that the soil fails along a straight line starting at the tip of the tool and reaching the soil surface at a distance ahead of the tool. The tool is a flat blade and the slope of the failure line depends on the internal friction angle of the soil. The soil resistance is approximated based on the quasi-static equilibrium of the soil wedge formed in front of the blade. By assuming passive soil failure, Rankine's passive earth pressure theory and the Mohr-Coulomb failure criterion can give an approximate value for the soil wedge angle which is valid for frictionless, vertical blades with a flat soil surface [3]. The soil wedge angle can also be estimated by minimizing the weight related N-factor in the FEE [4]. In Luengo's work [5], the soil-tool interaction was studied in an automated excavation scenario, where the FEE was modified to also include the effect of the slope of the terrain.

Similar models were developed to assess the bearing capacity of foundations. For shallow strip foundations Therzaghi's bearing capacity formula [6] can be used. To account for deviations from shallow strip foundations, several different bearing capacity and correction factors exist [7, 8, 9, 10]. Empirical shape factors are used to consider the actual shape (e.g., circular cross-subsection) of the pile. Several authors proposed similar formulas by considering installation depth and inclination. In soil augering and robotic drilling the soil/material moved by the flights of the auger and transported to the surface has an important effect on the resistance felt by the tool. In case of augering the flights are typically full of compacted soil and the soil-tool interaction is simultaneous on all the surfaces of the auger and the borehole. In case of drilling, separate particles are being transported, the motion of which also depend on the centrifugal effects at higher rotation speeds [11]. A special operation which can happen in case of soil augering is called cork-screwing. In this case the downward velocity is such that the penetration in one turn is equal to the pitch and the uplift velocity of the soil is zero. The drill screws into the soil as the worm of a corkscrew.

Compared to the cutting force resistance model, the estimation of the ground reaction for piles and drill bits are more complicated in the sense that the reaction force/torque has more components, and there is coupling between the reaction force and torque. When rotary displacement piles are pulled into the ground by the drill rig, the cylindrical surface of the pile comes into contact with the soil, and, in addition to the tip resistance, a depth dependent and distributed friction force develops along the embedded length of the pile. As the pile rotates, the direction of the resultant friction force changes with the instan-

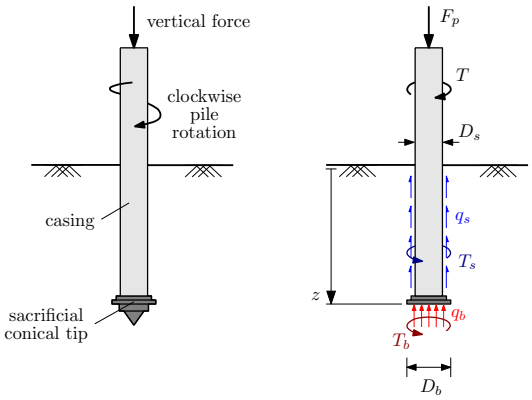


Figure 2. Loading and reactions of circular cross subsection pile

taneous pitch of the pile. At high rotational speeds with slow penetration this generates a large resistance torque, whereas at a low rotation rate the vertical component of the friction force will dominate. This phenomenon is used to lower the skin resistance below certain depths to reduce the necessary piling force.

2.2 Ground Resistance Force

When a pile is already installed in the ground, two main resistance force components can be distinguished. These are the base or tip resistance and the skin resistance. The base resistance is a localized property whereas the skin resistance, which depend on the embedded surface area of the pile, depend on the composition of the soil layers.

Due to the complexity of the problem, piles are often designed using empirical methods. This is based on on-site measurements, i.e., the bearing capacity of the pile is correlated with cone penetrometer test data. Another approach for estimating the resistance is to use analytical models that are based on certain failure criteria of the soil. An advantage of these analytical models is that they depend on a relatively limited number of parameters and can provide a tunable model for simulation. In case of the ground model developed for the IHC project the latter approach was followed.

Pile design formulas are valid under the assumption that the pile is advancing slowly and a quasi-static model is applicable. The formulas presented below are based on this assumption. Additional kinematic considerations were necessary to model the resistance of the different piles. Under slow rotation and penetration, the main ground reaction components are presented in Fig. 2, where F_p and T are the applied piling force and torque, q_b and q_s are the distributed base and skin reaction forces, and T_b and T_s are the resultant base and skin resistant torques.

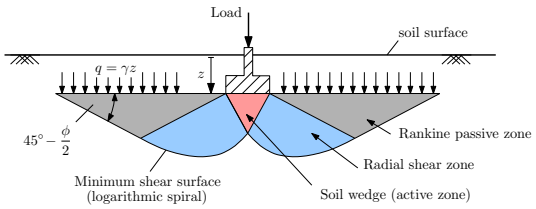


Figure 3. Assumed soil failure in case of shallow strip foundations

2.2.1 Base resistance

The resistance of the soil at the tip of the pile is often calculated by using Therzagi's bearing capacity formula [6]. This is valid for strip foundations at a shallow depth, but empirical shape factors are used to consider the actual shape (e.g., circular cross-subsection) of the pile. Terzaghi's assumption for shallow strip foundations is shown in Fig. 3. In this figure, z denotes the depth of the tip of the pile, γ is the specific weight of the soil, and ϕ its internal friction angle. In addition, q denotes the overburden pressure due to the weight of the soil. There are three distinct zones identified, and the soil is assumed to fail (be able to slide) along the edges of these zones. The bearing capacity of the soil can be calculated by analyzing the static equilibrium of the soil wedge shown below the pile.

The ultimate bearing capacity of foundations has the general form [7]:

$$q_b = c\, d_c S_c N_c + \gamma z\, d_q S_q N_q + 0.5 \gamma d_\gamma\, S_\gamma N_\gamma \quad (1)$$

where c, $q = \gamma z$ and γ are the cohesion, overburden pressure and specific weight of the soil, and N_c, N_q and N_γ are the corresponding bearing capacity factors. These factors have an intricate dependence on the internal friction angle of the soil. In addition, d_* and S_* denote different depth and shape factors that are used to adjust the contribution of the additive terms. The resulting base resistance pressure q_b can be used to estimate the resistance of the ground during piling as the force $Q_b = q_b A_b$ represent the limit force (bearing capacity) at which the pile should sink into the ground, where A_b is the area of the base of the pile.

According to Terzaghi [6], the bearing capacity factors are

$$\begin{aligned} N_q &= \frac{e^{2(3\pi/4-\phi/2)} \tan(\phi)}{2\cos(\pi/4+\phi/2)^2} \\ N_c &= \frac{N_q - 1}{\tan(\phi)} \\ N_\gamma &= \frac{\tan(\phi)}{\gamma}\left(\frac{K_{p\gamma}}{\cos(\phi)^2} - 1\right) \end{aligned} \quad (2)$$

where $K_{p\gamma}$ is the passive soil pressure coefficient usually given in the form of tabulated data in the function of the internal friction angle. A simpler form of the last bearing

capacity factor above was proposed by Coduto [12]:

$$N_\gamma = \frac{2(N_q + 1)\tan(\phi)}{1 + 0.4\sin(4\phi)} \tag{3}$$

The shape and depth factors in Eq. (1) for circular footing are

$$S_q = 1, \quad S_c = 1.3, \quad S_\gamma = 0.6 \quad \text{and} \quad d_q = d_c = d_\gamma = 1 \tag{4}$$

Considering deep foundations, and therefore different soil failure lines, Meyerhof proposed different bearing capacity factors [7] :

$$\begin{aligned} N_q &= \mathrm{e}^{\pi \tan(\phi)} \tan(\pi/4 + \phi/2)^2 \\ N_c &= \frac{N_q - 1}{\tan(\phi)} \\ N_\gamma &= (N_q - 1)\tan(1.4\phi) \end{aligned} \tag{5}$$

where the shape and depth factors are:

$$\begin{aligned} S_q &= 1 + 0.1K_p, \\ S_c &= 1 + 0.2K_p, \\ S_\gamma &= S_q \end{aligned} \tag{6}$$

with

$$\begin{aligned} K_p &= \tan(\pi/4 + \phi/2)^2 \\ d_q &= 1 + 0.1\sqrt{K_p}, \quad d_c = 1 + 0.2\sqrt{K_p}, \quad d_\gamma = d_q \end{aligned} \tag{7}$$

A more recent study by van Baars [10] proposes another expressions for the shape factors. These are:

$$\begin{aligned} S_q &= 1 + 1.5\sin(\phi) + 3\tan(\phi)^3 \,, \\ S_c &= S_q + (0.2 - 0.1\tan(\phi)) \,, \\ S_\gamma &= S_q \end{aligned} \tag{8}$$

It can be seen that the bearing capacity factors N_c and N_γ are typically expressed with the help of N_q. Therefore, in Fig. 4 the tip resistances obtained by different authors are presented by comparing this bearing capacity factor. By noticing the logarithmic scale of the vertical axis, it is clear that there are very significant difference between the models that are available in the literature, but at the same time, they are qualitatively similar.

During pile driving, the base resistance provided by these formulas can significantly drop due to the disturbance of the soil when the pile rotates. On the other hand, when the penetration speed is suddenly increased, the soil might temporarily have a higher resistance as the stresses around the tip of the pile does not have time to relax.

Assuming a constant base resistance, and a corresponding constant linear shear stress distribution under the tip of the pile, the base resistance force and torque can be expressed as:

$$Q_b = \frac{D_b \pi^2}{4} q_b \quad \text{and} \tag{9}$$

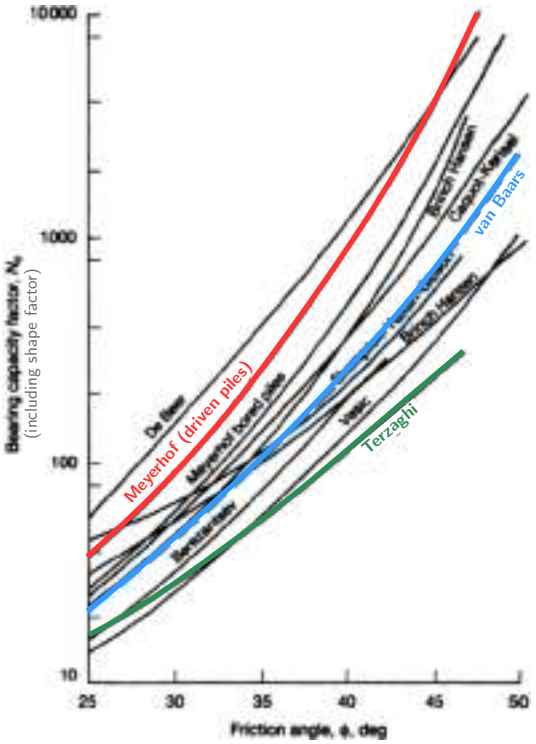


Figure 4. Different bearing capacity factors [10]

$$T_b = \int_0^{D_b/2} (\tan(\delta)\, q_b\, 2r\pi\, r)\ \mathrm{d}r = \tan(\delta)\, Q_b \frac{D_b}{3} \tag{10}$$

where δ is the friction angle between the soil and the pile's material and D_b is the diameter of the sacrificial tip.

2.2.2 Skin resistance

The skin resistance of piles is a frictional resistance force which depends on the pressure exerted by the soil on the surface of the pile. This lateral earth pressure is proportional to the overburden pressure, and changes significantly with the state of the soil. The active surface area also depend on embedded depth of the pile, so the depth dependence of the skin resistance is quadratic. When this resistance is calculated the composition of the soil layers has to be considered too. The total skin resistance force is the force which is acting on the complete embedded surface of the pile.

The lateral soil pressure is proportional to the overburden pressure. There are different theories to estimate the ratio between these. In case when the soil is undisturbed, the so-called in-situ soil pressure coefficient, K_0, can be calculated by using Jaky's simplified formula [13]. When the soil can expand and the stresses can relax, it is said to be active and the lateral pressure, K_a, is smaller than the in-situ pressure. When the soil is compressed (e.g.,

displaced by the pile) it is said to be passive and the lateral pressure, K_p, is significantly higher. The active and passive soil pressures are often estimated based on Rankine's theory. For cohesionless soils:

$$K_0 = 1-\sin(\phi), \quad K_a = \frac{1-\sin(\phi)}{1+\sin(\phi)}, \quad K_p = 1/K_a \quad (11)$$

In the pile design literature, various different formulas and tabulated experimental results exist for determining the lateral earth pressure coefficient, K, that can be used for the skin resistance calculation based on the piling technique and the specific soil type. Empirical factors are necessary to accompany the pure soil mechanics formulations to consider all the effects that can change the bearing capacity of a pile after it is inserted [14]. Recommendations for choosing the proper earth pressure coefficient, including averaging the formulae in Eq. (11), were proposed by Wrana [15]. For simulating the insertion process of a driven displacement pile in an operator training scenario, a good default value could be $K = K_p$ and introducing an additional scaling factor could allow the user to fine tune the soil behaviour. With these consideration the skin resistance force can be calculated as the integral:

$$Q_s(z) = \int_0^z (\tan(\delta) K q D_s \pi)\ \mathrm{d}\zeta \quad (12)$$

where the overburden pressure q, the lateral earth pressure coefficient K, and the friction coefficient $\tan(\delta)$ depend on the soil depth and the composition of soil layers. With the assumption of having constant soil properties per layer, this integral can be pre-computed in closed form for the real time simulation, or a simple 1 dimensional lookup table (with linear interpolation) can be used to quickly obtain the skin resistance force at specific depths. In order to incorporate the effect of pile rotation on the skin resistance, the instantaneous pitch of a rotary displacement pile needs to be considered. As the pile rotates, a point on it's surface moves relative to the ground with the sum of the tangential and axial velocities. The skin friction opposes this resultant velocity. The decomposition of the friction force into a radial and an axial component helps understanding, and captures the important effect, that in practice the rotation of a pile makes its insertion easier at the expense of increasing the overall resistance torque. The reduced skin resistance (the vertical component of the friction force) and the resistance torque which increases with the rotation speed can be computed in the form:

$$Q_{sv} = Q_s \frac{v_a + v_{min}}{v + v_{min}} \quad \text{with} \quad v = \sqrt{v_a^2 + v_t^2} \quad (13)$$

$$T_s = R_s Q_s \frac{v_t + v_{min}}{v + v_{min}} \quad (14)$$

where v_a is the axial velocity, $v_t = R_s\omega$ is the tangential velocity with radius R_s, and ω is the angular velocity of the pile. The minimum velocity v_{min} parameter was introduced to handle the special cases during simulation when one or both of the velocity components are zero. In case of the latter, when the pile is not moving, $Q_{sv} = Q_s$ and $T_s = R_s Q_s$ represent the maximum force and torque bounds of the constraints modelling the reaction torque and force, respectively.

3 Real Time Soil Augering Model

In our real time simulation model, we consider 4 different generalized reaction forces. These are the extraction and penetration resistance forces, the resistance torque and the pull-down force exerted by the auger on the piling equipment. The main characteristics of these are captured by simple models that provide an efficient formulation and qualitatively accurate results. During extraction, it is assumed that the auger does not rotate, and the weight of the pile is calculated considering the flights of the auger are completely filled with soil. In this case the extraction resistance force can simply be calculated by using the skin resistance force derived in the previous section plus the weight of the soil. For the skin resistance computation, one can simply use the borehole diameter as the skin diameter (D_s) and replace the external (soil-pile) friction angle (δ) with the internal friction angle (ϕ) of the soil.

For estimating the penetration resistance force, three different operation regimes can be considered. First, when the auger is rotating relatively fast and it is transporting soil to the surface. Second, when the instantaneous pitch is equal to pitch of the helix of the auger, and the auger is corkscrewing itself into the ground with a rapidly increasing resistance torque. And third, when the linear downward motion dominates and the auger displaces soil similar to a displacement pile. In this third case, the estimation of the penetration resistance is relatively simple as it ends up being the skin resistance calculated with the borehole diameter and the internal friction angle of the soil. When soil is transported or corkscrewing happens further considerations are necessary.

The helix of the auger can be seen as a long wedge along which the total mass of the soil is pushed upwards. The uplifting force is the resultant of force transmitted from the helix to the soil minus the resistance force due to friction with the wall of the borehole. In order to estimate these forces, the kinematic model in Figure 5 is considered. This shows that the $\vec{v}_{\text{up}}$ uplifting transport velocity of the soil is the sum of the auger's downward velocity $v_z\,\hat{k}$, the lateral velocity $\vec{v}_t = r\omega\,\hat{t}$ and the relative soil velocity with respect to the auger $\vec{v}_{\text{rel}}$. The magnitude of the uplifting velocity ranges from zero to $v_{\text{up,max}} = v_t\tan(\psi) - v_a$ where $v_a = -v_z$ is the penetration velocity. In the following, this maximum velocity is used to express the uplifting velocity $v_{\text{up}} = \eta\, v_{\text{up,max}}$ where η is an efficiency parameter. It is left to the user to define this as a constant, or feed back

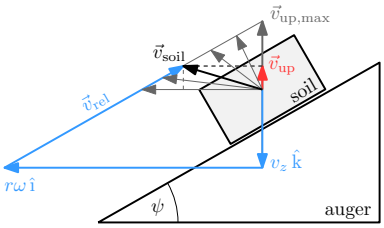


Figure 5. Model of soil uplift velocity of the auger

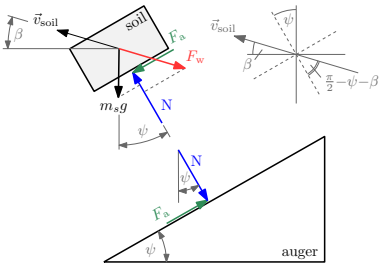


Figure 6. Free body diagram of the augering model

its value to the model using complementary calculations. With these considerations, the penetration resistance (Q_a) and the pull-down force (V_a) of the aguer can be expressed by using the free body diagrams shown in Figure 6 as

$$Q_a = F_a \sin(\psi) \tag{15}$$

$$V_a = N \cos(\psi) \tag{16}$$

with

$$N = m_s g \cos(\psi) + F_w \sin(\psi + \beta) \tag{17}$$

$$F_a = \tan(\delta) N \tag{18}$$

where N and F_a are the normal and lateral components of the reaction force between the soil and the auger. The total mass of the soil is m_s, g is the gravitational acceleration, and F_w is the total skin resistance force transmitted from borehole to the soil mass in the auger. Moreover, using the outer radius r of the auger, an upper bound on the augering resistance torque can be given in the form:

$$T_s = (F_a \cos(\psi) + N \sin(\psi))\, r \tag{19}$$

In the special case when the maximum uplifting velocity $v_{\text{up,max}}$ is zero, the auger is corkscrewing. This condition is identical to saying that the instantaneous pitch $v_a/v_t = \tan\psi = p/(2\pi)$ with p being the pitch (height difference) between the flights of the auger. When this happens the auger is pulling itself into the ground as a corkscrew into the cork, and the reaction force and torque increase until the machine limits are reached and the auger cannot be moved further into the soil. This dangerous situation must be avoided, and a trained operator needs to change the penetration and rotation speed in order to prevent this from happening. For such a training, we use a model, where a kinematic constraint is used to lock

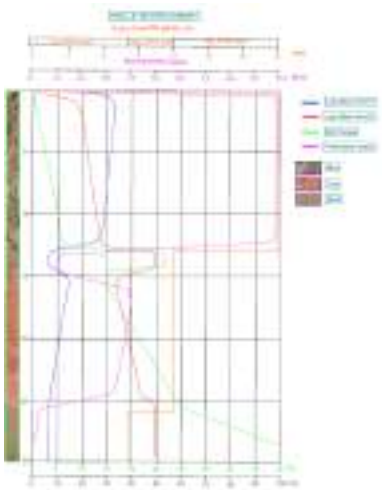


Figure 7. IHC reference data for displacement drill

the auger into the corkscrewing regime, with rapidly increasing reaction force and torque, when the instantaneous pitch is sufficiently close to geometric pitch (expressed in m/rad) of the auger. The maximum force and torque of this screw constraint is proportional to the corkscrewed length of the auger. This way, when the user sufficiently unscrews the auger the applied torque and force can overcome the constraint limits, and the auger can break out of corkscrewing.

4 Model Validation

The IHC FUNDEX F3500 multifunctional foundation rig was modelled and simulated. The main objective was to develop real-time, parametric and physically motivated ground resistance models for three different pile driving operations. These include the displacement drill, the continuous flight auger (CFA) and the hydrohammer piling techniques. In each case, the pile is driven into the ground by the drill rig and it penetrates through different layers of soil with different resistances. As the programmable logic controller (PLC) of the simulated drill rig is integrated into the system, it expects to receive physically realistic resistances. Therefore, selection and tuning of soil properties was an important element of the work.

For the validation of the model, IHC industry experts provided us with desired drill report curves. Figure 7 shows an example of the provided report for the expected soil resistance during displacement drilling. The desired behaviour was determined based on real drill reports by capturing the main characteristics of experimentally recorded data. The report shows vertical force and torque applied to the tool, as well as the penetration and angular velocities of the tool. There are three distinct soil layers that the tool passes through as depth increases: mud, clay, and sand.Capturing the effect of layer transitions is

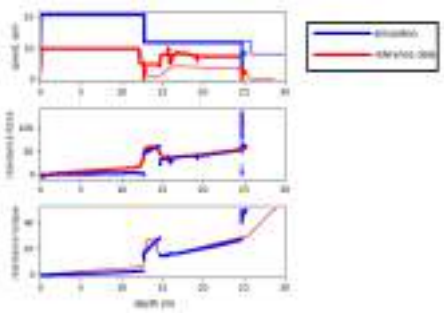


Figure 8. Machine in the loop simulation results

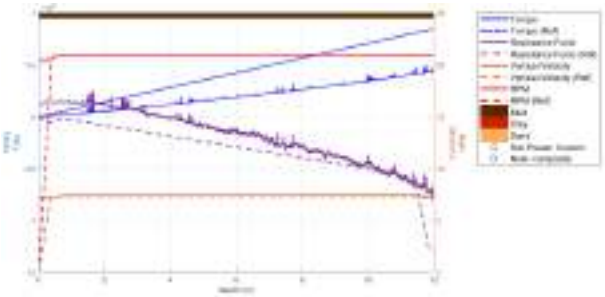


Figure 9. DEM Simulation results - CFA

an important aspect of the simulation. The reason we did not directly use some recorded drill reports is that in the simulated scenario there were only four layers with very different soil properties. The four layers and the composition of the layers can be chosen by the instructor of the training simulator according to the pedagogical needs.

In Fig. 8 the drill report of a displacement drill operation is shown. In the top panel the thin lines represent the planned penetration rate (speed) and rotation speed (rpm), and the thicker lines show the actual values commanded by the operator of the simulator. These are plotted as function of the penetration of the tip of the pile (not wrt. time). The middle panel depicts the observed (blue) resistance force vs. the expected one (red). Similarly, the comparison of the resistance torques can be seen in the bottom panel. Although, the penetration speed was generally slower than the desired value, the soil resistance matches the expected trends and values quite closely. Note, the equations used to estimate the ground reaction assume quasi-static conditions, therefore they cannot directly capture the effect of different penetration speeds.

5 DEM Model

Because DEM has a well documented ability to produce high-fidelity soil models [16], a DEM simulations was set up to show that the purely analytical model, as well as the reference data, was qualitatively in agreement with a DEM approach. In this section, we present the comparison for the CFA simulation using DEM.

The DEM model of the drill rig was built in the real time multibody simulation platform Vortex Studio [1] whose particle system is based on the Parallel Particles approach to DEM modelling proposed by Holz [17], based on the principles of position-based dynamics presented by Muller [18]. This method has previously been used for DEM soil simulations of wheels travelling over soil [19] [20] and blades excavating soil [20]. Due to current software limitations, the DEM simulations only consider the behaviour of the drill for the first 12 meters depth, at which point the reference data indicates a transition from the mud soil layer to clay. This is because the current version of this DEM implementation does not make it possible to model layers of soil with different properties. However, conclusions can still be drawn from the simulation of the first soil layer, which accounts for about a third of the total depth included in the IHC reference data.

Figure 9 shows a comparison between the reference data that was used in the previous section and the DEM simulation results for the continuous flight auger. Extensive tuning would be required in order to achieve a perfect quantitative match with the DEM simulation [19], but the simulation results with baseline soft soil parameters already showed good qualitative agreement. Without tuning of parameters, we are looking for the DEM model to replicate the overall trend of the outputs as depth increases, as well as the order of magnitude of the outputs. The doted lines on the plot show the reference values for torque, resistance force, penetration speed and RPM; while the solid lines show the results of the DEM simulation. Unlike the analytical model, this DEM model is a fully dynamic rigid body simulation coupled with a particles to model the soil. Velocity level constraints control the motion of the drill, allowing the simulation to exactly match the desired motion of the drill - illustrated in red and orange for the RPM and vertical velocity, respectively. In the interest of simulation time, the particles used in the simulation were rather coarse, which explains the noisy nature of the resistance force and torque outputs. While the order of magnitude and overall trend with respect to depth of the force and torque plots are consistent with the reference data, the inaccuracy of the exact numerical value of these outputs can be attributed to a lack of parameter tuning.

6 Conclusion

This paper proposes an analytical method for modelling piling operations for real time dynamic simulation which is accurate and efficient enough for the purposes of operator training. The model uses classical bearing capacity equations to compute the force components acting on the pile during insertion. To capture the reduction of the vertical resistance force when the pile is rotated, a simple

model was developed to decompose the overall friction force based on the instantaneous pitch. This is essential for building a realistic scenario for operator training when a rotary displacement pile is used.

Another important contribution of the paper is the simplified augering model that was presented. This model simulates the vertical pull down force of an auger reasonably well. The selected semi-analytical simulation approach was necessary to achieve the simulation speed and accuracy required for successful operator training. It enables real time training with PLC in the loop. This model is dependent on proper tuning of an efficiency parameter for the uplifting force that is left to the user to define, either as a constant value or through some complementary calculations. Setting this value as a constant and tuning it, the qualitative agreement between the presented models is realistic enough for human-in-the-loop drill operator training simulators. This claim was validated by comparing the analytical results to reference data provided by industry experts - data that was in turn qualitatively compared to a dynamic discrete element soil model for the first layer of soil.

References

[1] CM Labs Simulations. Vortex studio simulation platform, 2018. URL `https://www.cm-labs.com/vortex-studio/`.

[2] A.R. Reece. Paper 2: The fundamental equation of earth-moving mechanics. Proceedings of the Institution of Mechanical Engineers, 179(6):16–22, June 1, 1964.

[3] V.N.S Murthy. Geotechnical engineering: Principles and practices of soil mechanics and foundation engineering. CRC Press, Boca Raton, FL, 2002.

[4] Fiorenzo Malaguti. Soil machine interaction in digging and earthmoving automation. 12 1994. doi:10.1016/B978-0-444-82044-0.50029-1.

[5] O. Luengo. Modeling and identification of soil-tool interaction in automated excavation. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robotic Systems, page 7 pages, Victoria, BC, Canada, October 13-17, 1998.

[6] K. Terzaghi. Theoretical Soil Mechanics. Wiley, New York, 1943.

[7] G.G. Meyerhof. The ultimate bearing capacity of fundations. Geothecnique, 2(4):301–331, 1951.

[8] G.G. Meyerhof. Some recent research on the bearing capacity of foundations. Canadian Geotechnical Journal, 1(1):16–26, 1963.

[9] J.B. Hansen. A revised and extended formula for bearing capacity. In Bulletin No. 28, pages 5–11. The Danish Geotechnical Insititue, Copenhagen, 1970.

[10] S. van Baars. The inclination and shape factors for the bearing capacity of footings. Soils and Foundations, 54(5):985–992, 2014.

[11] Quan Qiquan, Tang Junyue, Yuan Fengpei, Jiang Shengyuan, and Deng Zongquan. Drilling load modeling and validation based on the filling rate of auger flute in planetary sampling. Chinese Journal of Aeronautics, 30(1):434–446, 2017.

[12] D. Coduto. Foundation Design. Prentice-Hall, 2001.

[13] J. Jaky. Pressure in silos. In Proceedings of the second international conference on soil mechanics and foundation engineering, pages 103–107, Rotterdam, June 21–30, 1948.

[14] Michael Tomlinson and John Woodward. Pile Design and Construction Practice. CRC Press, Boca Raton, FL, 6 edition, 2015.

[15] Bogumil Wrana. Pile load capacity – calculation methods. Studia Geotechnica et Mechanica, 37(4), 2015.

[16] William Smith, Daniel Melanz, Carmine Senatore, Karl Iagnemma, and Huei Peng. Comparison of discrete element method and traditional modeling methods for steady-state wheel-terrain interaction of small vehicles. Journal of Terramechanics, 56:61–75, 2014.

[17] D. Holz. Parallel particles (p2): A parallel position based approach for fast and stable simulation of granular materials. Workshop on Virtual Reality Interaction and Physical Simulation VRIPHYS, 2014.

[18] M. Muller, B. Heidelberger, M. Hennix, and J. Ratcliff. Position based dynamics. Journal of Engineering Mechanics, 2006.

[19] Eric Karpman, Jozsef Kovecses, Daniel Holz, and Krzysztof Skonieczny. Discrete element modelling for wheel-soil interaction and the analysis of the effect of gravity. Journal of Terramechanics, 91:139–153, 2020.

[20] Eric Karpman, Jozsef Kovecses, and Marek Teichmann. Terramechanics models augmented by machine learning representations. Journal of Terramechanics, 107:75–89, 2023.

Framework for Acoustic Comfort Analysis in Digital Twins

Hossein Pouresmaeeliasiabar[1] and Ali Motamedi[2]

[1] École de technologie supérieure, Université du Québec, Canada
[2] École de technologie supérieure, Université du Québec, Canada
hossein.pouresmaeeliasiabar.1@ens.etsmtl.ca, ali.motamedi@etsmtl.ca

Abstract –

Acoustic comfort, a critical yet often overlooked aspect of indoor environmental quality, plays a significant role in occupant health and productivity. Unlike other comfort dimensions, such as thermal or lighting, measuring acoustic comfort remains challenging due to its subjective nature and the complex interplay of physiological and psychological factors. Current approaches to assessing acoustic comfort in indoor environments often overlook the content of sound, despite its potential to be a decisive factor. For instance, the perceived comfort of listening to music at high sound levels differs significantly from that of hearing construction noise, even at lower sound levels. This research proposes a novel framework for integrating acoustic comfort analysis in the digital twin environment, which comprises psycho-acoustic metrics, sound event classification, and predictive analytics. The implemented system leverages sensor data, a sound event classification neural network, and advanced visualization methods to enable real-time and historical acoustic analysis. Privacy concerns are addressed through a privacy-by-design approach, ensuring data security by processing audio on the edge devices without storing raw sound. A case study in an office environment demonstrates the framework's effectiveness in monitoring and improving acoustic conditions. Microphones connected to edge devices classify sound events and calculate soundwave parameters such as relative sound pressure levels while integrating results into the digital twin.

Keywords –Digital Twins; Acoustic comfort; Psycho-acoustic metrics; Sound event classification; Acoustic monitoring; Comfort analysis

1 Introduction

There exist numerous definitions for the term “comfort” in various domains such as AEC-FM (architecture, engineering, construction, and facilities management) and healthcare. The common theme in the definitions of comfort in the built environment refers to it as a state of physical and psychological ease, characterized by the absence of discomfort or distress. Although this high-level definition understandably describes comfort, it fails to adequately pinpoint its defining measurable characteristics. The notion of comfort in buildings can be viewed in multiple distinct but interrelated aspects, such as thermal comfort, acoustic comfort, indoor air quality, and lighting comfort; each of which depends on various variables. However, the level of perceived comfort differs for each person, depending on the physiological and psychological state at that very moment in time. Maintaining comfort is a fundamental need in buildings; the indoor environment quality (IEQ) performance of a building, is a decisive factor in the health, productivity, and well-being of its occupants, and can be a determinant factor in lifecycle costs and energy consumption [1].

Among the different aspects of comfort, acoustic comfort received less attention [2]. A simple search on academic databases related to indoor comfort reveals a substantial disparity in the number of studies focusing on different aspects of comfort, with significantly fewer studies focusing on acoustic comfort compared to other aspects, such as lighting, thermal, and air quality. Acoustic comfort is generally defined as “a state of contentment with acoustic conditions” [3], and it can play a pivotal role in determining the overall comfort of an environment. Furthermore, it is interrelated to other aspects of comfort and the building's overall energy consumption. For instance, occupants may face the trade-off between the generated HVAC noise by the cooling system and thermal discomfort [4].

Similar to other aspects of comfort, measuring acoustic comfort is complex. Sound perception is subjective and varies among individuals based on their experiences, cultural backgrounds, and personal preferences [5]. Additionally, it highly depends on the psychological state of the occupants, which makes its objective assessment more difficult or impossible. Although the physical characteristics of sound waves can be adequately measured, there is no measurable characteristic showing the overall psychological state of a human being and how comfortable a person is while being exposed to acoustic stimuli. As a result, a certain

sound pattern may seem uncomfortable or annoying to a person and can be ignored or even considered pleasant by another person [6].

To estimate the acoustic comfort level of a space, traditionally, some of the basic variables of the acoustic waveform are measured. Such variables include the sound pressure levels (dB), the frequency (Hz), and the time that a person is exposed to the sound. Although these variables are essential for estimating acoustic comfort, and some regulations and guidelines are already developed recommending their acceptable ranges, "acoustic comfort" is by far more complex to measure. For example, the human hearing system does not respond to all the frequencies, and the sound pressure levels equally [7]; hence, the sounds with some frequencies may be perceived as louder than others. This shows the inadequacy of assessing acoustic comfort using only the sound (acoustic) pressure level.

Consequently, psycho-acoustic indicators such as loudness, A-weighted sound pressure level, sharpness, roughness, fluctuation strength, articulation index, and impulsiveness were introduced[8, 9]. These psycho-acoustic indicators consider the human hearing system in perceiving sounds, resulting in a better estimation of acoustic comfort. For example, to have a better understanding of "how loud a sound is", loudness is defined which is the property of the sound that can be "ordered on a scale extending from soft to loud" [10].

However, even the existing psycho-acoustic indicators would not be sufficient to adequately assess the acoustic comfort for a specific space. For example, although there might be a similarity between calculated psycho-acoustic indicators of audio patterns of certain musical pieces and construction tools (e.g., jackhammer), the perceived sensation for the occupants would be completely different. The feeling of acoustic comfort for occupants who are exposed to a musical piece heavily depends on their musical taste, whereas the noise of the construction tool can be unpleasant for all occupants.

The inadequacy of such metrics and indicators can be linked to their exclusive reliance on sound wave parameters, without accounting for the content or type of the sound. Consequently, researchers have sought to develop new psychoacoustic indicators and sound perception descriptors, such as satisfaction, dissatisfaction, noisiness, and pleasantness, as highlighted by Hossain et al. [11]. The recent advancement of the Deep-learning-based models allows a near real-time analysis and classification of sound events. Combining the classic metrics, psycho-acoustic indicators, and sound event categorization provides better monitoring and estimation of acoustic comfort for the built environment. Additionally, such a combinatory effect allows for defining customized indications of the comfort level considering individual occupants' preferences.

This research aims to provide a framework to perform acoustic comfort analysis by combining various metrics and indicators together with sound event category identification within a digital twin. The proposed framework comprises the integration of sensor data within a digital twin solution, the calculation of psycho-acoustic indicators, the near-real-time identification of sound event categories, and the predictive analytics capability. Additionally, the framework adopts the privacy-by-design approach in which the stored data cannot be used to recreate the original captured sound.

The remainder of this paper presents a brief review of the literature on acoustic environmental assessment, the application of psychoacoustic indicators, and the role of Sound Event Detection (SED) in such evaluations. This is followed by a detailed explanation of the proposed framework and its potential applications. A case study implementation is then described, along with the corresponding results and conclusions.

2 Related Work

Literature on assessing the acoustic suitability of environments reveals two interrelated yet distinct concepts: soundscape and acoustic comfort. While both address the auditory experience, their focus and applications differ. Acoustic comfort primarily pertains to enclosed spaces, emphasizing the reduction of annoyance and enhancement of auditory satisfaction within confined environments. In contrast, the concept of soundscape is broader, encompassing urban, natural, and built environments. It explores the dynamic interaction between environmental sounds and human experiences.

Both soundscape and acoustic comfort assessments rely on a combination of objective and subjective parameters. For example, Aletta et al. [12] emphasize the role of psychoacoustic indicators, such as loudness, sharpness, roughness, and fluctuation strength, in soundscape studies. Similarly, Engel et al. [13] conducted a systematic review of 46 peer-reviewed studies, demonstrating the widespread application of psychoacoustic indicators in soundscape research. These findings highlight the critical role of such indicators in capturing human perceptions of acoustic environments, thereby enhancing the accuracy and relevance of soundscape evaluations. Likewise, psychoacoustic indicators are integral to assessing acoustic comfort. For instance, the relationship between objective acoustic measurements and subjective responses to evaluate acoustic comfort in residential settings is investigated in [14]. Their study underscores the importance of indicators addressing low-frequency noise to better correlate technical data with perceived annoyance, thereby emphasizing the need to integrate human

perception into acoustic comfort assessments.

A review of the literature reveals the extensive use of psychoacoustic indicators in assessing acoustic environments. Zhang et al. [6] developed a high-precision model for predicting acoustic comfort in electric buses by integrating psychoacoustic indicators with machine learning techniques. Similarly, Herranz-Pascual [15] introduced the Acoustic Comfort Assessment Scale (ACAS-12), a psychometrically validated tool designed to assess acoustic comfort in urban environments. This scale incorporates indicators, such as pleasantness, eventfulness, familiarity, informational capacity, and congruence, demonstrating its reliability and effectiveness in capturing subjective perceptions of acoustic environments. Additionally, Kang et al. [16] proposed a framework for creating soundscape maps in smart cities. Their model predicts perceptual attributes, such as pleasantness and calmness, using sound profiling and linear regression analysis, offering a valuable tool for enhancing urban soundscape design.

Advancements in artificial intelligence have enabled the application of SED in soundscape assessment. For example, Espejo et al. [17] explored the use of short-time acoustic indices in combination with artificial neural networks (ANNs) to monitor and analyze soundscapes in urban-natural environments. Similarly, Bonet-Solà et al. [5] developed a predictive model for assessing acoustic comfort in urban settings. Their approach integrates SED using convolutional neural networks with noise data collected from wireless acoustic sensor networks (WASN), offering a sophisticated method for evaluating and improving urban acoustic environments.

To effectively visualize the acoustic status of a large environment, a 2D color-coded heat map based on soundscape evaluation is a practical approach, as demonstrated by Yue et al. [18]. They developed a visual soundscape prediction model for urban park design, which integrates sound pressure levels, sound source perception, and soundscape evaluation using machine learning techniques. This model provides an intuitive representation of acoustic conditions, facilitating the design and optimization of urban soundscapes.

3 The Proposed Framework

Potential Applications and Use Cases

Having a framework that allows monitoring the present and past values for a collection of psycho-acoustic indicators as well as sound event categories for each room of an indoor environment, facilitates the comfort analysis, prediction, and visualizations through the digital twin web interface. It allows facility managers to better plan for acoustic insulation and activity planning. The occupants can consult the monitoring data to decide if a certain room is suitable in terms of acoustic comfort for their intended use. The stored data allows the application of predictive algorithms to recommend rooms for certain activity types or to predict the acoustic comfort of a certain room in a selected time period in the future. The integrated assessment algorithms can be further enhanced by including other sources of data, such as activity schedules for each room.

The data can be used to identify sources of unwanted noise (such as background noise or noise pollution) and their effect on the neighboring rooms by cross-checking metrics of various neighboring spaces. Additionally, the system can accommodate user inputs to provide personalized metrics and Key Performance Indicators (KPIs) for individuals. Moreover, the system can communicate information about the acoustic comfort preference of neighboring occupants to the occupants of a certain room to adjust their noise-generating activities accordingly. Finally, it can give insights and predictions regarding outdoor noise.

Framework

The framework consists of a five-layered architecture, as depicted in Figure 1 (left). Starting from the bottom of the diagram:

1. In the data acquisition layer, sound is captured by microphones and converted to a digital signal via an analog-to-digital converter (ADC). Using the device-specific interface, the audio signal, in a binary stream format, along with the device-specific metadata such as sampling rate, microphone sensitivity, microphone type, and the location coordinates of the microphone, is prepared for subsequent processing. The data acquisition and data processing layers are distinct; however, they can be implemented on a single device, enabling high-speed data transfer through shared memory or inter-process communication (IPC) interface.
2. In the data processing layer, the audio data stream and device specifications from the data acquisition layer are utilized. This layer processes the input signal according to the requirements of its three modules, based on parameters such as the sampling rate and device specifications. The audio signal is resampled and formatted to meet the input requirements of the selected SED model. Additionally, soundwave parameters and psychoacoustic indicators are computed using the audio signal and accompanying metadata. The processed data is then sent to a cloud-based database to be stored as time-series data, utilizing either REST-based or SQL-based communication interfaces. Depending on the storage layer type,

data transfer can be conducted using GET and POST methods in the REST-based approach or directly via SQL queries.

3. The storage layer manages both historical and real-time data, ensuring continuous data availability for further analysis and processing. Notably, the recorded data includes metadata essential for the digital twin, such as microphone location coordinates, device identification, and room designation. This data can also be accessed through REST-based or SQL-based communication interfaces.
4. The calculated parameters and metadata are retrieved from the cloud-based database in the logic layer. Based on the nature of the environment, the defined KPIs, and user- or environment-specific settings, this layer processes the data and prepares it for visualization. The logic layer can be implemented at any host platform, whether on a local or a cloud platform. It uses user inputs and commands gathered in the presentation layer and the data managed in the storage layer to perform various calculations and predictions. Additionally, it can integrate external data sources, such as Building Information Modeling (BIM) models, to allow 3D visualizations as well as access metadata related to the built environment that is used for comfort analysis. Depending on the presentation layer and whether it is remotely connected to this layer or operates on the same device, the processed data can be accessed through various communication interfaces, including REST-based, file-based, or shared memory communication methods.
5. The presentation layer provides users with various visualization tools, including graphs, charts, and 2D or 3D heat maps. It also accepts user-defined settings and, through continuous communication with the logic layer, ensures that the system displays the desired outputs. This layer can be implemented either on an online digital twin platform or a local computer.

Notably, in a real-world scenario, collecting data from multiple rooms requires additional instances of the implemented data acquisition and data processing layers. This scalable approach facilitates the integration of multiple rooms without significant computational constraints. The performance on each edge device may vary depending on the complexity of the Sound Event Detection (SED) model and the computational power of the edge device, but it is not dependent on the number of rooms.

The proposed framework was adopted for the use case of sound event classification in an office building. A sound classification system that is integrated with the digital twin platform is developed, and a case study is performed to assess the applicability of the developed system.

3.1 Case Study

The case study was conducted in an office space within a public building. The room, measuring 106 square meters, has an L-shaped layout with a ceiling height of 2.96 meters. It is typically used for meetings, classes, and various gatherings, but it also occasionally serves as a study or group work area for students.

The objectives of this implementation are threefold:

1. To develop a system capable of measuring a collection of soundwave parameters, classifying sound events, and transmitting the collected data to a cloud-based platform
2. To develop a method for visualizing the collected data effectively
3. To integrate the developed acoustic comfort measurement module into the existing digital twin

The digital twin of the environment is implemented using Autodesk Platform Services. Time-series data is stored in Microsoft Azure Data Explorer, and the existing dashboard provides historical data visualization for parameters such as temperature, air pressure, air quality, and occupancy (Figure 2).

In the case study implementation, a sound event classification neural network was utilized to categorize audio events. YAMNet [19], [20] is a pre-trained deep learning model with 521 output classes. For the case study, the model needs to be customized. For example, the existing classes are grouped into a smaller number of categories, tailored to the case study environment. Additionally, the model is required to be fine-tuned to enhance its accuracy for the given application. For effective grouping, it is important to consider the environment's characteristics (e.g., a wild animal sound is unlikely to occur in an office, despite the existence of such a class in YAMNet sound categories). Based on the types of activities expected in the case study environment (an office), five main categories were defined (Table 1). Each YAMNet class was then reviewed and assigned to one of these categories.

To capture sound, microphones are installed in the room. To achieve reliable results, selecting the right microphone (or an array of microphones) is one of the key design considerations for the implemented system. For example, the microphones should not perform active noise control (ANC) (i.e., noise cancellation), or modify the sound balance. Additionally, to efficiently perform the measurement and calculation of metrics, the microphone should be precise and sensitive to adequately capture sounds within its range of operation. In the case study, multiple identical microphones were installed at

different points of the room, each is connected to an edge device that analyzes the sound. The selection of microphone types depends on their placement location. In this study, omnidirectional microphones were used, featuring a signal-to-noise ratio (SNR) of 80 dB and a maximum sound pressure level of 110 dB.

Table 1. The categorization of the possible sounds in the office and the number of corresponding YAMNet output classes

Category	Number of Classes
Speech, Meeting	14
Silence, Ambient Office Sounds	30
Construction and maintenance work, Interruptions	112
Gathering, Crowded, Music	179
Other Classes	186

Installing microphones in the rooms may raise privacy concerns for occupants. As mentioned, a privacy-by-design approach was followed in this implementation, in which the sound classification is performed on the edge devices in real-time using small snippets of the sound signal collected from microphones that are directly attached to them. The actual captured sound is not recorded or transferred over the network. In other words, the data acquisition layer and the data processing layer are implemented in one device, ensuring that the original sound signal remains secure and eliminating the risk of data leakage over data transmission in the network.

The implemented system accommodates the measurement of the relative sound pressure levels as well as the sound event category identification. The implemented system allows the integration of new customized KPIs, considering the preferences of occupants and their intended activity within the space. The results are integrated and visualized in the developed digital twin platform of the facility, which is the primary source of information and a tool for monitoring and performing predictive analytics. The architecture of this implementation is depicted in Figure 1(right).

3.2 Visualization of Sound Content

Although various types of data visualization methods can be easily created based on the data generated in the logic layer, a combination of bar and line charts is the preliminary visualization method (Figure 3). It allows visualizing classifications with color-coded bars overlaid by various line charts that show calculated metrics. As an example, the relative sound pressure level, defined in Equation (1), is illustrated in Figure 3. The graph has the date and time on the horizontal axes, which allows visual analysis of the changes in acoustic metrics and indicators during a certain day.

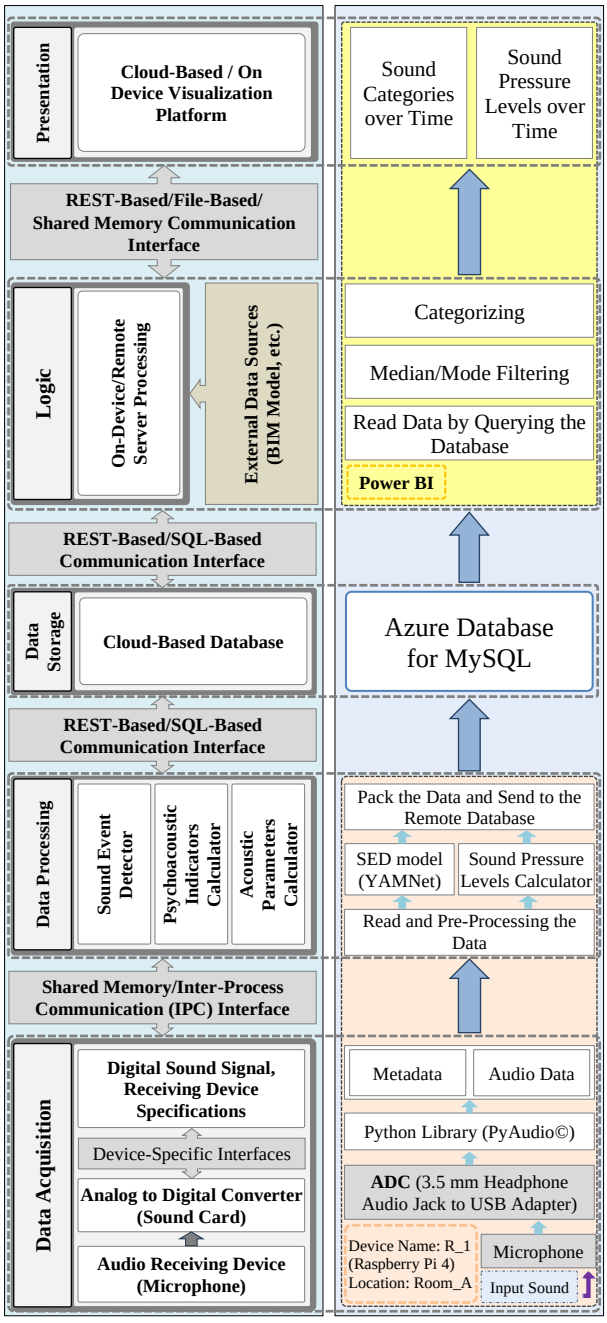


Figure 1. The Proposed Framework Layers (left) and the case study implementation (right)

Equation (1) represents the relative sound pressure level ($SPL_{relative}$). The input to the logarithm function is the Root Mean Square (RMS) value of the microphone's output, obtained using the PyAudio library at a 16 kHz sampling rate, and 16-bit audio depth, continuously

calculated over a defined interval (1-second) within the (-1, 1) range. The output of this equation falls within the (-inf, 0) dB range, and can be a relative measurement of the sound pressure level in an environment. Notably, this value cannot be directly used or compared with established comfort thresholds in the environment unless it is converted to absolute sound pressure levels.

$$SPL_{relative} = 20 * \log(RMS(normalized\ measured\ value)) \quad (1)$$



Figure 2. The existing digital twin of the office in Autodesk Platform Services

As depicted in the graph (Figure 3), the sound data over a seven-hour period is classified and visualized. The relative sound pressure level is represented in a dark blue line graph. The Construction and maintenance work/ Interruptions category is highlighted in red. The Gathering/Crowded/Music category is shown in purple, while the Silence/Ambient Office Sounds (such as printer noise) category is displayed in light blue. The "Other Classes" category is represented in dark grey.

Using these color codes and reading the graph from left to right, a meeting is detected from approximately 11:30 a.m. for two hours, represented in yellow. This is followed by a two-and-a-half-hour period of silence, shown in light blue. Finally, starting at 5:00 p.m., another meeting is observed for about two hours.

This chart is accessible through the digital twin platform's web interface by clicking on the acoustic monitoring icon available for each room.

4 Conclusions and Future Work

The focus of this paper was to provide a framework to perform acoustic comfort analysis by combining various metrics and indicators together with sound event category identification within a digital twin. Regarding SED models, YAMNet is used in the case study. However, various other models, such as SoundNet [21], Google's VGGish, HuBERT [22], OpenL3 [23], the CRNN proposed in [24], DENet [25], SincNet [26], and COPE [27] can also be used. YAMNet was selected for this study as it is a lightweight model suitable for implementation on edge devices, and it offers acceptable performance. Additionally, it is pre-trained and is adaptable to various types of environments. However, to get the best results, the performance of these networks should be evaluated and compared based on specific use case and fine-tuning.

The selection of both the microphone and sound interface plays a critical role in ensuring classification accuracy and precise measurement of acoustic parameters. Using a microphone with an inappropriate directional type in relation to its placement or one with a low signal-to-noise ratio (SNR) can limit the effective measurement range, creating blind spots in the system's sound classification capabilities. Furthermore, to fully utilize a microphone's potential, the sound interface must support high bit depth and sample rate, which are essential for capturing maximum acoustic data from the environment for further analyses.

The framework also included integrating acoustic measurements, such as acoustic waveform variables (e.g., sound pressure levels), psychoacoustic indicators (e.g., loudness), as well as sound event classification, into a digital twin platform. The potential applications of this

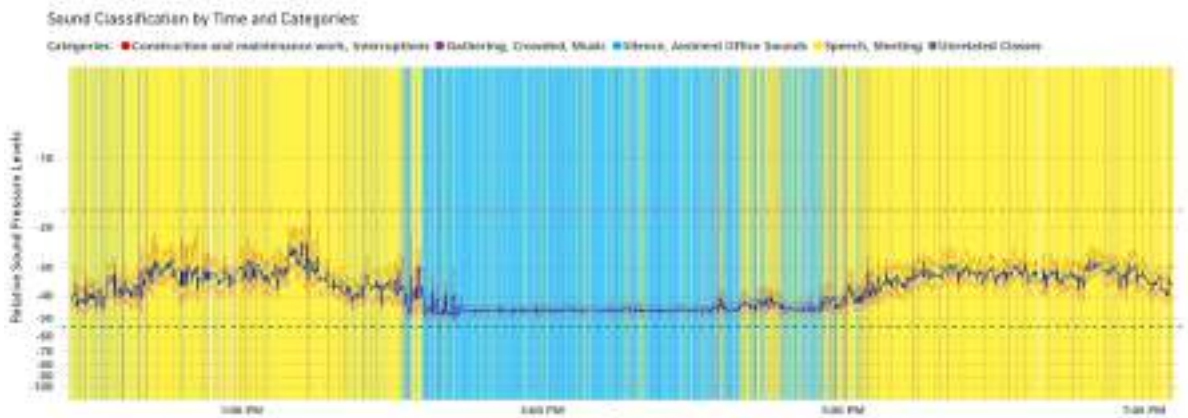


Figure 3. The Output Graph for Monitoring the Acoustic Status of the Case Study

framework were explored, and a case study was conducted in an office space, utilizing edge devices and an SED network to monitor both the current and historical status of the environment in a digital twin platform.

References

[1] Heinzerling, D., Schiavon, S., Webster, T., and Arens, E. Indoor environmental quality assessment models: A literature review and a proposed weighting and classification scheme. *Building and Environment*. 70 :210–222. 2013.

[2] Fasano, S., Fissore, V. I., Puglisi, G. E., Shtrepi, L., and Astolfi, A. Acoustic comfort contribution to the overall indoor environmental quality in workplaces.

[3] Navai, M. and Veitch, J. Acoustic Satisfaction in Open-Plan Offices: Review and Recommendations. 2003.

[4] Zhao, Y., Zhao, Q., Xia, L., Cheng, Z., Wang, F., and Song, F. A unified control framework of HVAC system for thermal and acoustic comforts in office building. *2013 IEEE International Conference on Automation Science and Engineering (CASE)*. pages 416–421. 2013.

[5] Bonet-Solà, D., Vidaña-Vila, E., and Alsina-Pagès, R. M. Acoustic Comfort Prediction: Integrating Sound Event Detection and Noise Levels from a Wireless Acoustic Sensor Network. *Sensors*. 24 (13) :4400. 2024.

[6] Zhang, E., Chen, Y., Chen, X., Zhang, J., Xu, P., and Zhuo, J. High-Precision Modeling and Prediction of Acoustic Comfort for Electric Bus Based on BPNN and XGBoost. *The International Journal of Acoustics and Vibration*. 28 (2) :158–164. 2023.

[7] Dobie, R. A. and Hemel, S. V. Hearing Loss: Determining Eligibility for Social Security Benefits. *Hearing Loss: Determining Eligibility for Social Security Benefits*. 2004.

[8] Kang, J., Aletta, F., Gjestland, T. T., Brown, L. A., Botteldooren, D., Schulte-Fortkamp, B., Lercher, P., Van Kamp, I., Genuit, K., Fiebig, A., Bento Coelho, J. L., Maffei, L., and Lavia, L. Ten questions on the soundscapes of the built environment. *Building and Environment*. 108 :284–294. 2016.

[9] Fastl, H. and Zwicker, E. *Psychoacoustics*. Berlin, Heidelberg, Springer Berlin Heidelberg. 2007.

[10] Rossing, T. D. *Springer handbook of acoustics*. New York, Springer. 2007.

[11] Hossain, M. R., Manohare, M., and King, E. A. Systematic review of indoor soundscape assessments: Activity-based psycho-acoustics analysis. *Building Acoustics*. 32 (1) :123–141. 2025.

[12] Aletta, F., Kang, J., and Axelsson, Ö. Soundscape descriptors and a conceptual framework for developing predictive soundscape models. *Landscape and Urban Planning*. 149 :65–74. 2016.

[13] Engel, M. S., Fiebig, A., Pfaffenbach, C., and Fels, J. A Review of the Use of Psychoacoustic Indicators on Soundscape Studies. *Current Pollution Reports*. 7 (3) :359–378. 2021.

[14] Vardaxis, N.-G., Bard, D., and Persson Waye, K. Review of acoustic comfort evaluation in dwellings—part I: Associations of acoustic field data to subjective responses from building surveys. *Building Acoustics*. 25 (2) :151–170. 2018.

[15] Herranz-Pascual, K., Iraurgi, I., Aspuru, I., Garcia-Pérez, I., Eguiguren, J. L., and Santander, Á. Development of the Acoustic Comfort Assessment Scale (ACAS-12): Psychometric properties, validity evidence and back-translation between Spanish and English. *PLOS ONE*. 18 (2) :e0281534. 2023.

[16] Kang, J., Aletta, F., Margaritis, E., and Yang, M. A model for implementing soundscape maps in smart cities. *Noise Mapping*. 5 (1) :46–59. 2018.

[17] Espejo, D., Vargas, V., Viveros-Muñoz, R., Labra, F. A., Huijse, P., and Poblete, V. Short-time acoustic indices for monitoring urban-natural environments using artificial neural networks. *Ecological Indicators*. 160 :111775. 2024.

[18] Yue, R., Meng, Q., Yang, D., Wu, Y., Liu, F., and Yan, W. A visualized soundscape prediction model for design processes in urban parks. *Building Simulation*. 16 (3) :337–356. 2023.

[19] Gemmeke, J. F., Ellis, D. P. W., Freedman, D., Jansen, A., Lawrence, W., Moore, R. C., Plakal, M., and Ritter, M. Audio Set: An ontology and human-labeled dataset for audio events. *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. pages 776–780. 2017.

[20] Hershey, S., Chaudhuri, S., Ellis, D. P. W., Gemmeke, J. F., Jansen, A., Moore, R. C., Plakal, M., Platt, D., Saurous, R. A., Seybold, B., Slaney, M., Weiss, R. J., and Wilson, K. CNN architectures for large-scale audio classification. *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. pages 131–135. 2017.

[21] Aytar, Y., Vondrick, C., and Torralba, A. SoundNet: Learning Sound Representations from Unlabeled Video. 2016.

[22] Hsu, W.-N., Bolte, B., Tsai, Y.-H. H., Lakhotia, K., Salakhutdinov, R., and Mohamed, A. HuBERT: Self-Supervised Speech Representation Learning by Masked Prediction of Hidden Units. 2021.

[23] Cramer, A. L., Wu, H.-H., Salamon, J., and Bello, J. P. Look, Listen, and Learn More: Design Choices for Deep Audio Embeddings. *ICASSP 2019 - 2019 IEEE International Conference on Acoustics,*

Speech and Signal Processing (ICASSP). pages 3852–3856. 2019.

[24] Çakır, E., Parascandolo, G., Heittola, T., Huttunen, H., and Virtanen, T. Convolutional Recurrent Neural Networks for Polyphonic Sound Event Detection. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 25 (6) :1291–1303. 2017.

[25] Greco, A., Roberto, A., Saggese, A., and Vento, M. DENet: a deep architecture for audio surveillance applications. *Neural Computing and Applications*. 33 (17) :11273–11284. 2021.

[26] Roberto, A., Saggese, A., and Vento, M. A deep convolutionary network for automatic detection of audio events. *Proceedings of the 3rd International Conference on Applications of Intelligent Systems*. In Las Palmas de Gran Canaria Spain. pages 1–6. 2020.

[27] Strisciuglio, N., Vento, M., and Petkov, N. Learning representations of sound using trainable COPE feature extractors. *Pattern Recognition*. 92 :25–36. 2019.

Towards Remote Construction Site Analysis: From Laser Point Clouds to Virtual Reality Applications

Trevor Neece[1], **Alessandro Fascetti**[1]

[1]Department of Civil and Environmental Engineering, University of Pittsburgh, United States of America

trevor.neece@pitt.edu, fascetti@pitt.edu

Abstract -

Efficient construction monitoring and project visualization are vital in civil engineering: they promote safety, minimize costs, and ensure adherence to design plans. With advances in reality capture techniques, such as LiDAR, point clouds have emerged as a powerful tool for digitizing large-scale scenes with excellent accuracy. However, the challenges of processing and visualizing large-scale point cloud data remain significant, particularly for applications in horizontal infrastructure projects. In this case study, a point cloud-based digital reconstruction of an interchange conversion project was developed to aid in defining best practices for data processing of large-scale point clouds and to develop novel digital tools to enable remote construction site inspection and analysis. The obtained models are leveraged to provide insights on three advanced methodologies for data exploration: (1) desktop interfaces for accessible remote monitoring, (2) holographic interfaces for collaborative visualization, and (3) immersive virtual environments for enhanced interaction and simulation. These three methods are compared for accessibility, opportunity, and their potential to improve decision-making and project outcomes in construction monitoring and analysis.

Keywords -

Point clouds; Laser Scanning; Virtual Reality; Construction Monitoring; Construction Simulation

1 Introduction

Infrastructure systems in the United States are poised for significant investments over the coming years, given the aging conditions and urgent need for renovation [1]. When developing and completing large construction projects, it is crucial to ensure that each step is executed accurately and according to specifications to maintain the structural integrity and overall performance of the completed systems [2]. Errors in construction processes can compromise safety, increase costs, and decrease the lifespan of the built infrastructure. To that end, there are broader consequences with poorly monitored projects, such as premature demolition and replacement - which have significant environmental impacts [3]. Given that greenhouse gas emissions are strongly correlated to the construction of public infrastructure [4], it is evident that poorly monitored projects, which may require early demolition and rebuilding, can exacerbate environmental impacts. If methods to extend infrastructure lifespan are not implemented, the repeated cycle of demolition and rebuilding will deplete environmental resources and contribute significantly to greenhouse gas emissions.

Recent advancements in digitalization and the rise of intelligent automation are driving a technological shift in the construction industry, enhancing the effectiveness of construction monitoring [5]. By leveraging innovations in reality capture, computer vision, robotics, and artificial intelligence, companies can improve data capture for complex infrastructure project management [6, 7]. In particular, rapid advancements in laser scanning technology have revolutionized the ability to capture detailed spatial information for infrastructure projects. LiDAR is a technique that uses laser beams to precisely measure distances and generate detailed three-dimensional representations of physical spaces. From the capture, data fusion with HDR imaging provides RGB information to each registered point in the cloud [8]. LiDAR devices are generally attached to a ground-based fixture (terrestrial laser scanning - TLS) or an aerial vehicle (aerial laser scanning - ALS). The capability of producing high-density scans at a relatively low cost enables the digitalization and automation of advanced construction management procedures [9]. LiDAR data can similarly streamline the capture of geometric information for computational analysis of infrastructure [10]. As such, research into advanced methods of data acquisition, exploration, and management [11] supports the development of accurate and reliable workflows for construction applications.

As data collection methods in the construction industry advance, adopting effective processes to convert data into accessible formats is crucial for effective visualization to enable remote inspection and monitoring. [12]. Since the information that operators need to extract from point clouds is very technical, there must be a clear workflow for processing and visualization to make it operable. Cleaning raw point cloud data for large-scale scenes is a laborious, time-consuming task, requiring manual processing and

expert judgment by skilled operators. Data storage considerations are also relevant [13], thus requiring a balance between accuracy and cost.

Leveraging point cloud data in civil engineering projects comes with many challenges. Due to the capture devices' technical constraints, point cloud data of active construction sites will contain several different types of noise [14]. When capturing ongoing projects, fully closing the site or nearby streets to traffic and workers is often impractical, resulting in the presence of unintended people and equipment in the scene [15]. Additionally, remote site monitoring requires a large amount of human judgment and expertise. As such, using effective methods to visualize data is highly important to complete the task. With technology improving, there is a need to test new methods and prove viability for applications in the industry [16].

Advanced visualization methods to explore site data are becoming more viable as consumer-grade virtual reality (VR) and holographic hardware become more accessible. Immersive virtual environments (IVEs) are a frequently used research tool to simulate and navigate digital environments [17]. When IVEs are built using computer graphics engines, custom modules and simulations based on point cloud data can fulfill the unique needs of individual projects. IVEs have successfully been used in applications such as health and safety training [18], digital twin modeling [19], and project visualization [20]. In this context, point cloud data obtained in real-life conditions can improve the IVEs by providing greater detail of the scenes and greater contextual realism of the specific settings being materialized [21].

There is a need for studies regarding data management, visualization, and exploration methods for point cloud data of large-scale construction scenes [9]. The capabilities for structural assessment have been well documented [22]; however, work to define the best practices for point cloud-based construction monitoring of large-scale horizontal infrastructure projects is needed. This paper introduces a novel workflow for processing and visualizing point cloud-based reconstructions of horizontal infrastructure projects, with a particular focus on identifying critical challenges in data processing and exploring effective methods for data exploration. A case study on an active highway construction site demonstrates the practical application of the proposed approach. The study presents a comprehensive analysis and comparison of various visualization and exploration techniques, including desktop computer programs, holographic 3D environments, and immersive virtual environments (IVEs), to evaluate their capabilities in extracting technical information and enabling remote construction site monitoring and inspection.

2 Methodology

2.1 Data Collection

An active construction site during an interchange conversion project (i.e., transition from a cloverleaf to a diverging diamond interchange) in Southwestern Pennsylvania was chosen as the case study. The location is of interest due to the active work zone, non-uniform terrain conditions, and various structures at different construction stages. The scan was performed on one cloverleaf loop to be replaced and the surrounding highway and state route.

The laser scan was conducted using the Leica BLK360 terrestrial scanner (see Figure 1 (A)). The BLK360 features a 360,000 points-per-second scan rate and can capture information at a range of up to 60 meters with a point accuracy of approximately 6 mm at 10 meters. The laser scanner has a 360° horizontal and a 300° vertical field of view. For each scanning iteration, the BLK360 (mounted on a tripod) was positioned in an area of interest, and the scanning cycle was initiated using a wireless connection through a portable device. After completion of the scanning cycle, the portable device provided a 360° image of the location and a low-density preview of the captured point cloud (see Figure 1 (C)).

The data collection process employed the following method: firstly, the scanner was placed at an initial location, and a scan was performed. For each successful scan, the device was moved to a nearby location where the following scan would have significant overlapping data such that an automated alignment link would be created. Links appear on top of a low-density site map, with green links found to be especially strong [23] (See Figure 1 (B)). Using the scan previews, the work continued until data was captured of the whole site. Special focus was provided to the portions of the site with the most active work ongoing.

2.2 Data Processing

The point clouds were then processed to simplify the data. For each scan, a two-step process was performed. (1) Automated coincident point removal among nearby scans and (2) removal of noise through automated statistical outlier removal and manual segmentation. Statistical outlier removal identifies and removes points that deviate significantly from their neighbors. It calculates the average distance of each point to its k nearest neighbors and removes those exceeding this average within a multiple of the standard deviation. The process is controlled by two input parameters: k, which defines the number of neighbors, and a multiplicative factor for the standard deviation, which determines the threshold for removal [24]. In this case, the parameters of $k = 6$ and a multiplication factor of 1 were used. The predominant occurrences of noise were vehicles driving down the interstate and around the site,

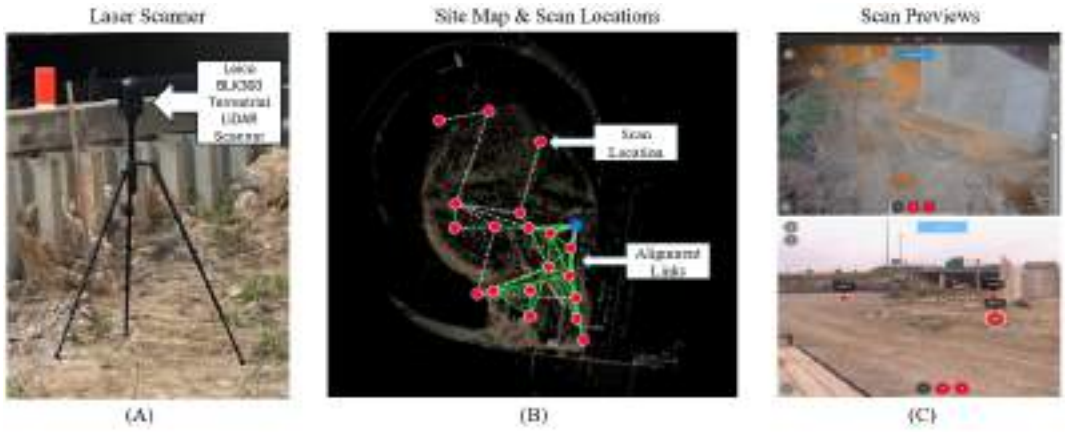


Figure 1. Leica BLK 360 Terrestrial Scanner (A) Site map with scan locations and alignment links (B) and scan preview via low-density point cloud and 3D image (C).

workers moving throughout the project, and lasers bouncing off reflective surfaces. In many cases, the scanner captured an object, but the object is inconsistent between scans and thus must be removed. Additionally, since the data collection process involved intentionally overlapping scans, nearby scans have many identical points. Coincident points included automated removal of overlapping data and, since the scanning device is less accurate as points are captured from further distances, low-density captures at further distances were removed during the manual segmentation process. Each scan was processed individually to remove noisy data, addressing the described instances of floating points, partially captured moving objects, and reflections.

For example, Figure 2 shows the cleaning of scan #1. In Figure 2 (A), the initial scan has a significant amount of noise from the nearby highway, inaccurate data from the reflection of a nearby truck, and workers in high visibility vests moving around the site. Figure 2 (B) shows the resultant point cloud after the coincident point removal and manual segmentation-noise deletion. By performing the cleaning workflow, the scan improved via the removal of reflection noise, partially captured workers and vehicles, and lower accuracy points captured far from the scan location. Along with being a more accurate capture of the true site conditions, this model also becomes more operable with a significantly lower file size ($\tilde{8}9\%$ reduction).

Another significant issue encountered was the misalignment of scans without vertical features. Scans which primarily consisted of the ground surface had inaccurate alignment in the vertical direction when performing automated adjustments, leading to inaccuracies in the dataset. Manual fine-tuning in the vertical direction was performed as a solution.



Figure 2. Example of data processing for scan #1. (A) Before and (B) after.

2.3 Data Exploration

The resulting high-fidelity point cloud can be leveraged for construction monitoring and project visualization activities. The following section describes three methods for data exploration, remote monitoring, and visualization: (1) desktop interface, (2) holographic interface, and (3) immersive virtual environments. A scenario-based evaluation was employed to validate the utility of these interfaces for data exploration to enable remote construction site analysis. This process did not involve comparison to project plans or reference measurements, but instead was a feasibility study for the visualization methods. In this process, the research team defined and performed standard tasks using the processed digital models to assess each interface's strengths, weaknesses, and capabilities. First, measurements of key structural elements—such as reinforced concrete walls, rebar spacing, wall angles, and foundation depths—were taken to verify that the built conditions could be measured. Second, feasibility of progress tracking and collaboration capabilities was assessed. Finally, the assessment of terrain elevations was undertaken to ensure that earth mobility could be captured. Additionally, informal interviews and visualization trials with construction operators conducted during on-site data col-

lection and through visits to the lab provided further insights into the methods' viability for construction monitoring tasks, and strengths or weaknesses as a collaboration and presentation tool.

2.3.1 Desktop Interface

For performing remote construction monitoring, the desktop interface (also known as 3D keyboard mouse) refers to a technician exploring the point cloud to make observations regarding project status, site layout, and gathering as-built measurements. Accessing the data in this capacity can be advantageous given the highly accessible nature of desktop computers and point cloud-compatible software. Additionally, many CAD and BIM workflows have compatible integration with point clouds [25, 26].

Considering point cloud data's high accuracy and precision, project status can be judged based on deviations from the planned design. At the same time, site layout assessments help verify spatial arrangements and identify potential conflicts or errors in construction. This method similarly helps maintain a safe and productive site layout. One significant object of interest in the captured data is the reinforced concrete wall, which will eventually support new sections of the roadway. Figure 3 (A) captures the measurements built from this construction feature. Such measurements include rebar spacing, wall angle, visible depth of foundation, and terrain elevation differences along the site. Measurements can be transformed into as-built drawings or 3D static mesh components.

The primary disadvantages of desktop workflows stem from the limited field of view and spatial understanding of the three-dimensional model viewed in a two-dimensional medium. As such, desktop workflows offer a limited sense of scale compared to 3D visualization methods. However, desktop workflows are typically more accessible and user-friendly, as they leverage widely familiar interfaces and are consistent with the majority of users' prior experience.

2.3.2 Holographic Interface

Another data exploration technique is using holographic models to represent the point cloud in 3D without relying on a head-mounted display (HMD). HMDs can be constraining when viewing 3D models, depending on the pixel density of the HMD. As such, holographic technology such as the Axiom Hologram Table [27] provides the technician with a semi-immersive representation of the project without sacrificing visual fidelity. An example of the construction site point cloud on the holographic table is shown in Figure 3 (B).

Holographic interfaces enable collaborative data exploration with multiple users. In the case of the Axiom device featured in Figure 3 (B), two can be wearing spatially tracked 3D glasses and controllers, and two more can be wearing non-tracked glasses. The difference between tracked and non-tracked glasses is that tracked glasses adjust the view of the hologram based on the user's physical orientation around the table. Non-tracked glasses copy the holographic view from the tracked glasses. As such, those with non-tracked glasses must stand near the tracked user to get an effective view of the scene. This device also features an audience view monitor, allowing spectators to understand what model is being presented to the users. Technicians leverage a remote controller to perform remote site construction site analysis operations, such as measuring construction features, rotating and scaling the environment, and creating presentations using multiple 3D files to display multiple project phases.

This workflow's disadvantages come from the technology's highly specific hardware and software requirements. This type of technology has a high entrance cost compared to the highly available consumer-grade desktops and HMDs. The models must be converted into a proprietary file type so the system can project them holographically. Some complex 3D files containing dynamic scenes are not compatible.

2.3.3 Immersive Environments

For this application, two types of immersive environments were explored: (1) viewing the three-dimensional point cloud in the first person and (2) a high-fidelity 3D reconstruction of the point cloud.

Point clouds can be loaded into applications such as Unreal Engine 5 (UE5) and navigated in VR [28]. Such applications can be advantageous compared to 3D images, as the high accuracy of point clouds should lead to improved spatial context. UE5 provides advanced modules and templates for VR applications and a dedicated LiDAR plugin. As such, point clouds can be uploaded directly and explored in the engine. The first-person point of view for this environment is shown in Figure 3 (C). VR-based toolsets are available for remote monitoring to capture measurement data [29], and can effectively reproduce the same construction monitoring tasks performed using the other visualization methods.

To build a high-fidelity 3D reconstruction of the site, UE5's environment modeling tools were leveraged to rebuild the model as a landscape component and several static mesh components. The landscape component represents the terrain of the site. The key difference between a static mesh and a landscape component is static meshes are non-deformable and predefined. In contrast, landscape components are dynamic - they can be modified in real time. Dynamic landscapes help simulate diverse construction scenarios, such as excavations. The point cloud was manually segmented to exclude everything but



Figure 3. Methods of data exploration for remote monitoring and viewing simulations: desktop interface (A), holographic interface (B), and immersive virtual environments (C).

the surface elevation to generate the landscape component. Given this landscape-only point cloud contained missing data, synthetic points were interpolated by generating a Delaunay 2.5D mesh [30] in those areas and sub-sampling the mesh into points. The obtained point cloud was recolored with a height ramp and rasterized into a 2.5D height map image. The height map image was uploaded directly into UE5 to build the landscape (see Figure 4 (A)). Once the landscape was generated, objects of interest from the point cloud were created to fill the environment with infrastructure components relevant to the scene. Objects of interest can be meshed using software such as Reality Capture [28]. Figure 5 shows an example of generating a mesh from the reinforced concrete wall central to the site. Additionally, the objects can be rebuilt manually using the measurements obtainable (see Figure 3 (A)). The roads of the highway interchange were built using a spline tool. Splines build a continuous path of iterated meshes along anchor points, locally adjusting the meshes and landscape to build highways. A similar technique was used to iterate center-line barriers and guide rails along the roads. Figure 4 (B) displays a sample of the VR scene.

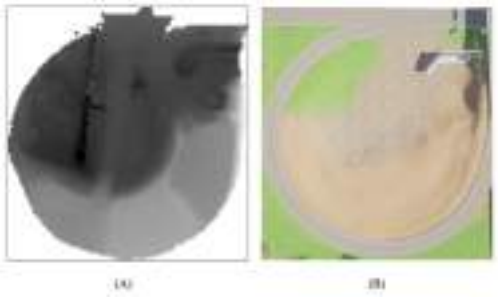


Figure 4. (A) Rasterized 2.5D image and (B) high-fidelity model for VR

This model is helpful for applications where the construction site is relevant as a foundation for a larger project. IVEs can be equipped with code to enable user interaction. This interaction increases the engagement of the model, which is known to increase knowledge retention for the user [31]. For example, various construction scenarios could be built on this scene for immersive construction health and safety training, contextually linked to the project. Additionally, coupled with data fusion from sensors, this technique can be used to build digital twin models.



Figure 5. Reinforced concrete wall meshed using Reality Capture.

The primary disadvantages of VR methods are the entrance price and physical comfort considerations. Although HMDs are becoming more available to consumers, there is an additional need for a high-performance computer to run the simulation. Additionally, only one user can use VR at a time, constraining the collaboration capabilities of the analysis. Regarding physical comfort, users may experience VR sickness (also known as motion sickness), given that the visual perception of their surroundings differs from the physical sensations. When artificial movement is added to a VR simulation, the visual and physical senses of the user do not match, causing motion sickness. Since approximately one-third of the population is highly susceptible to motion sickness [32], extended use of VR is not always practical.

3 Results and Discussion

The work's results are presented as an evaluation of data processing and recommendations for best practices in

visualization and data acquisition.

The original data capture comprised 20 successful scans, totaling ~500,000,000 unique points, and was roughly 10.7 Gb in storage. After performing the methodology explained in Section 2.2, instances of duplicate points and noisy data were removed and a clean point cloud was obtained (see Figure 6 (A)). In the first step, where coincident points were extracted, 75.9% of the initial data was removed. After performing the noise removal, there was a cumulative point reduction of 87.21% and a total file size reduction of 89.04% from the original scan, which shows the significant amount of unnecessary and inaccurate data that can be removed after data collection. This processed point cloud, along with being a more accurate representation of the scene, has lower computational costs and storage requirements than the unprocessed version. The obtained data was dense in points despite the size reduction, and, as shown in Figure 6 (B), there is a high concentration among the construction features of the work site. This high concentration is valuable for precise measurements of construction features.

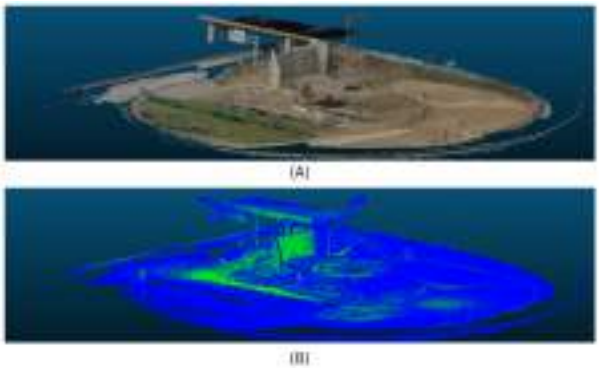


Figure 6. (A) Full point cloud after data processing and (B) Point density after data processing, visualized with a color scale: blue (lowest density) to red (highest density).

Three methods of data exploration were reviewed. Provided with a processed dataset, an expert can make judgments regarding the construction quality with easier access to many details of the site than would be available with an in-person visit. In all three methods, the standard tasks were able to be performed, thus displaying feasibility for these methods to enable remote inspection of construction sites. The efficiency and difficulty of each method for remote monitoring activities are highly dependent on the technicians' experience using the technology. Since most users are expected to have significantly more experience interacting with desktop interfaces, they were found to be the most accessible for remote monitoring. Holographic data visualization methods were found to be best used as a collaborative tool or for presentation purposes. Multiple users can review the 3D data in a semi-immersive environment. Given the depth perception and spatial rendering, participants can see the full scope of the data intuitively. As a presentation tool, the holographic interface is found to be a memorable and engaging method of presentation. For all methods, measurements of construction features were replicable. IVEs are effective for applications where the user should interact with the environment. Immersive virtual environments were found to be similarly engaging and memorable experiences for displaying projects compared to holographic ones.

Based on the described case study, the following best practices are proposed for future applications of virtual construction site analysis of horizontal infrastructure:

1. The TLS scans should be captured across the work zone so that nearby captures overlap in data, and areas with the most significant construction features are prioritized.
2. Engage in manually cleaning the data to remove artifacts such as floating points, partial objects, and reflections—additionally, leverage coincident point removal to eliminate overlapping data. While aligning, pay special attention to the vertical alignment links - especially in areas where the scan primarily captures terrain.
3. Leverage different visualization techniques depending on the use case of the data. For accessible remote monitoring, prioritize desktop interfaces. Use immersive techniques such as holographic interfaces and IVES for presentations, collaborative data exploration, or activities involving interaction with the model.
4. When building a high-fidelity IVE, separate the landscape from the construction features to create an environment with a well-defined terrain before adding construction features. Leverage the measurement techniques or meshing capabilities available from point clouds to recreate objects of interest accurately.

4 Conclusion

This study introduced a novel workflow for processing and visualizing point cloud-based reconstructions of horizontal infrastructure projects, through discussion of dealing with common types of noisy data relevant to construction sites and an exploration of data exploration methods. The data collection and processing were described to help define best practices for large-scale terrestrial point cloud capture of horizontal infrastructure. Additionally, a data processing framework was explored which highlights the significant amount of either unnecessary or noisy data within point clouds of active horizontal construction sites. To judge the viability of remote project monitoring and

simulation rendering, data exploration methods were surveyed, with an emphasis on the most advanced technological methods available. The highlighted methods were desktop interface, holographic interface, and immersive virtual environments. These advanced techniques proved promising for remote inspection and analysis of construction sites. Each method was found to be viable when performing standard construction monitoring tasks such as observing construction features, taking measurements, and presenting information. However, all methods require individual expertise for efficient operations in both site inspection and collaboration activities.

This work enables a formal quantitative study to improve understanding and analysis of the efficiency and difficulty of each method, along with assessment of cognitive load within larger groups of users. Future work will focus on developing the high-fidelity 3D model into practical use cases. Industry partners have been identified to guide this development. The work presented opens the possibility of defining a novel framework for developing IVEs from laser scans obtained from construction sites. In the future, this could provide a defined methodology to convert real-life projects into advanced, context-specific IVEs suitable for simulation.

Acknowledgements

The findings reported here are based on work performed with the support of the Impactful Resilient Infrastructure Science & Engineering (IRISE) consortium in the Department of Civil and Environmental Engineering, Swanson School of Engineering at the University of Pittsburgh.

References

[1] American Society of Civil Engineers. 2021 report card for america's infrastructure, March 2021. URL https://www.infrastructurereportcard.org/. Accessed: 2024-12-16.

[2] Guiwen Liu, Kexi Xu, Xiaoling Zhang, and Guomin Zhang. Factors influencing the service lifespan of buildings: An improved hedonic model. *Habitat international*, 43:274–282, 2014.

[3] Jingjing Wang, Yurong Zhang, and Yuanfeng Wang. Environmental impacts of short building lifespans in china considering time value. *Journal of Cleaner Production*, 203:696–707, 2018.

[4] Yichen Qian, Renjin Sun, Jun Zhao, Sasa Song, and Pinchen Wang. How does public infrastructure construction influence co2 emissions? a global case. *Frontiers in Environmental Science*, 10:1048297, 2022.

[5] Federica Geremicca and Melissa M Bilec. Searching for new urban metabolism techniques: A review towards future development for a city-scale urban metabolism digital twin. *Sustainable Cities and Society*, page 105445, 2024.

[6] Mohammadsaeid Parsamehr, Udara Sachinthana Perera, Tharindu C Dodanwala, Piyaruwan Perera, and Rajeev Ruparathna. A review of construction management challenges and bim-based solutions: perspectives from the schedule, cost, quality, and safety management. *Asian Journal of Civil Engineering*, 24(1):353–389, 2023.

[7] McKinsey & Company. Rise of the platform era: The next chapter in construction technology, 2023. URL https://www.mckinsey.com/industries/private-capital/our-insights/rise-of-the-platform-era-the-next-chapter-in-const Accessed: 2024-12-16.

[8] Imanol Munoz-Pandiella, Marc Comino Trinidad, Carlos Andújar, Oscar Argudo, Carles Bosch, Antonio Chica, and Beatriz Martínez. Gain compensation across lidar scans. *Computers & Graphics*, 106: 174–186, 2022.

[9] Yingbo Zhu, John C Brigham, and Alessandro Fascetti. Lidar-rgb data fusion for four-dimensional uav-based monitoring of reinforced concrete bridge construction: Case study of the fern hollow bridge reconstruction. *Journal of Construction Engineering and Management*, 151(1):05024016, 2025.

[10] Alessandro Fascetti and Caglar Oskay. Multiscale modeling of backward erosion piping in flood protection system infrastructure. *Computer-Aided Civil and Infrastructure Engineering*, 34(12):1071–1086, 2019.

[11] Osama Moselhi, Hassan Bardareh, and Zhenhua Zhu. Automated data acquisition in construction with remote sensing technologies. *Applied Sciences*, 10 (8):2846, 2020.

[12] Sajjad Einizinab, Kourosh Khoshelham, Stephan Winter, Philip Christopher, Yihai Fang, Eric Windholz, Marko Radanovic, and Songbo Hu. Enabling technologies for remote and virtual inspection of building work. *Automation in Construction*, 156: 105096, 2023.

[13] Chamin Nalinda Lokugam Hewage, Debra F Laefer, Anh-Vu Vo, Nhien-An Le-Khac, and Michela Bertolotto. Scalability and performance of lidar point cloud data management systems: A state-of-the-art review. *Remote Sensing*, 14(20):5277, 2022.

[14] Qian Wang, Yi Tan, and Zhongya Mei. Computational methods of acquisition and processing of 3d point cloud data for construction applications. *Archives of computational methods in engineering*, 27(2):479–499, 2020.

[15] Thomas Kanzok, Falk Süß, Lars Linsen, and Paul Rosenthal. Efficient removal of inconsistencies in large multi-scan point clouds. 2013.

[16] Juan Manuel Davila Delgado, Lukumon Oyedele, Thomas Beach, and Peter Demian. Augmented and virtual reality in construction: drivers and limitations for industry adoption. *Journal of construction engineering and management*, 146(7):04020079, 2020.

[17] Jack M Loomis, James J Blascovich, and Andrew C Beall. Immersive virtual environment technology as a basic research tool in psychology. *Behavior research methods, instruments, & computers*, 31(4): 557–564, 1999.

[18] Rafael Sacks, Amotz Perlman, and Ronen Barak. Construction safety training using immersive virtual reality. *Construction Management and Economics*, 31(9):1005–1017, 2013. ISSN 0144-6193.

[19] Johanna Pirker, Enrica Loria, Saeed Safikhani, Andreas Künz, and Sabrina Rosmann. Immersive virtual reality for virtual and digital twins: A literature review to identify state of the art and perspectives. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 114–115. IEEE, 2022.

[20] Lilia Potseluyko, Farzad Pour Rahimian, Nashwan Dawood, Faris Elghaish, and Aso Hajirasouli. Game-like interactive environment using bim-based virtual reality for the timber frame self-build housing sector. *Automation in Construction*, 142:104496, 2022. ISSN 0926-5805.

[21] Dimitrios Bolkas, Jeffrey Chiampi, John Chapman, and Vincent F Pavill. Creating a virtual reality environment with a fusion of suas and tls point-clouds. *International journal of image and data fusion*, 11 (2):136–161, 2020.

[22] Elise Kaartinen, Kyle Dunphy, and Ayan Sadhu. Lidar-based structural health monitoring: Applications in civil infrastructure systems. *Sensors*, 22(12): 4610, 2022.

[23] Leica Geosystems. Cyclone register 360 plus help / cyclone register 360 /review and optimize, 2024. URL `https://rcdocs.leica-geosystems.com/cyclone-register-360/latest/creating-links`. Accessed: 2024-12-19.

[24] CloudCompare Developers. Cloudcompare documentation, 2024. URL `https://www.cloudcompare.org/doc/`. Accessed: 2024-12-19.

[25] Autodesk, Inc. *Autodesk AutoCAD 2022 Help Documentation - About Working With Point Clouds*, 2022. URL `https://help.autodesk.com/view/ACD/2022/ENU/?guid=GUID-C0C610D0-9784-4E87-A857-F17F1F7FEEBE`. Accessed: 2024-12-20.

[26] Autodesk, Inc. *Autodesk Revit 2022 Help Documentation - About Point Clouds*, 2022. URL `https://help.autodesk.com/view/RVT/2022/ENU/?guid=GUID-BD499295-84DD-4BDE-B60D-73008AFBC791`. Accessed: 2024-12-20.

[27] Axiom Holographics. Axiom Holographics Hologram Table, 2024. URL `https://axiomholographics.com/devices/hologram-table/`. Accessed: 2024-12-18.

[28] Epic Games. Unreal engine 5.5 documentation, 2025. URL `https://dev.epicgames.com/documentation/en-us/unreal-engine/unreal-engine-5-5-documentation`. Accessed: February 27, 2025.

[29] Daniel Garrido, Rui Rodrigues, A Augusto Sousa, Joao Jacob, and Daniel Castro Silva. Point cloud interaction and manipulation in virtual reality. In *2021 5th International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, pages 15–20, 2021.

[30] CloudCompare. Cloud compare documentation: Alignment and registeration, 2015. URL `https://www.cloudcompare.org/doc/wiki/index.php/Alignment_and_Registration#Alignment`.

[31] Michael J Burke, Sue Ann Sarpy, Kristin Smith-Crowe, Suzanne Chan-Serafin, Rommel O Salvador, and Gazi Islam. Relative effectiveness of worker safety and health training methods. *American journal of public health*, 96(2):315–324, 2006. ISSN 0090-0036.

[32] Craig R Sherman. Motion sickness: review of causes and preventive strategies. *Journal of travel medicine*, 9(5):251–256, 2002. ISSN 1195-1982.

Computer Vision based Automated Timber Structural Defect Detection Framework

Afia Rasool[1], Qipei Mei[2], Rafiq Ahmad[1,*]

[1]Smart & Sustainable Manufacturing Systems Laboratory (SMART Lab), Department of Mechanical Engineering, University of Alberta, Edmonton, Canada
[2]Department of Civil and Environmental Engineering, University of Alberta, Edmonton, Canada
arasool@ualberta.ca, qipei.mei@ualberta.ca, rafiq.ahmad@ualberta.ca

Abstract –

In the wood construction industry, timber structural defect detection is usually considered a pre-manufacturing inspection step done manually. To address this issue, the proposed study discusses the timber structural defect detection method based on YOLOv8 variants. The evaluation matrices used are precision, recall, mAP.5, and mAP.5-.95, and the results indicate stable convergence and consistent accuracy on the complex dataset instances. This research contributes to the automation of timber defect detection for precise and robust manufacturing of timber structures. The proposed method further improves resource utilization and contributes towards eliminating waste in the residential construction industry.

Keywords –

Timber; Defect Detection; Mass-Timber Manufacturing; YOLOv8; Construction Industry; Computer Vision

1 Introduction

Timber, a finished wood product, is a major building material for residential and commercial buildings worldwide. Timber exhibits unique structural capacity and fineness, making it suitable for panels, trusses, ridges, beams, and staircases. Like every other wood type, natural timber exhibits multiple defects due to genetic and growth factors, processing phase, and environmental and storage conditions [1]. Defective timber can harm the overall building integrity and durability and must be excluded in final processing. Surface defects like knots and burls originate from growth conditions and can easily be identified at an early stage of timber handling. Contrarily, twisting, warping, and splits often occur during the manufacturing and drying phases of finished stud handling, and many construction workshops use experts' eye and appropriate angles for manual stud quality inspection and anomaly identification.

Therefore, a low-cost, fast solution is needed to identify complex structural irregularities like twists and bends [2-5]. Machine vision and deep learning solutions are popular for object detection, which utilize CNNs to mimic human brain learning behaviors [6-8]. Some research proposes generalized novel algorithms utilizing different layering of CNNs [6,9], whereas other researchers focus on utilizing and fine-tuning the proposed novel and hybrid algorithms to better tune them for specific tasks, environments, and datasets [10-12]. A recent study on defect detection across variant timber types has combined image preprocessing techniques with CNN-based architecture to build a vision-enabled defect detection platform [13]. Another similar approach was utilized for precise sawing and grading automation in timber processing [14]. A method called EMINet uses cross-fusion modules for the surface defect detection of sawn timber [15]. This algorithm is designed to learn and infer multi-scale features with greater speed and accuracy. Xi et al. have utilized SPD-Convolution, SiAFF-PA Net, and multi-attention mechanisms in their model to refine the localization and classification of timber defects [16]. Furthermore, different types of stained wood grain defects have been discussed in literature by incorporating CNN architecture to boost the quality inspection and grading mechanism defects [17].

Image detection algorithms can be divided into two categories: two-stage detection and one-stage detection. R-CNN (Region-based Convolutional Neural Network) [18], Fast R-CNN [19], and Faster R-CNN [20] are popular two-stage models comprising region proposal blocks and classification blocks. Whereas the one-stage object detection has a unified region proposal generation, and classification mechanisms are proposed in YOLO (You Only Look Once) [21-24] and SSD (Single Shot Multibox Detector) [25]. Regarding timber defect detection, the YOLO series has been enhanced in several ways to boost detection accuracy and specified classes. BDCS-YOLO is an enhanced method that showed a 12.3 percent improvement in timber defect detection and sawing for resource optimizations [13]. Another study

has used the SimAm module and Ghost convolution to improve YOLOv5, reducing the computational costs for wood surface defects [26]. Another study has focused on efficiently detecting cracks in wood surfaces by proposing ICDW-YOLO algorithm [27].

Besides multiple pieces of research exhibiting good performance on surface-level defects such as minor cracks, knots, and stains, it is clear from the literature study that no significant work has focused on utilizing computer vision to inspect physical defects such as bends, twists, and missing of wood parts within timber studs. This research is focused on generating industrial solutions for the precise localization of defective parts using the YOLOv8 algorithm. To automate the precise detection of structural defects in timber studs, this research begins with the data collection and generation of the timber stud structural defect detection dataset, specifically focusing on twists and bends. The developed dataset is used for training and prediction using the YOLOv8 image detection method. An extensive study of evaluation matrices is performed on multiple variants of YOLOv8 (YOLOv8nano, YOLOv8small, YOLOv8medium, YOLOv8large, YOLOv8x_large) to balance the overall system accuracies and computational cost [2]. Moreover, due to the complexity of the task nature, a prediction study is discussed, emphasizing the importance of initial setup settings during data collection in the overall success of deep learning training.

The precise and easy timber defect detection system can minimize the wasting of complete studs and maximize material utilization, potentially reducing the additional tree harvest. Furthermore, prior screening of timber quality can reduce repair overengineering resulting in durable construction. These factors contribute towards decarbonizing the construction industry and helping easily adopt natural timber as a cheap, sustainable material, lessening carbon emissions. This research can also directly advance industrial practices of automating timber inspection using computer vision and robotics.

The overall flow of the study is shown in Figure 1, and the key steps involve dataset preparation, explained in Section 2; model training on YOLOv8 variants detailed in Sections 3, 4, and 5; and model prediction results are presented in Section 6, followed by conclusion and discussion of future insights.

2 Dataset Preparation

Timber studs are used in the manufacturing of wood buildings as wall panels, roof trusses, floor units, and supporting ridges, and they vary in dimensions a lot. An average stud can be 2-4 meters long, 2-4 inches wide, and 4-8 inches in depth and have a face side (the wider and flatter surface) and an edge side (the narrower surface). They can vary hugely depending on the grade and species, moisture contents, load requirements, and code compliances. The changes in shapes and structural integrity of these timber studs, called structural defects, can occur due to drying, external stress, and improper seasoning. Bend (curvature along the length of stud), twist (spiral warping), bow (curve along the face side), cup (curve along the width), and warp (an amalgamation of bends, twists, and bows) are categorized as structural defects.

Table 1 shows the dataset preparation steps and their details. As the defect can form at any angle or side, we have used 3 camera heights (2 feet, 6 feet, and 18 feet) to

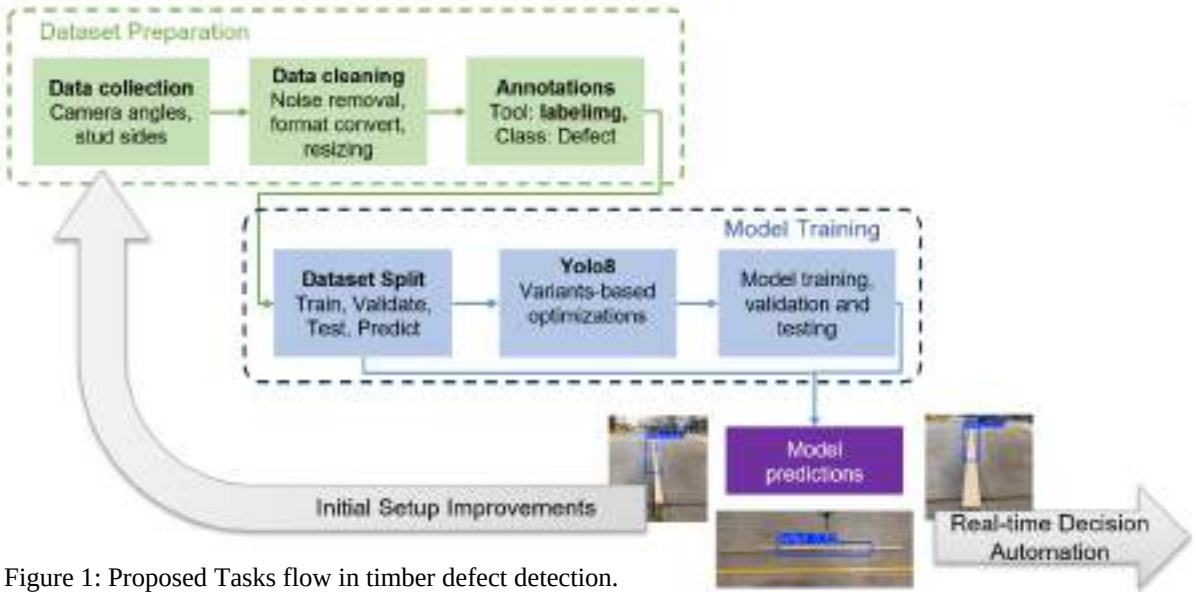


Figure 1: Proposed Tasks flow in timber defect detection.

capture the pictures of a stud on both sides (face and edge). A total of 15 studs, with 12 defective and 3 non-defective, are used to generate 400 multi-dimensional images. These images are manually cleaned to remove irrelevant details and are resized to 640 x 640 sizes to be fed to the YOLOv8 models. LabelImg [28] is a tool for ground truth generation for popular image detectors and is used here to annotate the dataset, creating defect class bounding boxes. The images showing no defect are not assigned to the defect class and are passed to the model training as background images. After cleaning and annotations, the final dataset comprising 285 images with defects and 65 background images is split stud-wise for training, validation, testing, and prediction at 65%, 15%, 15%, and 5%, respectively.

The dataset generation and model training for structural defects is tricky, as a defect visible on the face side at a particular angle may not be detectable across the edge and vice versa. The complexity of the task can be seen in Figure 2, which shows some real instances from the dataset generated.

Table 1: Timber Stud Surface Defect Detection Dataset preparation steps.

Phase	Parameter	Value
Initial Setup	Stud sides	2 (face, edge)
	Camera heights	3 (2 feet, 6 feet, 18 feet)
Data collection	Total studs	15
	Defective studs	12
	Non-defective studs	3
	Total images	400 (varying dimensions)
Resizing, cleaning, and annotations	Total images after data cleaning	350 (640 x 640)
	Background images	65
	Annotated with class: Defect	285
Model training	Training (65 %)	10 studs, 234 images
	Validation (15 %)	2 studs, 48 images
	Testing (15 %)	2 studs, 48 images
	Prediction (5%)	1 stud, 20 images

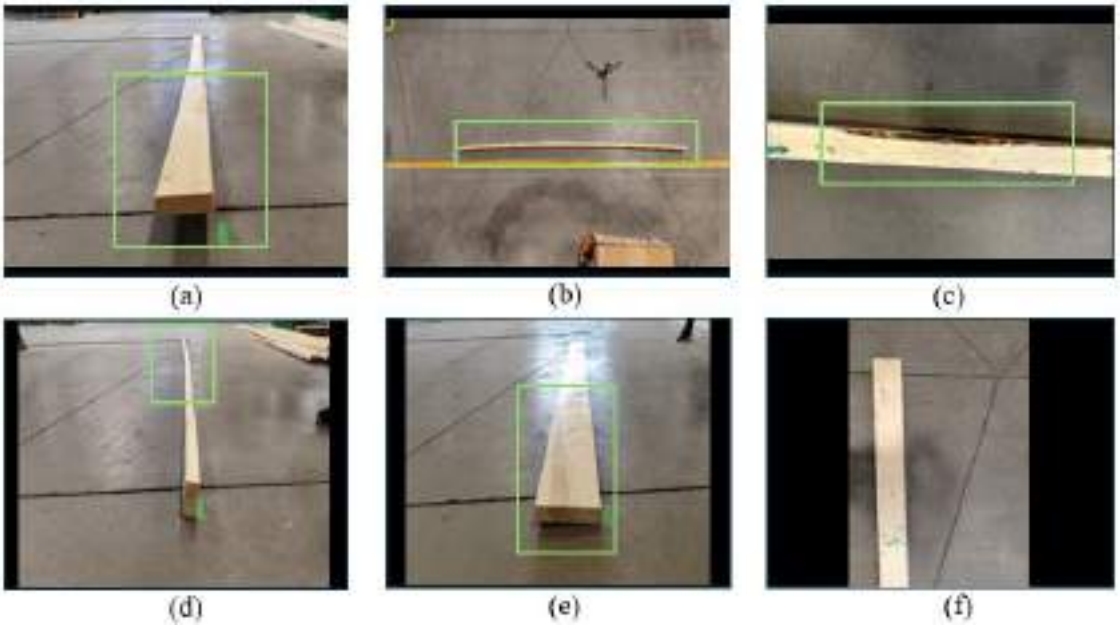


Figure 2: Instances of dataset (a) a bended stud captured from 2 feet height (b) the defect captured from 18 feet height (c) missing wood defect (d) bended and twisted stud from the edge (e) an example of twist and bend (f) background instance showing no defect

3 YOLOv8 Method

YOLO is a one-stage detection model that combines the predecessor's functionalities into improved architecture in its version series. YOLOv8 [2] is an adaptable and efficient model comprising three main modules. The initial backbone is a variant of CSPDarknet, used for multi-scale feature extraction in the model. The second unit, called the neck, implements the PAN-FPN module for feature aggregation, and finally, the head contributes to object classification and bounding box regression. One powerful feature of YOLOv8 is the real-time prediction generation that supports inspection-based applications for faster, lightweight, and more accurate real-time detections. The trade-off between speed and accuracy is well-managed in YOLOv8, which provides precise detections at the rate of high throughput. Moreover, the refined multi-scale feature fusion of this model makes it best suited for the current case of structural defect detection where the images are captured from different heights, offering multi-sized imperfection instances.

As the available dataset is of small size, the training is run on different variants of YOLOv8 to achieve high performance, considering factors of resource optimization, overfitting avoidance, comparative performance analysis, scalability and adaptation. These variants are YOLOv8 nano, YOLOv8 small, YOLOv8 medium, YOLOv8 large, and YOLOv8 extra-large, exhibiting different training speeds, computational requirements, and number of trainable parameters. From nano to large, these variants differ based on the layer count, sizes, and trainable parameters, and support limited to industry real-time applications.

4 Model Training

After the dataset preparation and splitting of the dataset, the data is passed to YOLOv8 variants for training and validation. Table 2 discusses the training machine specifications. The model is trained with the generated dataset on a machine with Central Processing Unit (CPU) Intel(R) Core(TM) Ultra 9 185H 2.50 GHz, Graphics Processing Unit (GPU) NVIDIA GeForce RTX 4070 Laptop GPU (8,188 MiB VRAM), 64 GB RAM (Random Access Memory), with installed version of CUDA 11.2, python 3.9.20 and Pytorch 2.5.1+cu121.

Table 2: Training machine specifications.

Property	Value
CPU	Intel(R) Core(TM) Ultra 9 185H 2.50 GHz
GPU	NVIDIA GeForce RTX 4070 Laptop GPU (8,188 MiB VRAM)
CUDA version	11.2
Operating System	Windows 11 Home
RAM	64 GB
Python version	3.9.20
PyTorch version	2.5.1+cu121

Table 3: Model Training parameters on YOLOv8

Parameter	Value	Parameter	Value
Image Size	640 x 640	Epochs	250
Batch Size	8	Augmentation	hsv_h=0.015
Learning rate	0.001		hsv_s=0.7
Momentum	0.937		hsv_v=0.4
Weight Decay	0.0005		mosaic=1.0

As the generated timber stud structural defect detection dataset is small and multifaceted, the training was performed on different hyperparameters to choose the best settings suitable for better accuracy, and Table 3 shows the carefully selected set of model training parameters on YOLOv8 variants. These hyperparameters were chosen to balance generalizability and computational efficiency, given the dataset constraints. The input image size is 640 x 640 with a batch size kept at 8 trained for 250 epochs during each model training. The batch size was selected based on the GPU availability to avoid any computational instability. The learning rate is initiated with 0.001, momentum of 0.937, and weight decay of 0.0005 to achieve a smoother convergence. The training was also set to early stopping to save the best-trained parameters for inference. During real-time image capturing for real-time inference, the brightness and hues vary greatly depending on the workshop environment and lighting conditions. To make the training robust of inference time conditions, the data augmentation is kept true during the training, and the augmentation parameters are passed as hue 0.015, saturation 0.7, and value 0.4, and the mosaic set is kept on. These data augmentation settings can avoid the overfitting of smaller datasets, making the model resilient to surrounding conditions by slightly changing the picture effects during training.

5 Evaluation Metrics

Typically, YOLOv8 does not support model testing on a different dataset. Instead, it generates the model predictions based on the trained weights. However, to assist the optimization study and hyperparameter evaluation, this study used a separate model testing mechanism by passing the unseen test images to a separate model evaluation method. This proved effective for understanding the real model accuracies and finetuning accordingly.

Table 4: Variants of YOLOv8 and their specifications

YOLOV8 Variants	#Layers	#Params	Val Precision (%)	Test Precision (%)	Val Recall (%)	Test Recall (%)
YOLO8nano	225	3,011,043	79.5	62.8	63.0	72.3
YOLO8sml	225	11,135,987	74.1	60.9	67.4	74.5
YOLO8med	295	25,856,899	93.0	74.8	60.9	72.3
YOLO8lg	365	43,630,611	85.7	79.8	60.9	67.1
YOLO8x_lrg	365	68,153,571	84.1	66.3	57.4	62.8

The evaluation metrics used for the inspection of model training and testing are precision (P), recall (R), and mean average precision (mAP) of the predicted bounding box of defects. Precision counts the correct predictions from all prediction instances, whereas recall counts the correct predictions out of all actual predictions. The formula for precision and recall is given in Equation (1) and Equation (2).

$$P = \frac{TP}{TP + FP} \quad (1)$$

$$R = \frac{TP}{TP + FN} \quad (2)$$

In Equation (1), *P* indicates precision, whereas *TP* and *FP* donate true positives and false positives, respectively. In Equation (2), *R* represents recall, and *TP* and *FN* present true positives and false negatives.

Mean average precision (mAP) considers the classification and location accuracy, and mAP.50 and mAP.5:.95 are reported here. mAP.50 reports the average precision of the model when the intersection over the union (IOU) threshold is 50% or up between predicted and ground truth bounding boxes. mAP@.50:.95 is another metric that computes the average precision for IoU thresholds ranging from 0.50 to 0.95 in 0.05-step increments. This metric is more critical than mAP@.50 since it penalizes projections far from ground truths and displays the model's ability to precisely locate objects. A high mAP@.50:.95 value indicates greater performance in localization and classification.

6 Discussion of Results

As indicated in Table 4, the number of layers and trainable parameters are increasing from YOLOv8nano to larger models. The validation and test results indicate varying accuracies of these trained models on the timber defect dataset. The validation and test precision for YOLOv8 nano, YOLOv8 small (sml), YOLOv8 medium (med), YOLOv8 large (lrg), and YOLOv8 extra-large (x_lrg) are 79.5% and 62.8%, 74.1% and 60.9%, 93.0% and 74.8%, 85.7% and 79.8%, 84.1% and 66.3% accordingly. The validation and test recall for YOLOv8 nano, YOLOv8 small, YOLOv8 medium, YOLOv8 large, and YOLOv8 extra-large are 63.0% and 72.3%, 67.4% and 74.5%, 60.9% and 72.3%, 60.9% and 67.1%, 57.4% and 62.8% accordingly. These results indicate that YOLOv8 medium and YOLOv8 large are best suitable for the current detection task, as the precision is high for these models, which means the model is detecting true instances and recall is comparatively better, demonstrating fewer missed detections.

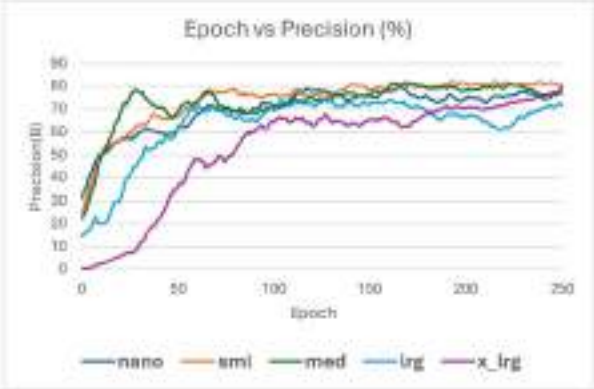


Figure 3: Training Precision in percentage

The training precision is shown in Figure 3, and it is visible that all models try to converge between 100-200 epochs with a slight increase in precision afterward, but the YOLOv8 large training curve indicates a loss in accuracy in later epochs, indicating that the model is still trying to learn new parameters due to its large size.

The training mAP.5 (B) and mAP.5:.95 (B) for all models in Figures 4 (a) and (b), respectively, indicate a similar pattern as precision(B) where models can achieve between 63% to 73% metrics for mAP.5 and between 35 % to 40% mAP.5-.95. These converged graphs show that models are neither over-fitting nor under-fitting, provided the toughness, small-scaled and complexity of the dataset.

Moreover, the final validation and test mAP.5 (B) and mAP.5:.95 (B) for all models are presented in Figure 5. This comparative graph indicates around 75% and 70% of validation and test mAP.5 (B) and around 45% and 38% achieve validation and test mAP.5:.95 (B). These values

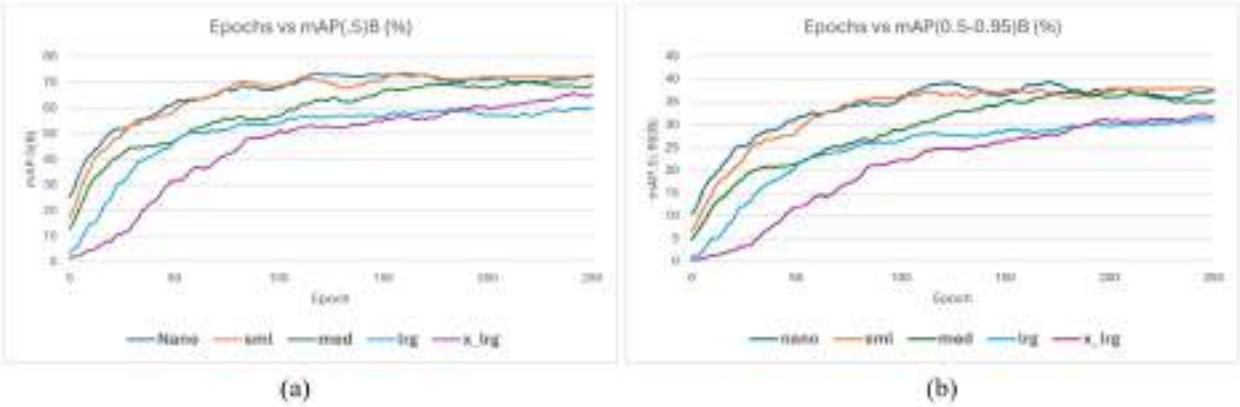


Figure 4: (a) Training mAP(.5)(B) in percentages (b) Training mAP(0.5-0.95)(B) in percentages

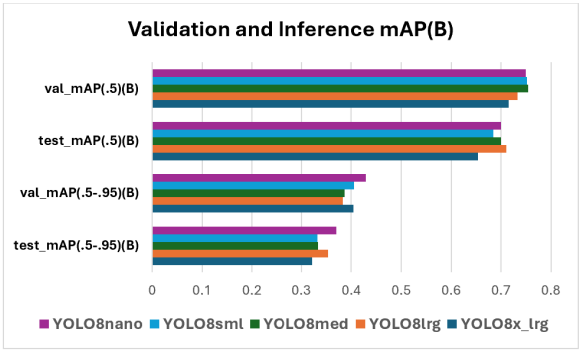


Figure 5: Validation and Inference results on YOLOv8 variants

demonstrate effective learning and consistently high capabilities, indicating that the models are generalized well on unseen data. Moreover, YOLOv8 medium and large are two modules that consistently achieved superior accuracies and precision, which confirms their ability to perform reliably in real-world scenarios.

Based on the collected results, YOLOv8 medium-trained weights are selected for model inference so that the images from one defective stud are passed to generate predictions, presented in Figure 6. The model generates predictions for the same defective studs differently when different angles of images are provided. The instances captured from 2 feet height generated a bounding box with high accuracy (a,b,c), but the top views 6 feet and 16 feet high indicate a gradual decrease in precision (d,e).

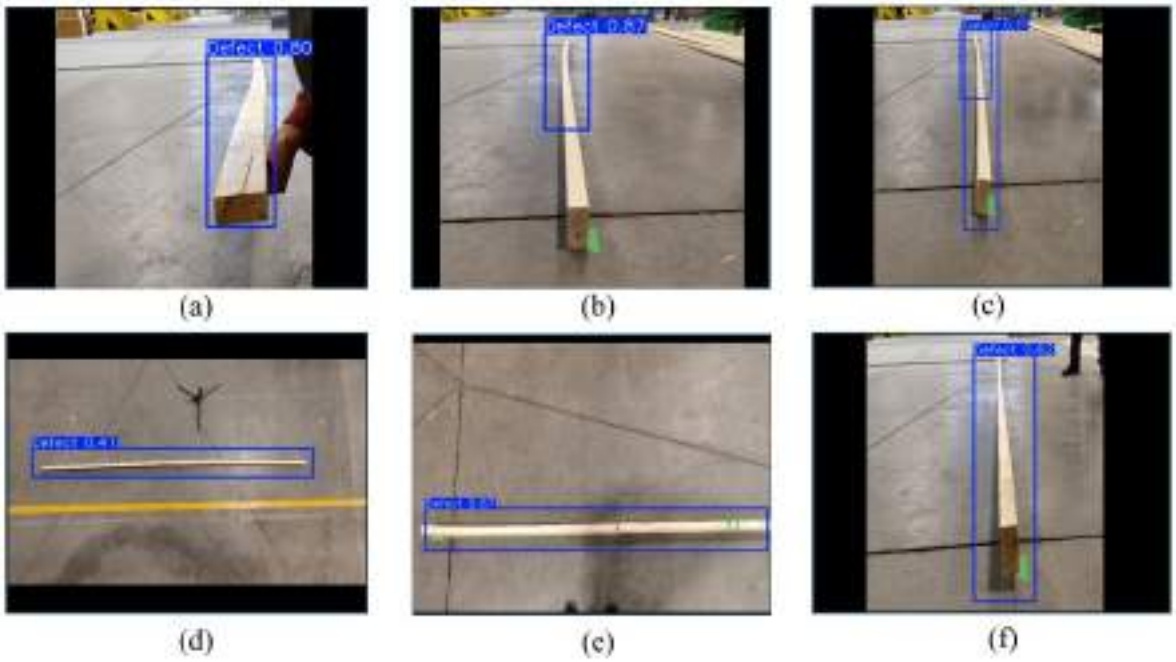


Figure 6: The difference in prediction accuracies based on different camera angles.

These observations provide valuable insight into the number of cameras, optimized camera installation location, and images needed for effective real-world predictions using the provided setup and methods.

As the number of defective studs in a working industrial setting is limited, so based on data availability, the number of defective and non-defective studs was adjusted to create a balanced trainable dataset, totalling 15 studs. The smaller datasets are highly prone to overfitting issues, compromising overall model authenticity and generalizability. To tackle this issue, a careful selection of hyperparameters and data augmentation was adopted while training with different modules of YOLOv8. In the future, this research can extend to collect and expand the dataset length to enhance accuracy and attain better results. Further, a performance comparison between other state-of-the-art and latest image detection training models can also be conducted.

7 Conclusion

This research discussed timber stud structural defect detection dataset preparation steps followed by the extensive model training on YOLOv8 variants, specifically focusing on detecting bends, twists, and wraps in timber studs used in wood house construction. The quantitative results indicated a balance between accuracy and computational efficiency, whereas YOLOv8 medium and YOLOv8 large demonstrated the high validation and test precision of 93.0%, and 74.8%, 85.7%, and 79.8%, respectively. The predictions generated provided knowledge about the best device placement and total frames for real-time prediction precision, and the predicted defect bounding boxes can be used for precise automated cutting of defective parts. This research will help reduce carbon emissions through resource utilization optimization and waste mitigation in the construction industry. Moreover, it advances computer vision and robotic automation in the Canadian wood construction industry. The proposed methods can be expanded to accommodate other types of timber defects, expand the dataset, and automate other pre- and post-manufacturing inspection steps.

References

[1] D. Li, Z. Zhang, B. Wang, C. Yang, and L. Deng. Detection method of timber defects based on target detection algorithm. *Measurement (Lond)*, 203:111987, 2022.

[2] Ultralytics. YOLOv8. On-line: https://github.com/ultralytics/ultralytics, Accessed: 25/12/2024.

[3] J. J. Cisneros-Gonzalez, A. Rasool, and R. Ahmad. Digital technologies and robotics in mass-timber manufacturing: a systematic literature review on construction 4.0/5.0. *Construction Robotics*, 8(2): 29, 2024.

[4] A. Rasool, G. R. Satsangee, L. Arickswamy, M. M. Ashfaq, and R. Ahmad. Winter-Safe Slip Prevention Rim for E-Scooter: Design to Production Lifecycle Analysis. In *1st International Conference on Industrial, Manufacturing, and Process Engineering (ICIMP-2024)*, page 88, Basel, Switzerland, 2024.

[5] J. J. Cisneros González, A. Rasool, Y. H. Chui, and R. Ahmad. Robotic cell design and simulation system for cross-laminated timber panel post-construction operations. In *Proceedings of the Transforming Construction with Off-site Methods and Technologies (TCOT) Conference*, pages 1–10, Fredericton, Canada, 2024.

[6] E. A. Holm, P. P. Choi, M. M. Nowell, and S. P. Lynch. "Overview: Computer Vision and Machine Learning for Microstructural Characterization and Analysis. *Metallurgical and Materials Transactions A*, 51(12):5985–5999, 2020.

[7] G. R. Satsangee, H. Al-Musaibeli, and R. Ahmad. A defect detection method based on YOLOv7 for automated remanufacturing. *Applied Sciences*, 14(13):5503, 2024.

[8] H. Z. Imam, H. Al-Musaibeli, Y. Zheng, P. Martinez, and R. Ahmad. Vision-based spatial damage localization method for autonomous robotic laser cladding repair processes. *Robotics and Computer-Integrated Manufacturing*, 80:102452, 2023.

[9] A. Rasool, M. M. Fraz, and S. Javed. Multiscale Unified Network for Simultaneous Segmentation of Nerves and Micro-vessels in Histology Images. In *Proceedings of the 2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2)*, pages 1–6, Islamabad, Pakistan, 2021.

[10] R. Wang, Y. Chen, F. Liang, B. Wang, X. Mou, and G. Zhang. BPN-YOLO: A Novel Method for Wood Defect Detection Based on YOLOv7. *Forests*, 15(7):1096, 2024.

[11] W. Meng and Y. Yuan. SGN-YOLO: Detecting Wood Defects with Improved YOLOv5 Based on Semi-Global Network. *Sensors*, 23(21):8705, 2023.

[12] R. Ehtisham, M. A. Khan, M. Sharif, S. L. Fernandes, and G. S. Siddiqui. Computing the characteristics of defects in wooden structures using image processing and CNN. *Automation in Construction*, 158:105211, 2024.

[13] C. Fan, Z. Zhuang, Y. Liu, Y. Yang, H. Zhou, and X. Wang. Bilateral Defect Cutting Strategy for Sawn Timber Based on Artificial Intelligence

Defect Detection Model. *Sensors*, 24(20):6697, 2024.

[14] M. Ji, W. Zhang, X. Diao, G. Wang, and H. Miao. Intelligent Automation Manufacturing for Betula Solid Timber Based on Machine Vision Detection and Optimization Grading System Applied to Building Materials. *Forests*, 14(7):1510, 2023.

[15] Y. Zhu, Z. Xu, Y. Lin, D. Chen, K. Zheng, and Y. Yuan. Surface defect detection of sawn timbers based on efficient multilevel feature integration. *Measurement Science and Technology*, 35(4):045401, 2024.

[16] H. Xi, R. Wang, F. Liang, Y. Chen, G. Zhang, and B. Wang. SiM-YOLO: A Wood Surface Defect Detection Method Based on the Improved YOLOv8. *Coatings*, 14(8):1001, 2024.

[17] J. Zhang, L. Bi, J. Lian and C. Guan. A Single-Trial Event-Related Potential Estimation Based on Independent Component Analysis and Kalman Smoother. *In Proceedings of the IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, pages 9–12, Auckland, New Zealand, 2018

[18] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. arXiv preprint arXiv:1311.2524, 2014.

[19] R. Girshick. Fast R-CNN. On-line: https://github.com/rbgirshick/, Accessed: 25/12/2024.

[20] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1137–1149, 2017.

[21] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *IEEE Computer Society*, pages 779-788, Las Vegas, USA, 2016.

[22] J. Redmon and A. Farhadi. YOLO9000: Better, faster, stronger. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6517–6525, Hawaii, USA, 2017.

[23] J. Redmon and A. Farhadi. YOLOv3: An Incremental Improvement. *arXiv preprint arXiv:1804.02767*, 2018.

[24] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv preprint arXiv:2004.10934*, 2020.

[25] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg. SSD: Single shot multibox detector. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 21–37, Amsterdam, Netherlands, 2016.

[26] J. Xu, H. Yang, Z. Wan, H. Mou, D. Qi, and S. Han. Wood Surface Defects Detection Based on the Improved YOLOv5-C3Ghost With SimAm Module," *IEEE Access*, 11:105281–105287, 2023.

[27] J. Zhou, J. Ning, Z. Xiang, and P. Yin. ICDW-YOLO: An Efficient Timber Construction Crack Detection Algorithm. *Sensors*, 24(13):4333, 2024.

[28] Tzutalin. LabelImg: Graphical Image Annotation Tool. On-line: https://github.com/tzutalin/labelImg, Accessed: 25/12/2024.

Comparative Study of Structure from Motion on Construction Site

Mingyun Kang[1] Sangmin Lee[1] Sebeen Yoon[1] and Taehoon Kim[1]

[1]Architecture Engineering, Seoul National University of Science and Technology, Republic of Korea

kmg4312@seoultech.ac.kr, lsm010714@seoultech.ac.kr, paul6452@seoultech.ac.kr, kimth@seoultech.ac.kr

Abstract

For progress monitoring in the construction industry, understanding the state of construction sites is crucial. However, traditional manual inspection methods are labor-intensive and time-consuming. To address these challenges, various methods for creating 3D models of job sites have been explored. Professional equipment such as LiDAR and laser scanners offer the most accurate means of generating point clouds (PCDs) and constructing 3D models. However, these tools are expensive, cumbersome, and often impractical for frequent use in dynamic construction environments. Recently, with the advancement of deep learning, 3D reconstruction techniques have been extensively studied and applied across various fields. Among these, Structure from Motion (SfM) stands out as a method capable of generating PCDs and estimating camera poses. Based on advancing capabilities of SfM, many research has been conducted to measure progress monitoring in construction field. However, most studies that have utilized SfM for progress monitoring have acquired a large number of images and ensured significant overlap in input data to enhance the robustness of the 3D model. While this approach provides a highly accurate 3D reconstruction, the image acquisition process itself introduces additional labor-intensive tasks. Therefore, this study aims to adhere to the fundamental nature of 3D reconstruction by evaluating the performance of various SfM models using only 26 images captured from a brief video recording at a construction site. The findings aim to evaluate the applicability of various SfM technologies with limited data, in real-world construction scenarios and finally provide insights into their potential and future directions.

Keywords –

Progress Monitoring, Structure from Motion, COLMAP, VGGSfM

1 Introduction

In the construction industry, progress monitoring is essential to ensure project efficiency, quality, and adherence to timelines. Traditional monitoring methods often rely on manual inspections and reports, which can be time-consuming, error-prone, and inconsistent. As a solution, 3D modeling techniques have gained attention for their ability to provide comprehensive information of construction sites [1, 2, 3].

To generate 3D models of construction sites, equipment such as LiDAR and laser scanners are widely used, offering the ability to produce highly accurate and dense point clouds (PCDs). For instance, Hu et al. utilized LiDAR to scan 3D surfaces of opencast quarries, enabling the analysis of slope stability. Similarly, Wang et al. employed laser scanners to rapidly perceive heavy construction equipment such as tower crane, improving safety and productivity in construction environments. Despite their accuracy and utility, these tools are expensive, bulky, require extensive training to operate, and involve significant time for scanning, making them less practical for dynamic and fast-paced construction site conditions [3, 4, 5].

These challenges have driven researchers to explore alternative methods for 3D reconstruction. With the advancement of deep learning and computer vision, techniques using 2D images to create 3D models have garnered significant interest [1]. While such methods may compromise lower accuracy compared to specialized equipment, their accessibility, ease of implementation, and reduced scanning time make them a promising alternative [1, 4]. Among these approaches, Structure from Motion (SfM) stands out as a photogrammetric technique capable of not only constructing 3D models from 2D images taken from varying perspectives, but also estimating camera poses which include their positions and orientations [6].

Using these strengths of SfM, research on progress monitoring based on SfM in the construction field has been actively conducted. However, these studies

focusing on the 3D reconstruction of actual construction sites, particularly interior environments, remain insufficiently explored. Additionally, achieving higher accuracy in SfM models often requires a large number of images and to be overlapped to some extend. Hence, acquiring such extensive image datasets in complex construction sites is challenging and may lead to even more cumbersome issues.

By utilizing image data captured within a very short time frame and comparing multiple SfM frameworks, this study aims to identify the challenges and opportunities of applying SfM in construction environments, thereby contributing to a deeper understanding of its potential in the industry.

2 Literature Review

2.1 Structure from Motion (SfM)

Structure-from-Motion (SfM) is a technique that builds 3D point clouds (structure) from cameras (motion). SfM is typically categorized into incremental and global approaches. In incremental SfM, the process begins by aligning an initial pair of images, followed by the gradual addition of new images. This method iteratively performs feature matching and bundle adjustment (BA), progressively constructing the 3D structure. In contrast, global SfM employs deep learning-based feature matching across multiple images, enabling a faster and more efficient reconstruction process. This approach estimates camera poses and reconstructs the 3D structure in a global, batch-wise manner rather than incrementally.

COLMAP [7] is one of the dominant incremental SfM tool in computer vision field due to its robustness and efficiency. It employs the Scale-Invariant Feature Transform (SIFT) algorithm, which ensures reliable feature matching even under image rotations and lighting variations. Furthermore, COLMAP uses the Ceres Solver-based BA to reduce accumulated errors, thereby enhancing reconstruction accuracy. As mentioned before, with the advancement of deep learning, global SfM methods also have been developed.

VGGSfM [8], one of the newest global SfM model, has been developed as an alternative to traditional incremental SfM methods. Unlike conventional approaches, VGGSfM jointly estimates all camera poses using a Transformer-based architecture, offering a simpler and more differentiable process compared to the combinatorial correspondence chaining step. Additionally, for BA, VGGSfM replaces the nondifferentiable Ceres Solver with the fully differentiable second-order Theseus Solver, further improving computational efficiency and scalability.

2.2 SfM in Construction Site

Although SfM cannot build 3D models as accurately as TLS or LiDAR, it has been widely studied due to its affordability, accessibility, and ability to perform spatial mapping to a certain extent. Ding et al. explored indoor scene understanding using 304 images captured with a widely used consumer-level camera and conducted 3D reconstruction using VisualSfM [5]. The strong 3D reconstruction capability significantly contributes to scene understanding, which has led to extensive research on its applications in progress monitoring.

Studies have also been conducted to evaluate COLMAP's powerful performance within the construction domain, emphasizing its potential applicability in this field. Keyvanfar et al. applied nine different 3D reconstruction models, including COLMAP, to visualize construction sites [9]. In their study, they utilized 138 images captured using an Unmanned Aerial Vehicle (UAV), ensuring a more robust 3D reconstruction by acquiring images with 60–80% overlap.

In addition to studies that solely utilized SfM for progress monitoring, research integrating SfM with other technologies has also been conducted. Mahami et al. and Han et al. employed SfM combined with Multi-View Stereo (MVS) to construct 3D models and align them with as-built Building Information Model (BIM) elements for progress monitoring [10, 11]. Additionally, some studies have leveraged SfM-based camera pose estimation to generate denser 3D models. For instance, Pal et al. developed an SfM-MVS-based 3D reconstruction pipeline to estimate camera parameters from 2D images and incorporated Neural Radiance Field (NeRF) and semantic segmentation to predict progress completion [1].

Although numerous studies have explored the application of SfM in the construction industry, many of these approaches have prioritized achieving more robust and dense 3D reconstructions by utilizing high-resolution cameras or capturing images with significant overlap to enhance feature matching. Consequently, rather than fully exploiting SfM's inherent advantages, these methods introduce constraints related to data acquisition, limiting their practical applicability.

3 Case Study

3.1 Overview

To conduct a 3D reconstruction study in a real construction environment, we visited a construction site located in Seongdong-gu, Seoul, and acquired image data. A specific room within the construction site was selected to compare the performance of different 3D reconstruction methods. Image acquisition was

performed using the Insta ONE X2 fisheye camera, which facilitated rapid data collection by capturing images in all directions with a single shot. The captured footage was converted into perspective images, and a total of 26 images from the selected room were used for the comparative analysis between COLMAP and VGGSfM. All of the converted images have resolution of 4320 × 2880. The workstation used for the case study was equipped with 32 GB of RAM, a Core i7-12700KF CPU, NVIDIA GeForce RTX 3080 Ti GPU and CUDA version 11.3 was utilized for GPU acceleration.

3.2 Evaluation Metrics

To quantitatively compare the results of 3D reconstruction, not only the number of 3D points and processing time but also three key metrics were used: mean track length (MTL), mean observations per image (MOI), and mean reprojection error (MRE). These metrics provide insights into the robustness, contribution, and accuracy of the reconstruction process [12].

MTL is the average number of images in which a 3D point is observed. It measures the frequency with which 3D points are visible across the set of input images, providing insight into the robustness of feature matching and triangulation in the reconstruction process.

$$\mathrm{MTL} = \frac{\sum_{i=1}^{N} L_i}{N} \tag{1}$$

Where N denotes total number of 3D points in reconstruction, L_i the number of images in which the i-th 3D point is observed.

MOI is the average number of 3D points observed per image. This metric evaluates how much each image contributes to the 3D reconstruction, indicating the distribution of feature matches across the dataset.

$$\mathrm{MOI} = \frac{\text{Total Observations}}{\text{Number of Images}} \tag{2}$$

MRE quantifies the average distance, in pixels, between the observed 2D points and their corresponding 2D projections from the 3D points. It serves as a key indicator of the accuracy of the reconstruction and the quality of the camera parameters.

$$\mathrm{MRE} = \frac{\sum_{i=1}^{N} |\boldsymbol{p_i} - \widehat{\boldsymbol{p}_i}|}{N} \tag{3}$$

Where N denotes total number of 2D points, p_i the observed 2D coordinates of the i-th point, $\hat{p}_i$ the reprojected 2D coordinates of the i-th point obtained from the estimated 3D point and camera parameters.

4 Result Analysis

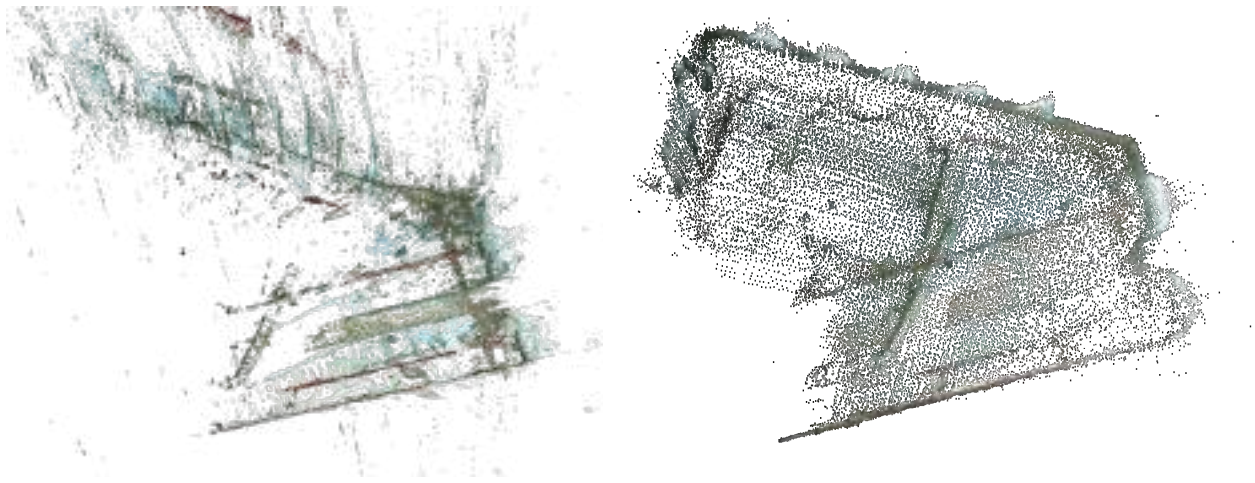
4.1 Quantitative analysis

The quantitative comparison of COLMAP and VGGSfM using the aforementioned metrics is summarized in the following Table 1.

Table 1. Summary of quantitative comparison of COLMAP and VGGSfM

Metrics	COLMAP	VGGSfM
Number of 3D points	13,765	**69,612**
Processing time	30m 33s	**10m 35s**
MTL	**1.22082**	0.66221
MOI	1,377.38	**1,993**
MRE	1.08028	**0.01852**

* Bold values indicate better performance



(a) COLMAP (b) VGGSfM

Figure 1 Top view perspective of 3D reconstruction model of COLMAP(a) and VGGSfM(b)

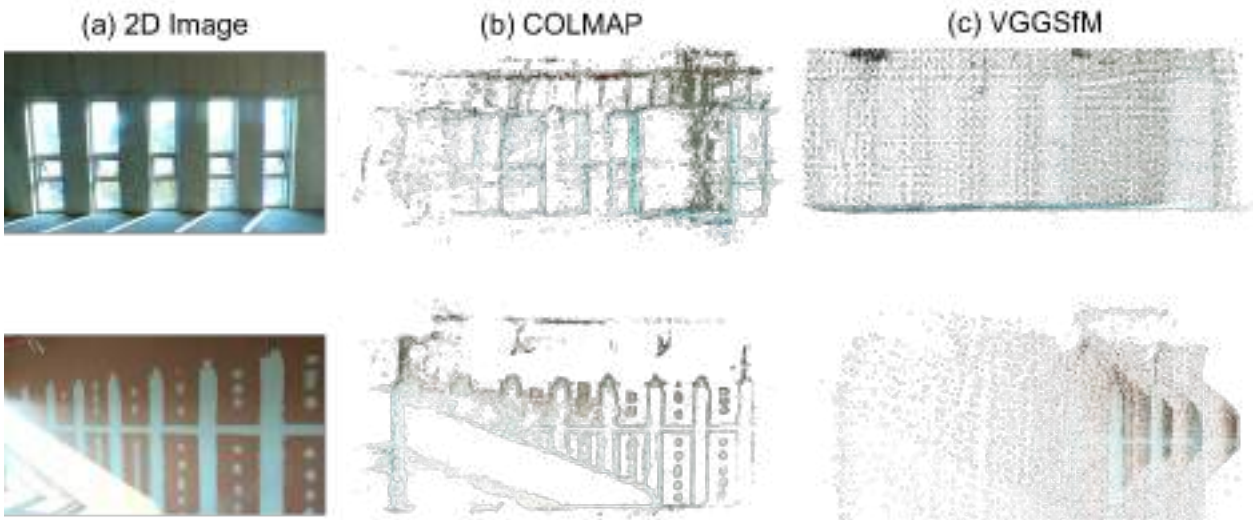


Figure 2 Reconstruction outputs of wall: (a) 2D image of wall (b) COLMAP (c) VGGSfM

The number of generated PCDs was significantly higher for VGGSfM, with 69,612 PCDs compared to 13,765 PCDs for COLMAP, a difference of approximately 55,000 points. In terms of processing time, VGGSfM was approximately 20 minutes faster than COLMAP, completing the task in 10 minutes and 35 seconds compared to COLMAP's 30 minutes and 33 seconds.

While MTL was 0.6 higher for COLMAP, for MOI, VGGSfM showed a higher value by about 600, indicating that images contributed more effectively to PCD generation. Moreover, the MRE for VGGSfM was about 1 lower than COLMAP, suggesting that VGGSfM's PCDs better explain the observed data.

Overall, VGGSfM demonstrated superior performance compared to COLMAP in all metrics except for MTL, which can be attributed to dataset-specific characteristics rather than model performance.

4.2 Qualitive analysis

Figure 1 presents the results of 3D reconstruction projected from a top-view perspective. In the case of COLMAP (Figure 1(a)), the room's shape appears inconsistent and scattered compared to VGGSfM (Figure 1(b)). Additionally, the floor area in the middle of the room is poorly reconstructed in COLMAP, whereas VGGSfM provides a more complete representation.

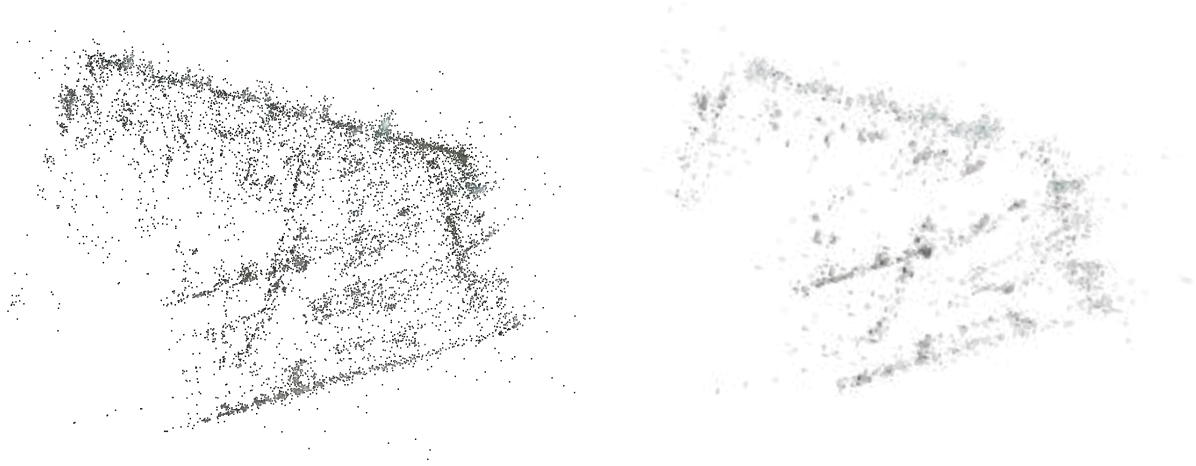
Figure 2 compares the reconstruction outputs of a wall using a 2D image (Figure 2(a)) as input, with results from COLMAP (Figure 2(b)) and VGGSfM (Figure 2(c)). As shown, COLMAP effectively captures distinctive lines and patterns, demonstrating its strength in feature-specific reconstruction. In contrast, VGGSfM produces a more uniformly distributed PCDs, reflecting a balanced and comprehensive reconstruction.

On the other hand, both models share a common limitation compared to laser scanners, as they rely solely on 2D input data for reconstruction. This constraint results in missing information in areas with bright occlusion or insufficient feature data.

5 Discussion

In this section, we compare not only COLMAP and VGGSfM but also various other SfM models. Meshroom [13] is an incremental SfM model that offers ease of use through a graphical user interface (GUI), whereas OpenMVG [14] is a global SfM model based on a command-line interface (CLI). The same set of 26 images used in the case study was used as input data for both models.

The processing time for Meshroom was approximately 21 seconds, while OpenMVG took 19 seconds. The number of generated 3D points was 8,587 for Meshroom and 3,174 for OpenMVG. Although both models demonstrated significantly faster processing times compared to COLMAP and VGGSfM, their reconstructed 3D models were insufficient for scene understanding, as shown in Figure 3. On the other hand, considering the perspective of SfM's camera pose estimation accompanying with fast processing speed, these models can serve as effective preprocessing tools for generating dense 3D models using methods such as MVS or NeRF. Their ability to provide camera positions makes them suitable for applications where precise pose estimation is essential for subsequent reconstruction processes.



(a) Meshroom (b) OpenMVG

Figure 3 Top view perspective of 3D reconstruction model of Meshroom(a) and OpenMVG(b)

Table 2. Summary of Meshroom and OpenMVG

Metrics	Meshroom	OpenMVG
Number of 3D points	**8,587**	3,174
Processing time	21s	**19s**

* Bold values indicate better performance

6 Conclusion

In this study, we explored the applicability of SfM models, in real construction environments. For this purpose, we compared two SfM models: COLMAP, one of the most widely used SfM models, and VGGSfM, a more recently developed framework. Comparison revealed that VGGSfM outperformed COLMAP in terms of the number of 3D points and processing time. Additionally, VGGSfM demonstrated superior performance in all metrics except for MTL.

Qualitative comparisons also highlighted VGGSfM's advantages. VGGSfM more accurately depicted the arrangement of windows and walls, whereas COLMAP exhibited scattered and poorly defined spaces. In contrast, VGGSfM effectively delineated a finite outline of the room, successfully reconstructing the enclosed space.

In discussion section, we also utilized two more SfM model of Meshroom and OpenMVG. These SfM has highly faster processing time of no more than 1 min, however, their number of 3D points were in proportion to the processing time. This result indicates that these models are not adequate for 3D reconstruction of scene understanding, on the other hand, in the perspective of preprocessing for building dense 3D models, they can serve camera poses of images in a short time.

Besides, this study remains certain limitations. Since SfM relies solely on 2D images as input, bright occlusions can lead to a lack of PCDs. Also, the 3D reconstruction was conducted for a single room within the construction site. When attempting to reconstruct multiple rooms, challenges arise due to the inability to extract overlapping features from images capturing transitions through openings, such as doors. Consequently, reconstructing an entire construction site remains a challenge. Future studies will focus on overcoming these limitations by developing methods to integrate reconstructions of multiple rooms, ultimately facilitating a comprehensive understanding of construction sites and contributing to progress monitoring efforts.

7 Acknowledgements

This research was supported by a grant (RS-2022-00143493) from Digital-Based Building Construction and Safety Supervision Technology Research Program funded by Ministry of Land, Infrastructure and Transport of Korean Government.

References

[1] A. Pal, J. J. Lin, S. H. Hsieh, and M. Golparvar-Fard, “Activity-level construction progress monitoring through semantic segmentation of 3D-informed orthographic images,” *Autom Constr*, vol. 157, Jan. 2024, doi: 10.1016/j.autcon.2023.105157.

[2] T. Wang and V. J. L. Gan, “Enhancing 3D reconstruction of textureless indoor scenes with IndoReal multi-view stereo (MVS),” *Autom*

Constr, vol. 166, Oct. 2024, doi: 10.1016/j.autcon.2024.105600.

[3] Y. Jeon, A. S. Kulinan, D. Q. Tran, M. Park, and S. Park, "NeRF-Con : Neural Radiance Fields for Automated Construction Progress Monitoring," Jun. 2024. doi: 10.22260/ISARC2024/0149.

[4] Y. Xu and J. Zhang, "UAV-based bridge geometric shape measurement using automatic bridge component detection and distributed multi-view reconstruction," *Autom Constr*, vol. 140, Aug. 2022, doi: 10.1016/j.autcon.2022.104376.

[5] Y. Ding, X. Zheng, Y. Zhou, H. Xiong, and J. Gong, "Low-cost and efficient indoor 3D reconstruction through annotated hierarchical Structure-from-Motion," *Remote Sens (Basel)*, vol. 11, no. 1, Jan. 2019, doi: 10.3390/rs11010058.

[6] N. Jiang, Z. Cui, and P. Tan, "A Global Linear Method for Camera Pose Registration," in *2013 IEEE International Conference on Computer Vision*, IEEE, Dec. 2013, pp. 481–488. doi: 10.1109/ICCV.2013.66.

[7] J. L. Schonberger and J.-M. Frahm, "Structure-from-Motion Revisited," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2016, pp. 4104–4113. doi: 10.1109/CVPR.2016.445.

[8] J. Wang, N. Karaev, C. Rupprecht, and D. Novotny, "Visual Geometry Grounded Deep Structure From Motion," Dec. 2023, [Online]. Available: http://arxiv.org/abs/2312.04563

[9] A. Keyvanfar, A. Shafaghat, and M. S. F. Rosley, "Performance comparison analysis of 3D reconstruction modeling software in construction site visualization and mapping," *International Journal of Architectural Computing*, vol. 20, no. 2, pp. 453–475, Jun. 2022, doi: 10.1177/14780771211066876.

[10] H. Mahami, F. Nasirzadeh, A. H. Ahmadabadian, and S. Nahavandi, "Automated progress controlling and monitoring using daily site images and building information modelling," *Buildings*, vol. 9, no. 3, 2019, doi: 10.3390/buildings9030070.

[11] K. Han, J. Degol, M. Golparvar-Fard, and A. M. Asce, "Geometry-and Appearance-Based Reasoning of Construction Progress Monitoring," 2017, doi: 10.1061/(ASCE).

[12] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[13] C. Griwodz *et al.*, "AliceVision Meshroom," Association for Computing Machinery (ACM), Jun. 2021, pp. 241–247. doi: 10.1145/3458305.3478443.

[14] P. Moulon, P. Monasse, R. Perrot, and R. Marlet, "OpenMVG: Open Multiple View Geometry," vol. 10214, B. Kerautret, M. Colom, and P. Monasse, Eds., in Lecture Notes in Computer Science, vol. 10214. , Cham: Springer International Publishing, 2017, pp. 60–74. doi: 10.1007/978-3-319-56414-2_5.

Automatic Reconstruction of Aligned Indoor Building and Façade Elements from Point Cloud Data

Mudan Wang[1], Yuandong Pan[1], Anson, T.K. Chan[1], Sven Auerswald[2], Seyedehmaedeh Aghili[2], Ioannis Brilakis[1]

[1] Department of Engineering, University of Cambridge, 7a JJ Thomson Ave, Cambridge CB3 0FA, United Kingdom
[2] Sustainable Systems Engineering Building Systems Technology, Fraunhofer Institut for Solar Energy Systems ISE Heidenhofstrasse 2, D-79110 Freiburg, Germany
mw923@cam.ac.uk, yp296@cam.ac.uk, tkc30@cam.ac.uk, sven.auerswald@ise.fraunhofer.de, seyedehmaedeh.aghili@ise.fraunhofer.de, ib340@cam.ac.uk

Abstract

The 3D building reconstruction from point clouds and images has significant applications in the field of built environment. Previous studies have investigated the methods of the indoor environment reconstruction and other studies explored the building façade reconstruction. However, limited studies have been conducted on aligningindoor and outdoor elements from point cloud data, which is critical for the building energy analysis and retrofitting plan during the operation stage. This research aims to develop a method that can automatically align the indoor building and façade elements from point cloud data. The results can enable the integration of buildings through the identified openings on both indoor and façade. The proposed method is validated on a real case and the experiment results show the effectiveness of the method. This research contributes to 3D reconstruction of the buildings with more details for automating building energy analysis and retrofitting design.

Keywords –

Point Cloud; Indoor Environment; Building Façade; 3D Reconstruction; Opening Detection

1 Introduction

Building reconstruction from point clouds and images has significant applications in the field of built environment, such as facility management, building performance analysis, and retrofitting design [1, 2]. Previous studies have investigated the methods of the indoor environment reconstruction and other studies explored the building façade reconstruction. For example, [3] proposed methods to reconstruct the indoor structural elements and generate geometric digital twin models. [4] presented method for the extraction of façade features from point cloud data. These methods are mainly focused on either indoor or façade data. Limited studies have been conducted to link the indoor and façade and integrate the building point cloud data as a whole.

3D laser scanning has been leveraged to capture the accurate as-is conditions of buildings, providing high-resolution spatial data that supports advanced modeling [5]. The BIM model can be then generated with processed point cloud clusters, which is known as the scan-to-BIM process [6]. Many studies have explored in this field to quickly and automatically generate BIM models from the point cloud data. For example, [7] presents the mobile laser scanning in reconstructing the indoor building environment and conduct semantic point cloud segmentation to reconstruct the buildings.

Despite advances in point cloud and image processing, as well as BIM-based workflows, detailed geometrical representations of indoor buildings often lackconnecting the indoor environment with building envelope enrichment. The integration of geometrical and spatial relationships between indoor spaces and façades into digital twin models remains a significant challenge. Achieving such integration is crucial for developing a more holistic understanding of buildings, enabling applications such as automated building energy analysis and retrofitting design. This research addresses this gap by proposing a novel method to align indoor building and façade elements from point cloud data, thereby contributing to more detailed and interconnected 3D building reconstructions.

2 Background

Previous studies have developed automatic methods for reconstructing indoor environments, primarily focusing on structural elements. For example, [3] proposed a method for creating geometric digital twins of buildings using a void-growing approach, which effectively identifies walls, ceilings, windows, floors,

and doors. Similarly, [8] reconstructed indoor environments from point clouds collected by mobile laser scanners. Structural elements such as doors, windows, walls, floors, and ceilings were extracted based on geometric and semantic features. Furthermore, [9] presented an automatic approach for segmenting indoor point clouds and performing parametric reconstruction of structural elements, such as walls and floors. In summary, the segmentation of indoor point clouds has been extensively explored, with numerous algorithms and approaches proposed for identifying structural elements.

Other studies have investigated building façade reconstruction from point cloud data. For example, [4] identified openings in façades made of reflective materials and estimated their geometric properties and locations. [10] introduced a large-scale 3D point cloud dataset of building façades for semantic segmentation, enabling the identification of façade features. Additionally, [11] developed a method for generating building façade geometry from point clouds and implemented an automatic approach for building energy performance simulation. In summary, previous research has explored the extraction of building façades from point cloud data, with a focus on identifying façade openings.

The 3D reconstruction of the physical environment into a digital format enables multiple use cases, such as VR-based room tour, distant visualization of tourist spots and conservation of dilapidated cultural heritage sites. However, several challenges are faced when aligning outdoor and indoor point cloud data (PCD). Firstly, it is difficult to find common overlapping between indoor and outdoor environments since they are often disconnected [12-14]. Another challenge is the significant lighting variations when transitioning between outdoor and indoor environments [15, 16]. Traditionally, geometric models are constructed by laser scanning and photogrammetry. For example, Scianna and La Guardia [17] re-created the 3D model of the Manfredonic Castle of Mussomeli in Sicily leveraging the photos collected from drones. Despite the generation of the castle in 3D format, it focused solely on the external layout of the castle and had not yet connected to the internal castle features due to complexities and technical barriers. Kaminsky, Snavely [18] presented a preliminary algorithm that aligned indoor and outdoor structures observed in satellite photos and floor plans. Geotags were then assigned to the photos to match their locations.

After detecting common areas between indoor and outdoor elements, rough estimations were computed to match the scenes using an objective function. In order to match the indoor and outdoor environments, Yang, Fang [19] put forward a Semantic-Geometric Descriptor (SGD) that utilized both semantic and spatial information obtained from PCD. An indoor and outdoor alignment framework was presented after conducting empirical experiments with their experimental PCD and public PCD. It showcased the possibilities to align both indoor and outdoor areas by some shared building features, such as windows and doors which enhanced the accuracy and speed. To tackle slow and error-prone manual generation of as-built BIM models, Tang, Huber [20] reviewed the state-of-the-art as-built BIM model generation methods in recent years. Object recognition, parametric modelling and relationship modelling are the most common methodologies for as-built BIM automation. Although significant works have been carried out to reconstruct the indoor room layout inside BIM, there is still a gap to integrate the indoor room layout into the outdoor environment. Cohen, Schönberger [15] exploited the semantic information from detecting windows in various scenes by a pre-trained classifier. Separate window projections were computed altogether to enhance consistency and the results were evaluated by a cost function in their 3D reconstruction.

In conclusion, limited studies have been conducted on aligning indoor and outdoor elements from the point cloud data, which is critical for the building energy analysis and retrofitting plan during the operation stage. This research aims to develop a method that can automatically align the indoor building and façade elements from point cloud data.

3 Methodology

This research proposes an automatic openings detection and align the indoor and outdoor together from point cloud data. This alignment is based on the openings (e.g., windows and doors) between outdoor and indoor point clouds.

3.1 Wall Segmentation

The input of the proposed method consists of façade point cloud data and the indoor point cloud data. The first step is to identify the walls and separate the walls into individual planes. The region growing and RANSAC algorithms are adopted to segment the point cloud and identify the walls [21]. Since there are typically more walls in the indoor environment, the segmented façade walls are used as a reference to select the corresponding walls in the indoor environment PCD. The walls, after filtering, are then used for further detection of openings. This process is conducted for both the indoor and outdoor data. After the processing, the input point cloud data are processed to segmented walls with the openings.

3.2 Openings Detection

The second step is to detect the openings on the walls for both indoor and outdoor point cloud data. Compared

with the indoor environment, the openings on the façade are more obvious than the openings in the indoor environment [4]. This is because there is more furniture in the indoor environment and can cause more occlusions of the point cloud. This makes it more difficult to identify openings in indoor environments.

Therefore, the proposed method starts by identifying the openings from the façade point cloud data, after the opening detection and the configuration of the openings on the façade, the results will be used to assist the refinement detection results of the openings detection in the indoor environment. The underlying principle of the opening detection is that openings appear as areas with no or very sparse points. Therefore, this research uses grid representation to detect the boundary of the openings.

To improve the accuracy of indoor environment data processing, a key step is to identify the openings from indoor and outdoor point cloud data. Since walls have a certain thickness, their indoor and outdoor surfaces represent two sides of the same structural element. The matching of indoor and outdoor point cloud data is based on their geometry features, such as the normal vector of the wall, and the coordinate position [19].

3.3 Openings Alignment

The final step is to align the openings on the same wall from indoor and outdoor point cloud data. Therefore, the indoor and outdoor building elements can be accurately matched. This aligned point cloud data can be further used for the 3D building reconstruction. Figure 1 presents the pipeline of the alignment of the indoor building and façade elements from point cloud data.

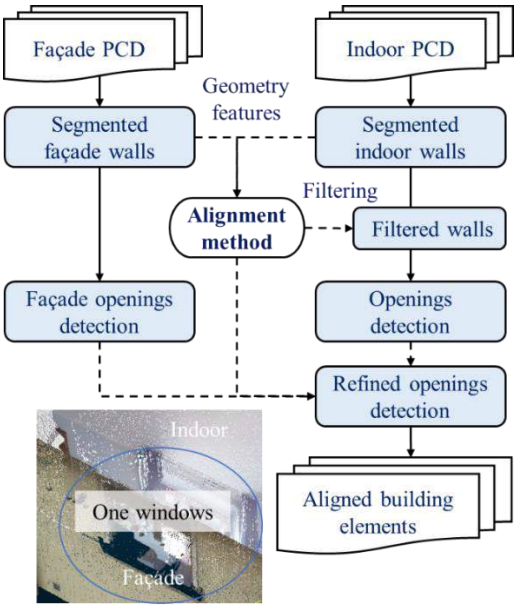


Figure 1. Pipeline of the proposed method

4 Experiments and Results

4.1 Case Background

The data used in this research is the point cloud data of the house building in the European. The materials used to build the building constructed in 1980 are concrete for the structure and bricks for the facades combined with windows in aluminum. The retrofitting proposed for the building is aimed to provide renewable energy to cover all the energy requirements from a single family, to reduce the energy demand acting in the insulation of the envelope and improving the air quality/heating/cooling by including a new heating/cooling/ ventilation system in the facade. Improved air quality and comfort in the different spaces will be the main objectives of the proposed. The project will retrofit completely the envelope of the building. Apart from this, a new ventilation/heating/cooling system will be integrated in the façade and connected to the heat pumps with the possibility to heat/cool the building depending on the requirements. The first floor of the case is used for the experiments to validate the effectiveness of the proposed method.

The point cloud data is collected by using the Leica BLK360 imaging scanner [22]. The registration of point clouds collected from multiple stations was done with the Leica Cyclone REGISTER 360 PLUS [23].

4.2 Indoor Wall Segmentation

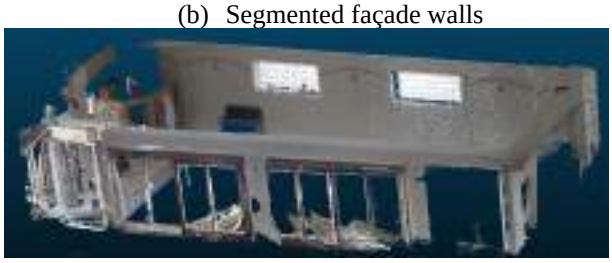
After the region growing and RANSAC algorithms, the walls in the indoor environment PCD and façade PCD are segmented into individual and separate point clusters, which represent the walls of the building. The results of the segmented walls for the two PCD datasets are presented in the Figure 2. Figure 2a presents the original point cloud data of the ground floor. Figure 2b shows the segmentation results of the façade walls while Figure 2c shows the results of the indoor walls after segmentation. For the indoor environment PCD, there are more walls than the façade datasets. After the filtering process, the filtered walls that are connected to the façade are presented in Figure 2d.



(a) Original point cloud data



(b) Segmented façade walls



(c) Segmented indoor walls



(d) Filtered indoor walls

Figure 2. Segmented walls. (a). Original point cloud data, (b). Segmented façade walls, (c) segmented indoor walls, (d) Filtered indoor walls.

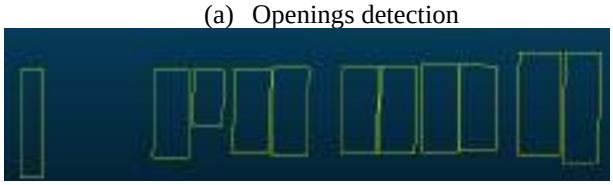
The results of the point cloud segmentation show that the indoor segmentation resulted in more walls as it has more rooms inside while the façade was an envelope of the building point cloud. By comparing the normal vectors of the wall clusters using RANSAC [24], the walls with similar normal vectors will be grouped as the same wall and will be remained. Therefore, the rest walls were filtered and removed. Only the walls that can match with the façade point cloud were further processed for the opening detection.

4.3 Façade Openings Detection

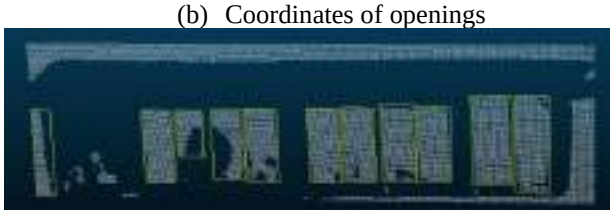
After the wall identification, the openings of the façade are detected using the opening detection method. The results of the opening detection of the façade are given in figure 3. Figure 3a presents the point cloud data for opening detection. Figure 3b shows the results of the detected openings. Figure 3c presents the geometry and the position of the openings. The results are obtained in the rectangle format, which contain the four corner coordinates of the openings. Therefore, the location and the geometry information can be computed with the rectangle information representing the openings.



(a) Openings detection



(b) Coordinates of openings



(c) Detected openings

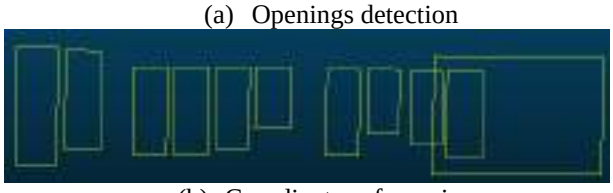
Figure 3. Façade openings detection

4.4 Indoor Openings Detection

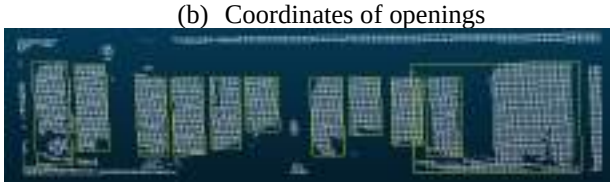
Similarly, the openings on the indoor point cloud can be obtained. The results of the opening detection of the façade are given in figure 4. Figure 4a presents the point cloud data for opening detection. Figure 4b shows the results of the detected openings. Figure 4c presents the geometry and the position of the openings.



(a) Openings detection



(b) Coordinates of openings



(c) Detected openings

Figure 4. Indoor openings detection

4.5 Aligned Building Elements

The last step is to align the results of the openings on the indoor openings detection with the detection results of façade point cloud. Figure 5 shows the alignment

results of the openings on both point cloud. The detected openings are aligned based on their coordinates of the rectangles representing the openings. The red points represent openings detected in the façade point cloud, and white points correspond to openings identified in the indoor point cloud. The alignment is performed based on the similarity of detected openings, considering both the least distance between corresponding openings and their overall layout. By matching openings with minimal spatial deviation and maintaining consistency in their distribution, the method ensures accurate geometric integration of indoor and façade elements.



Figure 5. Aligned openings in indoor and façade point clouds

4.6 Discussions

The detection results presented in Figures 3 and 4 highlight key issues that contribute to errors and reduced accuracy. Notably, in Figure 4c, the largest rectangle detected is not a window. This error is attributed to occlusion in the point cloud, which occurred during data collection in the indoor environment. Occlusion is a common challenge in such settings, as furniture and other building elements obstruct the view of certain features, leading to incomplete or misinterpreted data [25].Improved noise filtering techniques, such as adaptive thresholding or advanced denoising algorithms, can help retain useful points while minimizing the impact of noise [26]. Additionally, data augmentation methods, including multi-view point cloud merging and the integration of photogrammetry with LiDAR, can enhance the reconstruction of occluded areas [27].

Another significant issue arises from the reflection points of glass in windows. These reflections create ambiguities in the data and are almost unavoidable when capturing point cloud data of windows. The reflective surfaces interfere with accurate feature extraction and increase detection errors. Machine learning-based segmentation methods could be trained to differentiate between real structural features and reflection artifacts in windows [28].

In addition, the inherent noise within the point cloud data further impacts detection accuracy. While algorithms designed to filter out noise can enhance overall data quality, they may inadvertently remove some useful points. This, in turn, affects the detection of openings and reduces the geometric accuracy of identified features. It is challenging to preserve essential data while removing noise.

Addressing these issues requires enhanced data preprocessing methods, improved algorithms for handling occlusions in point cloud data, and techniques to minimize the impact of reflective surfaces during data acquisition [29]. Deep learning techniques, such as convolutional neural networks and transformer-based segmentation models, have shown promising results in point cloud processing. Examples such as KPConv [30] and Point Net [31] have been widely used for point cloud segmentation and feature extraction.

The point cloud data used in this study were collected using a terrestrial laser scanner, which provides higher data quality and accuracy and is recommended for data collection when applying this method. However, while this study successfully demonstrates the feasibility of detecting openings using point cloud data, the façade surface area analyzed is relatively small and lacks significant geometric variations. Since the proposed method focuses on identifying the presence of openings rather than reconstructing their precise shape, it has the potential to be extended to a wider range of building structures. To ensure its generalizability, further validation across diverse building types, including those with complex façade geometries and irregular opening configurations, is necessary.

Although the case study validates the proposed method's feasibility, the lack of a quantitative comparison with existing point cloud alignment methods remains a limitation. A systematic performance evaluation would provide a more comprehensive assessment of the method's effectiveness. Future research will address this by incorporating quantitative evaluations, specifically assessing accuracy and computational efficiency, to strengthen the validation and facilitate comparisons with alternative approaches. Additionally, expanding the range of case studies will be essential to evaluate the method's scalability and reliability under different architectural and environmental conditions, further enhancing its applicability to real-world scenarios.

5 Conclusions

This research introduces a method for automatically aligning indoor building elements with façade features using point cloud data. The approach emphasizes the identification and integration of openings, such as windows and doors, to seamlessly connect indoor and outdoor geometries. The validation conducted through a real-world case study highlights the effectiveness of the proposed method.

The results demonstrate that façade point cloud data effectively improves indoor opening detection and

achieves accurate alignment, significantly contributing to detailed 3D building reconstruction. This enhanced reconstruction not only improves the representation of building geometry but also supports advanced applications such as automated building energy analysis and retrofitting design.

However, there are still some limitations, including occlusions, noises, and reflective surfaces, which are common challenges in indoor point cloud data collection and processing. Future work will focus on refining the algorithms to better handle noise and occlusions, particularly in complex indoor environments, and further minimizing the impact of reflective surfaces during data acquisition. Additionally, integrating advanced preprocessing techniques could enhance detection accuracy and expand the applicability of the method to a wider range of building scenarios.

6 Acknowledgement

The authors acknowledge the funding received from the European Union's Horizon Europe research and innovation program under grant agreement No 101079961 (AEGIR project), and the Marie Sklodowska-Curie grant agreement No 101034337.

References

[1] Hamidreza, A., et al., Enhancing Building Maintenance Efficiency Through BIM-LiDAR Integration, in Proceedings of the 2024 European Conference on Computing in Construction. 2024, European Council on Computing in Construction.

[2] Sanzana, M.R., et al., Application of deep learning in facility management and maintenance for heating, ventilation, and air conditioning. Automation in Construction, 2022. 141: p. 104445.

[3] Yuandong, P., et al., 3D deep-learning-enhanced void-growing approach in creating geometric digital twins of buildings. Proceedings of the Institution of Civil Engineers - Smart Infrastructure and Construction, 2023. 176(1): p. 24-40.

[4] Mudan, W., et al., Automatic Extraction of Building Features for Building Façades Based on Laser Scanning Technology, in Proceedings of the 2024 European Conference on Computing in Construction. 2024, European Council on Computing in Construction.

[5] Wang, M., et al., A Systematic Review of Digital Technology Adoption in Off-Site Construction: Current Status and Future Direction towards Industry 4.0. Buildings, 2020. 10(11): p. 204.

[6] Wang, Q., J. Guo, and M.-K. Kim, An Application Oriented Scan-to-BIM Framework. Remote Sensing, 2019. 11(3): p. 365.

[7] Hu, D., V.J.L. Gan, and C. Yin, Robot-assisted mobile scanning for automated 3D reconstruction and point cloud semantic segmentation of building interiors. Automation in Construction, 2023. 152: p. 104949.

[8] Cui, Y., et al., Automatic 3-D Reconstruction of Indoor Environment With Mobile Laser Scanning Point Clouds. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2019. 12(8): p. 3117-3130.

[9] Martens, J. and J. Blankenbach, VOX2BIM+ - A Fast and Robust Approach for Automated Indoor Point Cloud Segmentation and Building Model Generation. PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science, 2023. 91(4): p. 273-294.

[10] Su, Y., et al., DLA-Net: Learning dual local attention features for semantic segmentation of large-scale building facade point clouds. Pattern Recognition, 2022. 123: p. 108372.

[11] O'Donnell, J., et al., LiDAR point-cloud mapping of building façades for building energy performance simulation. Automation in Construction, 2019. 107: p. 102905.

[12] Le, S., et al., Door and Window Detection in 3D Point Cloud of Indoor Scenes. Journal of Computer-Aided Design & Computer Graphics, 2019. 31(9): p. 1494-1501.

[13] Lu, T., et al., A novel methodology for the path alignment of visual SLAM in indoor construction inspection. Automation in Construction, 2021. 127: p. 103723-103723.

[14] Zhang, R. and A. Zakhor. Automatic identification of window regions on indoor point clouds using LiDAR and cameras. 2014. WACV 2014.

[15] Cohen, A., et al. Indoor-Outdoor 3D Reconstruction Alignment. in Computer Vision – ECCV 2016. 2016. Cham: Springer International Publishing.

[16] Ochmann, S., et al., Automatic reconstruction of parametric building models from indoor point clouds. Computers & Graphics, 2016. 54: p. 94-103.

[17] Scianna, A. and M. La Guardia, Survey and Photogrammetric Restitution of Monumental Complexes: Issues and Solutions—The Case of the Manfredonic Castle of Mussomeli. Heritage, 2019. 2(1): p. 774-786.

[18] Kaminsky, R.S., et al. Alignment of 3D point clouds to overhead images. 2010.

[19] Yang, Y., et al., Indoor–Outdoor Point Cloud Alignment Using Semantic–Geometric Descriptor. Remote Sensing, 2022. 14(20): p. 5119.

[20] Tang, P., et al., Automatic reconstruction of as-built building information models from laser-scanned point clouds: A review of related techniques. Automation in Construction, 2010. 19(7): p. 829-

843.
[21] Wang, M., et al., Onsite Quality Check for Installation of Prefabricated Wall Panels Using Laser Scanning. Buildings, 2021. 11(9): p. 412.
[22] Leica BLK360 Imaging Laser Scanner. Leica BLK360 Imaging Laser Scanner. [cited 2024 07 May]; Available from: https://leica-geosystems.com/products/laser-scanners/scanners/blk360.
[23] Leica Geosystems. Leica Cyclone REGISTER 360 PLUS - 3D Laser Scanning Point Cloud Registration Software,. [cited 2024 07 May]; Available from: https://leica-geosystems.com/products/laser-scanners/software/leica-cyclone/leica-cyclone-register-360.
[24] Wang, M., et al., A Streamlined Laser Scanning Verticality Check Method for Installation of Prefabricated Wall Panels. Journal of Construction Engineering and Management, 2024. 150(11): p. 04024159.
[25] Park, J., et al., Deep Learning–Based Automation of Scan-to-BIM with Modeling Objects from Occluded Point Clouds. Journal of Management in Engineering, 2022. **38**(4): p. 04022025.
[26] Ding, S., et al., A noise-reduction algorithm for raw 3D point cloud data of asphalt pavement surface texture. Scientific Reports, 2024. **14**(1): p. 16633.
[27] Zhang, J. and X. Lin, Advances in fusion of optical imagery and LiDAR point cloud applied to photogrammetry and remote sensing. International Journal of Image and Data Fusion, 2017. **8**(1): p. 1-31.
[28] Pan, Z., et al., Image-based surface scratch detection on architectural glass panels using deep learning approach. Construction and Building Materials, 2021. **282**: p. 122717.
[29] Berger, M., et al., A Survey of Surface Reconstruction from Point Clouds. Computer Graphics Forum, 2017. **36**(1): p. 301-329.
[30] Thomas, H., et al. Kpconv: Flexible and deformable convolution for point clouds. in Proceedings of the IEEE/CVF international conference on computer vision. 2019.
[31] Qi, C., et al., PointNet++: Deep Hierarchical Feature Learning on, in NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017. p. 5105-5114.

Microscale digital twinning of rubble stone masonry walls using robotic assembly

Qianqing Wang[1], Jingwen Wang[2], Stefana Parascho[2], Katrin Beyer[1]

[1]Earthquake Engineering and Structural Dynamics Laboratory (EESD), EPFL, 1015 Lausanne, Switzerland
[2]Lab for Creative Computation (CRCL), EPFL, 1015 Lausanne, Switzerland

qianqing.wang@epfl.ch, jingwen.wang@epfl.ch, stefana.parascho@epfl.ch, katrin.beyer@epfl.ch

Abstract -

This paper presents a novel method for generating microscale digital twins of rubble stone masonry walls using a robotic construction pipeline. The pipeline integrates stone stock digitalization, stone layout planning, robotic stone assembly, and a dynamic digital twinning process. To validate the approach, a $0.7 \times 0.7 \times 0.4$ m^3 wall was physically constructed in the lab alongside its digital twin, which provided a precise representation of the wall's microstructures. The results show that the proposed pipeline achieved accurate stone assembly with an averaged ±1.2 cm tolerance, producing walls comparable to traditional masonry in terms of compressive strength.

Keywords -

Stone masonry wall; Digital twin; Robotic assembly; Natural stone; Simple compression test

1 Introduction

The construction industry urgently requires innovative solutions to reduce its environmental impact. Utilizing locally available natural materials, such as stone, offers a sustainable path by minimizing energy inputs and carbon emissions associated with material production and transportation [1, 2]. However, traditional stone construction poses challenges for the structural analysis. The irregular shapes of stone units and their placement based on local masonry practices create a highly variable microstructure, which directly impacts the structural behavior of stone masonry building [3]. Current practice for the structural analysis of rubble stone masonry buildings rely on fixed-length classifications based on typologies, which are subjective and introduce ambiguity in the analysis [3].

With advancements in digital technologies, it is now possible to create a digital twin, a 3D model of a physical asset. In masonry structures, digital twinning at the building level—where entire structures are scanned, digitized, and used for numerical analysis—has been explored by several works [4, 5, 6]. However, digital twinning at the stone-level, capturing the as-built microstructure [3], remains rare. This lack of detailed digital twins for stone masonry microstructures poses a significant bottleneck in the development of micro-scale numerical methods for analyzing rubble stone masonry structures [3].

This paper introduces an end-to-end pipeline for generating the microscale digital twin of stone masonry structures, utilizing a novel construction method that eliminates the need for masons. It starts from digitalizing the irregular shape of stones using photogrammetry. The 3D geometry of stones is used by a planning algorithm to determine the positions and orientation of stones. The execution of the plan, i.e., placing stones from the initial position to the final position in the wall, is carried out by a robotic system composed of a vacuum gripper, an on-site scanning system, and a robotic arm. A digital twining pipeline is carried out at the same time as the physical construction, recording the as-built position of stones, which feedbacks the planning algorithm to achieve a more precise and accurate construction.

2 Method

This section presents the main components of the proposed pipeline for the digital construction of mortar-joint stone masonry walls. Section 2.1 presents the methods and results of using photogrammetry to build the digital stone stock. The construction process is explained in Section 2.2, which includes the optimization of the stone layout, the physical construction, and the creation of the digital twin.

2.1 Digitization of initial stone stock

. We scanned 150 stones that were recycled from the demolition waste of stone masonry walls built by masons. The stones were scanned by capturing 100 photographs of one side, then flipping them over and taking an additional 100 photos of the opposite side. The photos were taken with a robotic arm that rotates around the object and takes photos at every 18-degree interval. The 200 photos were used for 3D model reconstruction with the software Agisoft Metashape [7]. A scaled bar is also visible in all photos with specific markers supported by Agisoft Metashape [7]. Figure 1 shows the distribution of stone shapes and stone sizes in the initial stock. Stones that are not placed in the constructed wall are indicated by circles in Figure 1. Stone flatness is measured by the ratio of the shortest

dimension to the intermediate dimension, and elongation is measured by the ratio of the intermediate dimension to the longest dimension, with dimension sizes determined through the principal component analysis [8].

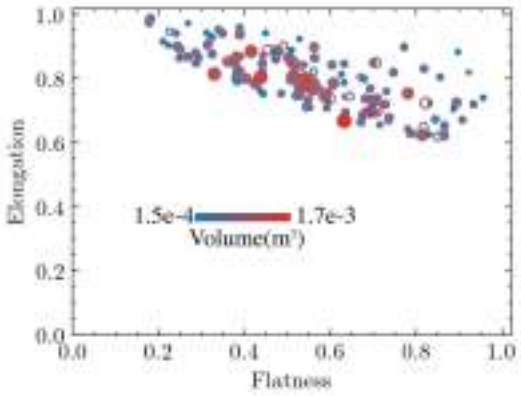


Figure 1. Elongation, flatness, and volume of stones in the initial stock. Stones not used in the final wall are shown by unfilled circles, while others are shown by filled points.

2.2 Microscale digital twinning by segment-wise construction

We propose a segment-wise construction and digital twinning process, where one segment corresponds to elevating the wall with a certain height according to the construction progress during the day. Figure 2 presents the steps for constructing one segment. Before the physical construction of a segment, the layout planning algorithm determines the stone layout for the current segment considering the current landscape, which is composed of the digital twin of previous segments and the boundary condition. The planning of stone layout is followed by the physical construction, where a robotic arm places stones to their planned positions. The robotic assembly of stones is paused every time a new layer of stones was placed on top of an existing one. During the pause, the physical wall is scanned by taking photos around it to capture the as-built positions of stones. We generate the digital twin of the segment based on the scanned layers, which serves as the input for planning and constructing the subsequent segments. Mortar layers are applied manually between layers. We will present in the following sections the details regarding the layout planning, robotic construction, and digital twinning.

2.2.1 Stone layout planning

To generate the stone layout in each segment, the planning algorithm developed in our previous work [9, 10] is used. We adapt the algorithm for segment-wise planning by including two additional functions. One is to apply a mask on top of every segment to avoid stones being placed beyond the segment height. It constrains the algorithm to find feasible positions only below the upper bound. The mask is implemented as assigning the value 1 to voxels above the current segment in the landscape occupancy tensor $L_{\text{occupancy}}$. The other improvement is to use the digital twin of stones from previous segments instead of using their planned positions to account for deviations of stone positions that can arise during the physical construction. As an example, Figure 3 shows the landscape for constructing one segment, composed of five parts: the as-built position of already placed stones from previous segments (in green), the planning positions of stones of the current segment (in blue), the boundaries of the wall (in grey), and the mask for limiting the upper bound (in black).

2.2.2 Robotic stone assembly

We choose a vacuum gripper instead of the conventional two-finger gripper due to its flexibility in handling objects with varying lengths and widths. After analyzing the shape of the stone, considering its curvature, center of gravity, suction area, and avoidance of collisions with stones placed previously, we determined the optimal picking planes for the stones. A team member manually attaches the stone to the gripper at the calculated position, after which we re-scan the stone to precisely locate its position and orientation relative to the robot end-effector. This data enables accurate calculation of the robot's placement trajectory.

2.2.3 Digital twinning

After placing each layer of stones, we scan the half-built wall with photogrammetry and reconstruct the 3D model with Agisoft Metashape [7]. The mesh is then transformed and rescaled according to the positions of markers in the robot frame, measured at the beginning of the construction. We register the stone mesh (a complete stone mesh) to the landscape mesh (where stones can be observed partially) using ICP [11], transforming the stone from the planning position to the as-built position, as shown in Figure 4.

3 Results

We applied the proposed digital construction pipeline to build a stone masonry wall of size $0.7 \times 0.7 \times 0.4$ m^3. The dimensions were chosen to match the size of mason-built walls of a similar typology, as in [3], ensuring a fair comparison of construction results. Figure 5 shows a photo and the 3D surface mesh of the final wall, as well

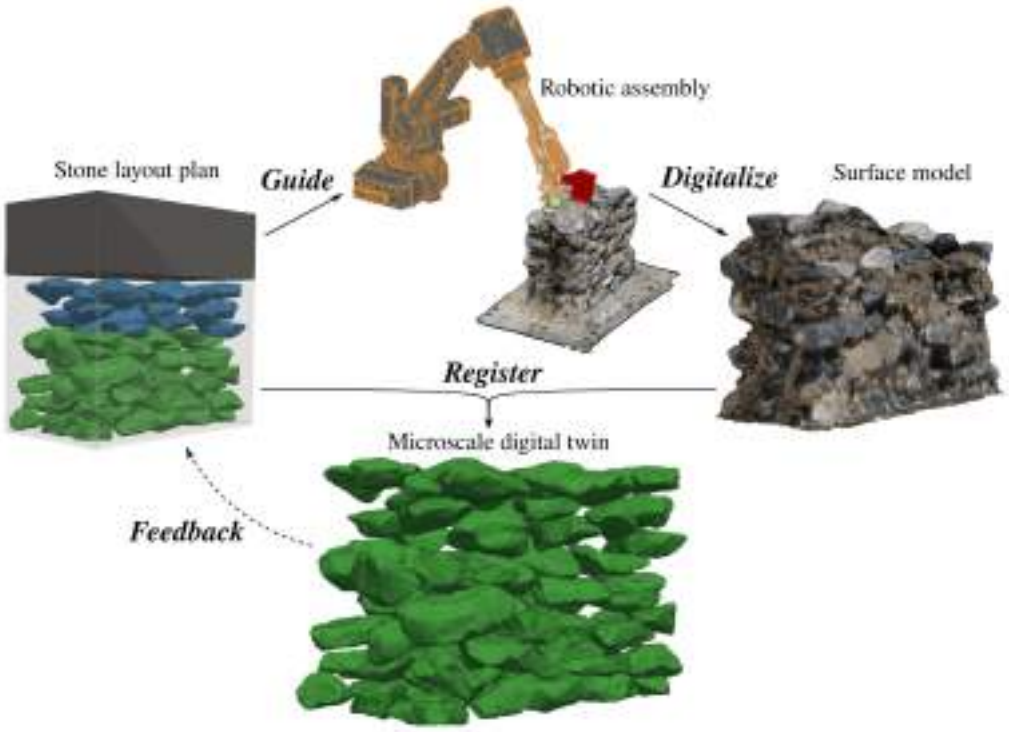


Figure 2. Microscale digital twinning process: starting with the generation of the stone layout plan, followed by robotic assembly guided by the plan, digitalization of the wall surface, and finally, registering the planned stones to the digitized wall to create the microscale digital twin of the as-built stone layout.

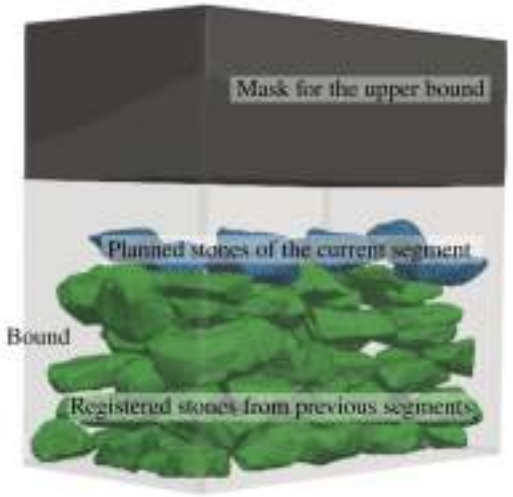


Figure 3. Landscape for planning one segment.

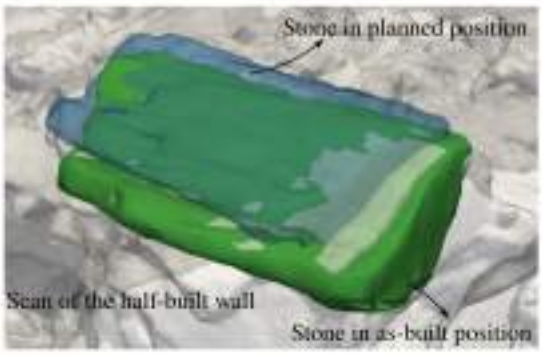


Figure 4. Transforming stones from planned position to as-built position by registering the stone mesh to the landscape mesh.

as its microscale digital twin composed of stone meshes in their as-built positions.

3.1 Simple compression test of the constructed wall

In masonry construction, walls primarily support compressive forces from the weight of the structure above. In this case, compressive strength is a crucial index that directly reflects their load-bearing capacity and resistance to failure under vertical loads. To obtain the compressive strength of the constructed wall, a simple compression test with no confinements at the top and bottom of the specimen was conducted according to the described procedures in EN1015-Part 1 [12]. In the compression tests, three loading/unloading cycles were applied before the peak force. The Young's modulus was derived by fitting a line to each of the three stress–strain cycles and taking the average value of the three slopes as presented in [13]. We obtained a compressive strength of 0.74 MPa, and a Young's modulus of 263 MPa. The strength is within the range reported in [14], between 0.43 MPa and 0.82 MPa.

3.2 Precision in stone assembly

To quantify the precision of the stone assembly process, we made use of the microscale digital twin obtained as the end of the construction, measuring the distances between the as-built positions of stones and their planned positions. We observed an averaged distance between stone centers of 0.012 m and a root mean square error of 0.014 m. The difference between the as-built geometry and planned geometry can be attributed to the following reasons:

1. When the robot reaches the force limit, it will release the stone before it arrives to its planned position.
2. The planned position of the stone is not necessarily a stable position. The stone can move under gravity after robot release it.
3. There are systematic errors in the digital twinning process, including the downsampling of stone meshes for robotic planning and layout planning, the registration of stone mesh to gripping position, the reconstruction of landscape mesh, and the registration of stone mesh to landscape mesh.

Situation 1 consistently happened when placing the first layer of stones. This is because the planning assumes the foundation to be perfectly flat, while the small bumps on the foundation stop the stone placement earlier. The error didn't propagate to the next segment of stones as the planning algorithm considered an updated landscape. For the following layers, stones were mostly placed below their planned position due to the effect of gravity, especially stones on the corners.

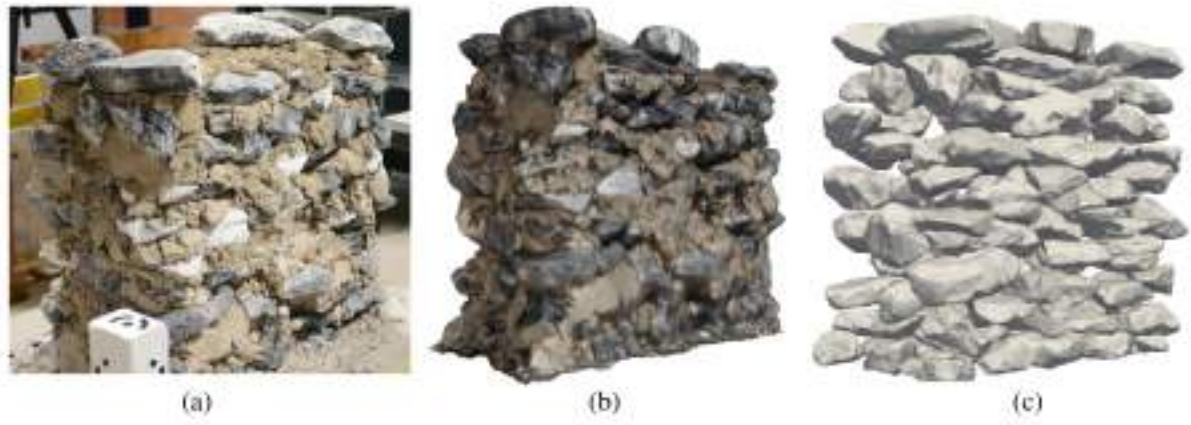


Figure 5. The constructed wall. (a) Photo of the constructed wall. (b) 3D surface model of the final wall. (c) Stones in the final wall on their as-built positions.

4 Conclusions

The traditional construction of masonry structures using natural stones often faces challenges in structural analysis due to the uncertainty and variability of the microstructure. This study introduces a pioneering pipeline for generating the microscale digital twin of stone masonry walls, which replicates the microstructure of physical walls. The pipeline makes use of the recent advance in stone layout optimization [9], to instruct a robotic system for stone assembly. The microscale digital twin is created at the same time of the physical construction, providing dynamic feedback for the construction process.

A three-leaf masonry wall was built as a case study to show the feasibility of the proposed method. Results show that the constructed wall achieved compressive strength of 0.74 MPa, and the digital twinning process had an averaged distance error of 0.012 m. While this work demonstrates the feasibility of digital masonry construction, further advancements in the digitalization of mortar are essential to fully realize the pipeline's potential.

References

[1] J.C Morel, A Mesbah, M Oggero, and P Walker. Building houses with local materials: means to drastically reduce the environmental impact of construction. *Building and Environment*, 36 (10):1119–1126, dec 2001. doi:10.1016/s0360-1323(00)00054-8. URL https://doi.org/10.1016%2Fs0360-1323%2800%2900054-8.

[2] R. Přikryl, Á. Török, M. Theodoridou, M. Gomez-Heras, and K. Miskovsky. Geomaterials in construction and their sustainability: understanding their role in modern society. *Geological Society, London, Special Publications*, 416(1):1–22, January 2016. ISSN 2041-4927. doi:10.1144/sp416.21. URL http://dx.doi.org/10.1144/sp416.21.

[3] Savvas Saloustros, Andrea Settimi, Andrea Cabriada Ascencio, Julien Gamerro, Yves Weinand, and Katrin Beyer. Geometrical digital twins of the as-built microstructure of three-leaf stone masonry walls with laser scanning. *Scientific Data*, 10(1), August 2023. ISSN 2052-4463. doi:10.1038/s41597-023-02417-3. URL http://dx.doi.org/10.1038/s41597-023-02417-3.

[4] Emerson Cuadros-Rojas, Savvas Saloustros, Nicola Tarque, and Luca Pelà. Photogrammetry-aided numerical seismic assessment of historical structures composed of adobe, stone and brick masonry. application to the san juan bautista church built on the inca temple of huaytará, peru. *Engineering Failure Analysis*, 158:107984, April 2024. ISSN 1350-6307. doi:10.1016/j.engfailanal.2024.107984. URL http://dx.doi.org/10.1016/j.engfailanal.2024.107984.

[5] Bryan German Pantoja-Rosero, Radhakrishna Achanta, and Katrin Beyer. Automated image-based generation of finite element models for masonry buildings. *Bulletin of Earthquake Engineering*, 22(7):3441–3469, June 2023. ISSN 1573-1456. doi:10.1007/s10518-023-01726-7. URL http://dx.doi.org/10.1007/s10518-023-01726-7.

[6] A.M. D'Altri, S. de Miranda, G. Castellazzi, and B. Glisic. Numerical modelling-based damage diagnostics in cultural heritage structures. *Jour-*

nal of Cultural Heritage, 61:1–12, May 2023. ISSN 1296-2074. doi:10.1016/j.culher.2023.02.004. URL http://dx.doi.org/10.1016/j.culher.2023.02.004.

[7] AgiSoft. Photoscan professional, 2016. Software.

[8] Deheng Wei, Jianfeng Wang, and Budi Zhao. A simple method for particle shape generation with spherical harmonics. *Powder Technology*, 330:284–291, May 2018. ISSN 0032-5910. doi:10.1016/j.powtec.2018.02.006. URL http://dx.doi.org/10.1016/j.powtec.2018.02.006.

[9] Qianqing Wang, Ketson R.M. dos Santos, and Katrin Beyer. Geometric planning for masonry wall construction with natural stones using discrete convolution. *Engineering Structures*, 329:119695, April 2025. ISSN 0141-0296. doi:10.1016/j.engstruct.2025.119695. URL http://dx.doi.org/10.1016/j.engstruct.2025.119695.

[10] Qianqing Wang, Bryan German Pantoja-Rosero, Ketson R.M. dos Santos, and Katrin Beyer. An image convolution-based method for the irregular stone packing problem in masonry wall construction. *European Journal of Operational Research*, 316(2):733–753, July 2024. ISSN 0377-2217. doi:10.1016/j.ejor.2024.01.037. URL http://dx.doi.org/10.1016/j.ejor.2024.01.037.

[11] Paul J. Besl and Neil D. McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992. doi:10.1109/34.121791.

[12] BS EN. 1052-1: Methods of test for masonry, part 1: Determination of compressive strength. *British Standards Institution, London, UK*, 1999.

[13] Amir Rezaie, Michele Godio, and Katrin Beyer. Experimental investigation of strength, stiffness and drift capacity of rubble stone masonry walls. *Construction and Building Materials*, 251:118972, August 2020. ISSN 0950-0618. doi:10.1016/j.conbuildmat.2020.118972. URL http://dx.doi.org/10.1016/j.conbuildmat.2020.118972.

[14] Savvas Saloustros and Katrin Beyer. *Experimental Investigation on Size-Effect of Rubble Stone Masonry Walls Under In-Plane Horizontal Loading: Overview and Preliminary Results*, page 463–471. Springer Nature Switzerland, September 2023. ISBN 9783031396038. doi:10.1007/978-3-031-39603-8_38. URL http://dx.doi.org/10.1007/978-3-031-39603-8_38.

Integrating BIM and Photogrammetry for Digital Twin Development: A Case Study in Road Infrastructure

Khalil Ahmed Bin Mushtaq[1], Uzair Manzoor[1], Muhammad Usman Hassan[1], Muhammad Fawad[2], Qian Chen[2], and Muhammad Farhan Jahangir[3]

[1] School of Civil and Environmental Engineering, National University of Sciences and Technology, Pakistan.
[2] Faculty of Applied Sciences, University of British Columbia, Okanagan Campus, Canada.
[3] Department of Civil and Architectural Engineering, Aarhus University, Denmark.

Khalil.cemphd21nit@student.nust.edu.pk, umanzoor.ms22nice@student.nust.edu.pk, usman.hassan@nice.nust.edu.pk, muhammad.fawad@ubc.ca, qian.chen@ubc.ca, mfj@cae.au.dk

Abstract –

Building Information Modelling (BIM) has wide application in building construction; however, its adoption in road construction remains limited. Although some studies have investigated the integration of BIM and photogrammetry, this integration remains largely an under-explored and overlooked area in specific environment of road infrastructure construction, with significant gaps in research and application. The novelty of this study lies in bridging this research gap by exploring the practical integration of BIM and photogrammetry, leveraging commercial-grade drone technology to develop a BIM-enabled Digital Twin (DT) tailored to road infrastructure construction. In this study, a 3D BIM model of a 500-meter-long road project was developed using Autodesk InfraWorks 360. The modeling process was followed by the acquisition and processing of 816 high-resolution images, captured over four separate drone flights on different dates, using a DJI Mavic 3 Enterprise commercial drone. The photogrammetric data was processed in Agisoft Metashape, yielding a dense point cloud, a Digital Elevation Model (DEM), and ortho-mosaics. These ortho-mosaics were integrated into BIM model to create a DT. The synchronization of multimodal data demonstrated that a BIM-enabled DT powered by photogrammetry not only enhanced visualization but also proved to be a promising tool for reliable and efficient monitoring of road construction progress.

Keywords –

Digital Twin; BIM; Ortho mosaic; Drone; UAV

1 Introduction

The demand for road infrastructure has grown significantly because these initiatives are vital to a nation's competitiveness, economy, health, education, and quality of life [1]. Road projects frequently face challenges such as schedule delays, cost overruns, and inadequate construction monitoring. These issues compromise the projects' capacity to succeed [2]. Road infrastructure projects consistently perform behind project baseline, resulting in an estimated $82.6 billion in annual cost overruns worldwide [3]. To mitigate these issues, efficient monitoring of construction progress is essential. However, conventional manual tracking methods are often inefficient, time-consuming, and susceptible to errors [4, 5]. These traditional methods generally rely on checklists and daily reports, which are open to subjective interpretation and can fail to provide accurate physical measurements of on-site work, keeping progress monitoring largely a 2D process [6].

While Building Information Modelling (BIM) has been widely adopted in vertical construction, its application in road infrastructure remains limited, leading to missed opportunities for efficiency and cost-effectiveness. Global acceptance of BIM as a management approach and digital construction tool has grown, and it is increasingly becoming an important tool in project management [7]. The foundation of a Digital Twin (DT) lies in its digital representation, with BIM serving as the crucial technology to construct this representation. BIM can be integrated with advanced technologies such as laser scanning, photogrammetry, drones, and sensors to improve the accuracy, efficiency, and real-time updating of site information [8].

The most common data collection technologies include robotic total stations, the Global Navigation Satellite System (GNSS) receivers, Light Detection and Ranging (LiDAR) scanners, and aerial photogrammetry etc. However, the scale of most linear transportation projects requires methods that can efficiently survey large areas with sufficient accuracy and resolution [9]. Photogrammetry has emerged as a cost-effective, rapidly

deployable data collection technology with numerous applications in different fields [9, 10].

Rapid advancements in drone technology have enabled wider and enhanced utilization of photogrammetry in construction projects. Drones are flying robots that fall under the category of Unmanned Aerial vehicles (UAVs). UAVs have become a popular and recognized technology in recent years because of their outstanding ability and flexibility [11]. They can be operated remotely by people or by software using pre-programmed flight plans [12]. UAVs have a range of applications, including aerial photography, and are utilized across many industries [13].

Using UAVs in combination with photogrammetry offers an efficient solution for digital mapping of projects [14]. By taking overlapping, georeferenced vertical and oblique photos, drone-based image acquisition makes large-scale surveys easier and quicker. These photos can then be utilized to produce precise, dense three-dimensional models. This approach is significantly more cost-effective than LiDAR and often achieves higher spatial resolution. Additionally, the portability of the hardware allows access to a greater variety of sites.

2 Related Work

Photogrammetry techniques can be used to generate a 3D model from data collected from drone pictures, while point clouds from 3D scanning of a construction site can also be used to create 3D models [8]. Then, during different phases of construction, this "drone model" can be compared with the Building Information Model to track and evaluate the status of the project [15]. Ortho-mosaic maps and 3D models produced through photogrammetry enhance visualization, providing clear and detailed visual representations of the site that support more informed decision-making. The shared model brings together engineers, contractors, and owners [16].

For effective progress monitoring, it is essential to integrate and visualize both as-planned and as-built information to gain a comprehensive overview of all construction processes. A visualization technique using BIM and Photogrammetry highlights progress deviations by superimposing an as-planned model onto real-time-lapsed photographs, resulting in a single, comprehensive visual representation [17].

While many studies have investigated the integration of BIM and photogrammetry for construction monitoring, they still have limitations and gaps that warrant further research. A review of 134 previous studies by Castañeda, Sánchez [8] aimed at identifying key complementary technologies to BIM in road projects found that photogrammetry was examined in merely eight (8) studies. This finding underscores that integration of photogrammetry with BIM remains a largely overlooked area in the specific context of road infrastructure projects.

Turkan, Bosche [4] developed a system that integrated spatial data collected from 3D laser scanning with schedule data for construction progress tracking; however, it was limited to vertical construction and it neither extended to linear projects like roads nor incorporated photogrammetry. Vick and Brilakis [9] and Lo, Zhang [5] conducted studies on utilization of aerial photogrammetry for road layer construction progress monitoring, but their studies did not integrate the point cloud data into a BIM model to create a DT, and they did not use a commercial-grade UAV. Puri and Turkan [18] proposed a framework using mobile LiDAR and 4D design models, but their study focused primarily on bridge construction only and lacked photogrammetry.

Jacob-Loyola, Muñoz-La Rivera [19] conducted a study on construction progress tracking by periodically recording physical site conditions using UAVs and generating point clouds through photogrammetry. The data was then integrated with 4D BIM for monitoring purposes. However, a key limitation of the study was its primary focus on vertical construction, limiting its applicability to linear infrastructure projects.

Foregoing in view, this study is aimed at addressing the limitations and gaps in previous research by focusing on the integration of BIM and aerial photogrammetry for road infrastructure using a commercial-level drone and the concept of DT—a novel approach that remains overlooked in existing literature.

3 Proposed Framework

This study aims to investigate the integration of BIM with photogrammetry in the specific context of road infrastructure construction using commercial-grade drone. This framework for BIM integration with road infrastructure is shown in Figure 1 and outlined as under:

- Selection of an under-construction road project, collection of relevant drawings, and development of a 3D BIM model using Autodesk InfraWorks 360.
- Acquisition of 816 images of the project through photogrammetry using DJI Mavic 3 Enterprise, spanned over four flights on different dates.
- Processing photogrammetric data using Agisoft Metashape to obtain high-resolution ortho-mosaic.
- Integration of ortho-mosaic with 3D BIM Model in Autodesk InfraWorks 360 to create a DT.
- Synchronization of data between actual project and DT.

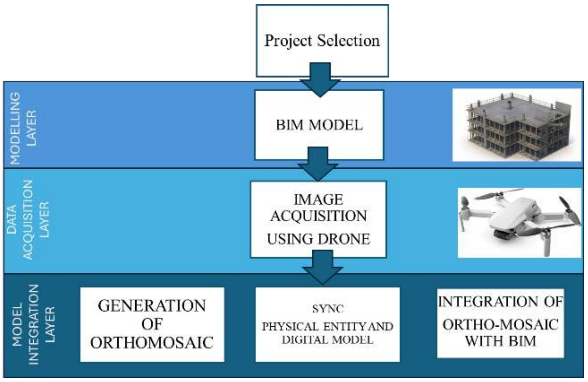


Figure 1. Proposed BIM-Photogrammetry framework for road construction monitoring.

3.1 Data Collection

The DJI Mavic 3 Enterprise was selected for capturing high-resolution images of the construction site due to its affordability, reliability, and capability to cover large areas efficiently. Specifically designed for high-precision surveying and mapping, this drone features a real-time kinematic (RTK) system, which uses satellite navigation to obtain accurate positional data. Other features include: a mechanical shutter, 56× zoom camera, 20MP sensor, and RTK module, all integrated into a compact and mobile platform. Additionally, the camera's ability to capture images at intervals of 0.7 seconds significantly enhances the efficiency of mapping operations, enabling faster data collection [20].

3.1.1 Flight Planning

Flight planning is performed to ensure that manual input is not necessary for acquisition of data, ensure reproducibility of results and limiting the effect of data on the skill of operator. A flight plan outlines the specific route and procedures that the UAV will adhere to during its flight. First, the mission was planned through DJI Remote Control by logging into DJI application. Subsequently, a spacious, obstacle-free launch area was carefully selected. The equipment was meticulously assembled, with all propellers securely fastened. The batteries for both the drone and the remote control were thoroughly inspected to ensure they were fully charged. Once the remote control and the drone were powered on, their connection was verified to confirm proper functionality.

Prior to take-off, the system validated the availability of an adequate number of satellites to ensure a stable and reliable flight. Figure 2 shows a snapshot of flight planning. All necessary permissions required for operation of done were acquired before the conduct of flight and a professional drone pilot was tasked to overview safe operation of the drone.



Figure 2. UAV flight planning for photogrammetry-based monitoring

3.1.2 Image Acquisition

UAV flights were conducted over the study area on four separate dates: 26th August, 6th September, 30th September and 15th October 2024. Each flight was executed with a consistent flying altitude of 60 meters to ensure uniform data collection across all sessions. A total of 816 images (204 images per flight) were captured at a resolution of 1.87cm/pixel, providing high-detail visuals necessary for precise analysis of the project's development over time.

Figure 3 presents a series of high-resolution images captured over the study area from a 90-degree angle. Each image contained detailed geospatial data reflecting the current state of the site, highlighting key features such as access routes, construction modules, and the terrain's topography. Additionally, these images provided essential georeferencing information, facilitating the integration of data for creating a precise point cloud model.



Figure 3. High Resolution images acquired using UAV

3.2 Image Processing

The process of generating ortho-mosaics from the high-resolution drone images was carried out in multiple steps using Agisoft Metashape, which is a robust photogrammetry tool commonly used for 3D reconstruction from 2D images [21].

4 Case Study

The study area selected for this research is a 500-meter linear stretch of the service road located parallel to the Islamabad Expressway in Islamabad, Pakistan. Figure 4 shows the study area. The site is part of an interchange project at Islamabad Highway. The images were taken on site for further processing and integration with BIM for progress monitoring. The coordinates marking the boundaries of the study area are 33°33'19.01"N, 73°10'0.23"E at one end, and 33°33'3.62"N, 73°10'11.52"E at the other end.



Figure 4. Geographic representation of the 500-meter study area along the Islamabad Expressway.

The process of generating ortho-mosaics from the high-resolution drone images was carried out in multiple steps elaborated as follows.

4.1.1 Data Import

The data processing began by importing the captured images into Metashape, ensuring that all high-resolution images, captured at a consistent altitude with a side overlap ratio of 70% and a frontal overlap ratio of 80%, were prepared for further processing.

4.1.2 Image Alignment

The next step involved image alignment, where Structure-from-Motion (SfM) algorithms were utilized to reconstruct the relative positions of the images based on their overlapping features. These algorithms detect matching key points across multiple images, allowing the software to calculate their spatial relationships and reconstruct the 3D geometry of the scene. This process relies on bundle adjustment, an optimization technique that refines the camera positions and orientations to minimize errors and ensure alignment accuracy [22]. Incorporation of RTK positioning data significantly enhanced this process by providing precise geospatial information. The RTK module embedded in the drone ensured centimetre-level accuracy for the positional data of each image, eliminating the need for a large number of Ground Control Points (GCPs).

4.1.3 Point Cloud Data

After image alignment, a dense point cloud was generated for three-dimensional representation of the scanned area. This process first produced a sparse point cloud, containing fewer points spread across the scene, which was used for initial alignment. Once the alignment was complete, the point cloud was refined into a dense point cloud, containing a higher density of points that more accurately reflected the physical environment. These dense point clouds provided finer details of the site, crucial for accurate modeling and progress monitoring.

The data obtained from the point cloud included essential information about the terrain and site structures, offering a foundation for further steps in the workflow. Following the generation of the point cloud, a Digital Elevation Model (DEM) was created to represent the terrain's surface without any objects, such as buildings or vegetation, and was used to understand the topographical aspects of the study area.

4.2 Development of Ortho-Mosaics

Finally, the ortho-mosaics were created from the images. These high-resolution, geometrically corrected images provided a comprehensive view of the entire site, ensuring that every pixel was aligned with its true ground position. Figure 5 shows the DEM and corresponding ortho-mosaics for two out of the four data collection intervals.

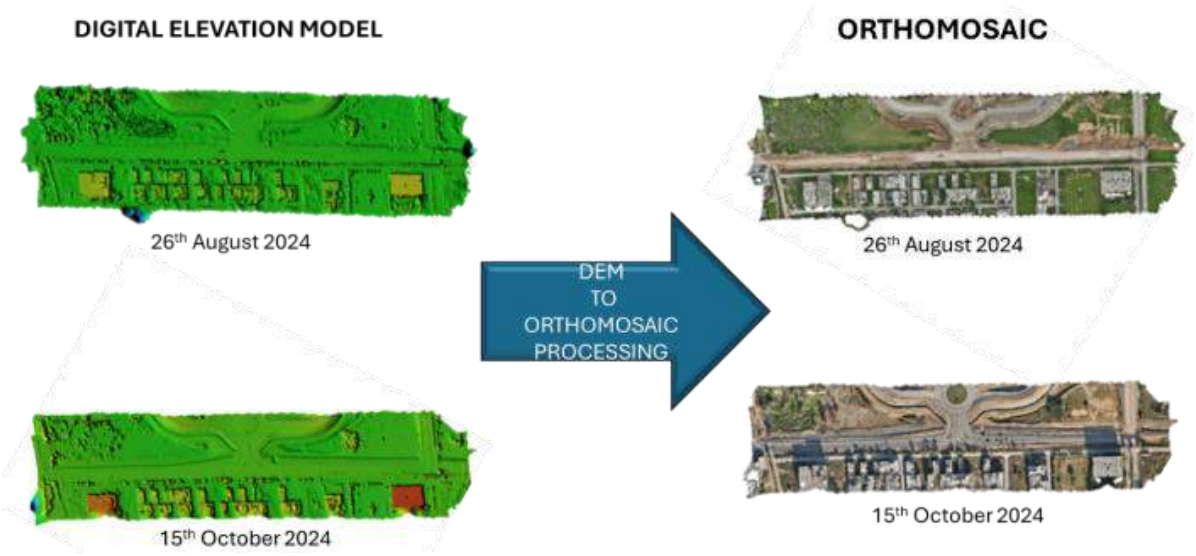


Figure 5. DEM and Ortho mosaic of data acquired on 26th August and 15th October

4.3 Integration with BIM

The processed ortho-mosaic data was exported in formats compatible with various BIM software applications to ensure seamless integration with project workflows. One of the primary formats used for export was TIFF (.tiff), which is widely supported across many Geographic Information System (GIS) and BIM platforms due to its high-resolution capabilities and geospatial referencing attributes. This export process enabled the ortho-mosaic to be incorporated into BIM environments for enhanced visualization, design, and analysis. By exporting the data in such formats, it allowed precise geospatial data into BIM models, improving the accuracy of site planning, progress monitoring, and decision-making throughout the project's lifecycle.

The processed photogrammetry data was integrated into the 3D BIM model using Autodesk InfraWorks 360. This process involved overlaying the high-resolution ortho-mosaic onto the 3D BIM model of the road, enabling precise mapping of the current site conditions. The ortho-mosaic was imported and georeferenced within the InfraWorks environment, creating a comprehensive DTof the construction site.

The creation of layers in the BIM model enabled effective monitoring of the progress of the project. The road layers, such as the base course, sub-base course, and surface course were carefully mapped and tracked to assess their completion status. This integration facilitated a visual comparison of the planned design with the actual progress on-site, enabling accurate and reliable progress monitoring. The combination of photogrammetry data and the BIM model enhanced and ensured the alignment between the planned design and evolving site conditions.

4.3.1 Data Synchronization

Regular updates to the BIM model were conducted by integrating the latest photogrammetry data after each drone flight. This was pivotal in enabling the DT of the project, ensuring that the BIM model accurately reflected the current state of the construction site. By continually synchronizing the model with multimodal real-time data, it provided an up-to-date visualization that enabled effective monitoring of progress and site conditions. Table 1 depicts construction progress on the dates of data acquisition, and Figure 6 shows the digital representation.

Table 1. Construction progress of pavement computed through Digital Twin

Date	Layer	Area (m^2)	Volume (m^3)
26th August 2024	Sub-base	2,461	492
	Base	6,171	1,234
6th September 2024	Base	8,632	1,726
30th September 2024	Base	3,953	791
	Surface	4,726	945
15th October 2024	Surface	8,632	1,726

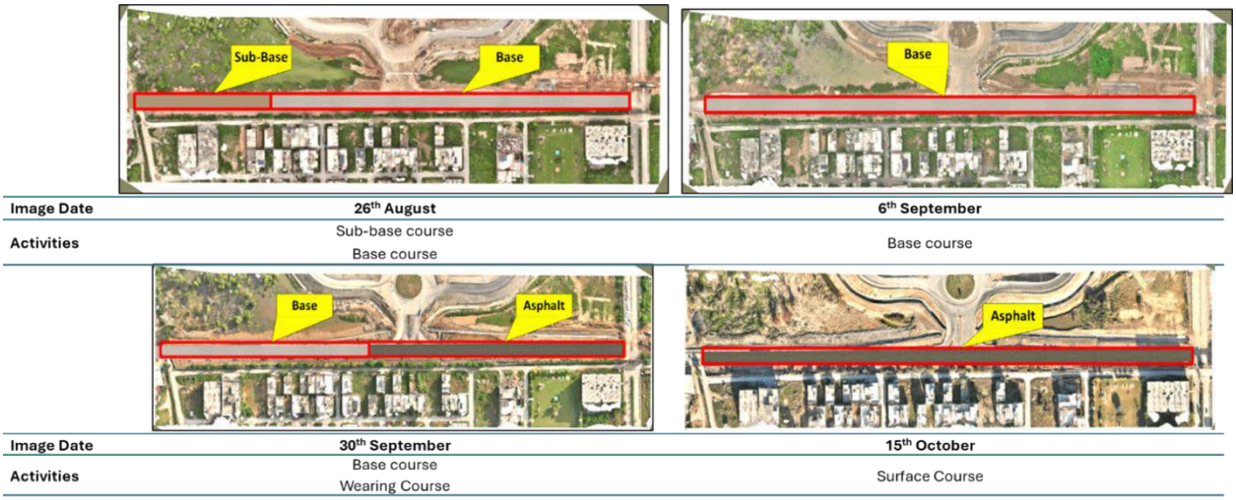


Figure 6. Digital twin of the project site on the mentioned dates along with the status of site.

The end result is represented in Figure 7 where the DT is rendered in 3D along with its environment.



Figure 7. 3D visualization of a preliminary digital twin platform

5 Discussion

Although numerous studies have explored the integration of BIM and photogrammetry, its application in road infrastructure construction remains largely under-explored and overlooked. This study addresses the mentioned research gap by integrating BIM and photogrammetry in the specific environment of road infrastructure and showing practical application in progress monitoring using existing technologies.

The development of 3D BIM model was based on construction drawings of the project. Model development process was followed by the collection of data through aerial photogrammetry technology using a commercial grade drone. Utilization of such RTK-equipped drone ensured precise geospatial information and high resolution of the collected data i.e. 1.87cm/pixel. Moreover, such drones are capable of being used on large-scale projects such as roads which are generally spread over numerous kilometers. The photogrammetric data was processed yielding a dense point cloud and DEM. The data obtained from the point cloud included essential information about project terrain offering a solid foundation for further steps in the workflow.

The DEM was processed to subsequently generate ortho-mosaics. A BIM-enabled Digital Twin (DT) was then developed by integrating high-resolution ortho-mosaic images into Autodesk InfraWorks at regular intervals. The DT was used for successfully computing and monitoring construction progress of road layers.

Previous studies, such as Turkan, Bosche [4] have used photogrammetry and UAV for progress measurement, However, most research has focused on vertical infrastructure, lacking BIM and DT integration. This study leveraged ortho-mosaics to build a reliable and dynamic representation of the project site, adding to its novelty. This study found photogrammetric data collection technology to be cost-effective as well as rapidly deployable. While data interoperability issues were not faced during data collection and processing, such challenges cannot be ruled out if BIM-enabled DT of the road is fed with further data related to scheduling and cost data etc.

6 Conclusion

This study has several notable contributions to the body of knowledge. First, it confirms the extent to which the integration of photogrammetry data and BIM can be successfully implemented in the road construction environment. Additionally, it demonstrates how DTs can be effectively utilized in road construction using existing technological capabilities. Furthermore, it verifies that BIM-enabled DT powered by photogrammetry can not

only facilitate effective road visualization but also help in reliable monitoring and measurement of progress across large-scale construction sites efficiently. Finally, it highlights how commercial-grade drones can be efficiently utilized in road construction for applications that extend far beyond basic videography, offering significant practical benefits.

Future research should explore the integration of Artificial Intelligence (AI) with BIM-Photogrammetry in the specific environment of road construction. Future research can also explore how BIM-Photogrammetry integration can be leveraged for addressing challenges related to construction quality of roads.

References

[1] Herrera, R.F., et al., *Cost overrun causative factors in road infrastructure projects: A frequency and importance analysis*, in *Applied Sciences (Switzerland)*. 2020.

[2] Aziz, R.F. and A.A. Abdel-Hakam, *Exploring delay causes of road construction projects in Egypt*, in *Alexandria Engineering Journal*. 2016, Faculty of Engineering, Alexandria University. p. 1515-1539.

[3] Vick, S.M. and I. Brilakis, *A Review of Linear Transportation Construction Progress Monitoring Techniques*, in *16th International Conference on Computing in Civil and Building Engineering, ICCCBE2016*. 2016. p. 1106-1113.

[4] Turkan, Y., et al., *Automated progress tracking using 4D schedule and 3D sensing technologies.* Automation in Construction, 2012. **22**: p. 414-421.

[5] Lo, Y., et al., *Monitoring Road base course construction progress by photogrammetry-based 3D reconstruction.* International Journal of Construction Management, 2023. **23**(12): p. 2087-2101.

[6] Vick, S. and I. Brilakis, *Asphalt Road Layer Detection for Construction Progress Monitoring*. 2017. p. 271-278.

[7] Wang, T. and H.M. Chen, *Integration of building information modeling and project management in construction project life cycle*, in *Automation in Construction*. 2023, Elsevier B.V. p. 104832.

[8] Castañeda, K., et al. *Building Information Modeling Uses and Complementary Technologies in Road Projects: A Systematic Review*. Buildings, 2024. **14**, DOI: 10.3390/buildings14030563.

[9] Vick, S. and I. Brilakis, *Road Design Layer Detection in Point Cloud Data for Construction Progress Monitoring.* Journal of Computing in Civil Engineering, 2018. **32**(5): p. 04018029.

[10] Roosevelt, C.H., et al., *Excavation is destruction digitization: Advances in archaeological practice*, in *Journal of Field Archaeology*. 2015. p. 325-346.

[11] Li, J., et al., *Non-Terrestrial Networks-Enabled Internet of Things: UAV-Centric Architectures, Applications, and Open Issues*. 2022. p. 1-24.

[12] Cavoukian, A., *Scopus - Document details - Privacy and Drones: Unmanned Aerial Vehicles | Signed in*, in *Information & Privacy, Ontario Canada*. 2012.

[13] Fina, L., et al., *Entropy-Based Distributed Behavior Modeling for Multi-Agent UAVs*, in *Drones*. 2022. p. 1-19.

[14] Zidek, K., et al., *Digital twin of experimental workplace for quality control with cloud platform support*, in *EAI/Springer Innovations in Communication and Computing*. 2020. p. 135-145.

[15] Bhise, G.S., *Construction Monitoring and Reporting using Drones and Unmanned Vehicles (UAV'S)*, in *International Journal for Research in Applied Science and Engineering Technology*. 2024. p. 5872-5884.

[16] Kiriiak, N., *Development and implementation of technical decision for digital support of construction using photogrammetry methods*, in *Nuclear Engineering and Design*. 2021, Elsevier B.V. p. 111366.

[17] Golparvar-Fard, M., et al., *Visualization of Construction Progress Monitoring with 4D Simulation Model Overlaid on Time-Lapsed Photographs*, in *Journal of Computing in Civil Engineering*. 2009.

[18] Puri, N. and Y. Turkan, *Bridge construction progress monitoring using lidar and 4D design models.* Automation in Construction, 2020. **109**: p. 102961.

[19] Jacob-Loyola, N., et al. *Unmanned Aerial Vehicles (UAVs) for Physical Progress Monitoring of Construction*. Sensors, 2021. **21**, DOI: 10.3390/s21124227.

[20] Nwaogu, J.M., et al., *Application of drones in the architecture, engineering, and construction (AEC) industry*, in *Automation in Construction*. 2023, Elsevier B.V. p. 104827.

[21] Bori, M.M. and Z.E. Hussein, *Integration the low cost camera images with the Google earth dataset to create a 3D model*, in *Civil Engineering Journal (Iran)*. 2020. p. 446-458.

[22] Murtiyoso, A., et al., *Open source and independent methods for bundle adjustment assessment in close-range UAV photogrammetry*, in *Drones*. 2018. p. 1-18.

Are Open-Vocabulary Models Ready for Detection of MEP Elements on Construction Sites

Abdalwhab Abdalwhab[1], Ali Imran[1], Sina Heydarian[2], Ivanka Iordanova[2] and David St-Onge[1]

[1]Lab INIT Robots, Department of Mechanical Engineering, ETS Montreal, Canada
[2]GRIDD, Department of Construction Engineering, ETS Montreal, Canada
abdalwhab-bakheet-mohamed.abdalwhab.1@ens.etsmtl.ca

Abstract -

The construction industry has long explored robotics and computer vision, yet their deployment on construction sites remains very limited. These technologies have the potential to revolutionize traditional workflows by enhancing accuracy, efficiency, and safety in construction management. Ground robots equipped with advanced vision systems could automate tasks such as monitoring mechanical, electrical, and plumbing (MEP) systems. The present research evaluates the applicability of open-vocabulary vision-language models compared to fine-tuned, lightweight, closed-set object detectors for detecting MEP components using a mobile ground robotic platform. A dataset collected with cameras mounted on a ground robot was manually annotated and analyzed to compare model performance. The results demonstrate that, despite the versatility of vision-language models, fine-tuned lightweight models still largely outperform them in specialized environments and for domain-specific tasks.

Keywords -

Robotics; MEP Detection; Open-Vocabulary Models

1 Introduction

Recent surveys highlight the growing adoption of mobile robots in the construction industry as a means to address labor shortages and improve safety [1]. Despite these advancements, significant challenges remain for real-world deployment. As noted by [2], collaboration between developers and construction firms is essential to ensure these technologies contribute effectively to productivity and safety. Among the most repetitive and time-consuming tasks in construction workflows, site surveys—critical for monitoring progress and supporting decision-making—are still largely performed manually [3]. Robust detection technologies could significantly enhance the automation of these processes, but detecting Mechanical, Electrical, and Plumbing (MEP) systems remains a complex challenge.

MEP monitoring involves tracking and assessing the installation and functionality of critical systems during construction. This process is essential for identifying potential issues early, preventing costly rework, and avoiding delays [4]. While supervised learning algorithms, particularly convolutional neural networks (CNNs), have significantly advanced object detection in construction, these methods face limitations in dynamic and unstructured environments [5]. Models like YOLO (You Only Look Once) have shown success in controlled conditions but struggle with occlusions, variable lighting, and the introduction of novel objects often encountered on construction sites [6, 7].



Figure 1. Our Journeybot platform for data collection: a Jackal rover base from Clearpath equipped with several cameras and LiDARs (not used in this study).

Open vocabulary object detection presents an alternative approach, offering the ability to generalize across diverse datasets and identify objects not seen during training [8]. These open-vocabulary models, which integrate Large Language Models (LLMs) with computer vision, have demonstrated potential in tasks such as generating detailed inspection reports and monitoring construction sites [9]. Another example, Omni-Scan2BIM [10], employs open-vocabulary models to recognize and segment MEP components for automatic as-built BIM generation. However, their focus on specific classes, such as pipes and ducts, limits their applicability to broader MEP monitoring tasks. Trained on diverse datasets, open-vocabulary models can adapt to new environments and identify emerging components and materials [11], an essential capability for dynamic construction sites where variability is high.

This paper evaluates the potential of open vocabulary

object detection models for MEP monitoring in construction environments. To this end, we constructed a real-world dataset of MEP components using a mobile robot equipped with multiple sensors. The dataset was used to train and evaluate a fine-tuned supervised object detection model, YOLO11 [12], which as a closed-set detector works only for the specific classes that it was trained or fine-tuned on. Then, compare its performance with three open vocabulary object detection models: Grounding-SAM2 [13], Grounding DINO [14], and DETIC [15]. This comparative analysis provides a comprehensive assessment of fine-tuned, lightweight closed-set detectors versus pre-trained open-vocabulary models in real-world MEP monitoring tasks. Our findings offer valuable insights into the effectiveness of deploying open-vocabulary models compared to smaller, task-specific models for dynamic construction environments. As LLMs and open-vocabulary models continue to evolve, this work contributes to understanding their role in advancing construction site automation.

2 Experimental Setup

Robotic Platform: This study utilized the Journeybot robotic platform, a four-wheeled unmanned ground vehicle (UGV) equipped with a hybrid vision and laser sensing system (Figure 1). While capable of semi-autonomous navigation, the robot was manually operated during data collection for safety reasons.

The sensing system was tailored for comprehensive environmental monitoring and digital twin data collection. It included four Intel Realsense D435i stereo cameras covering all sides of the robot, with one facing upward, and a front-facing Intel Realsense T265 tracking camera for enhanced localization. Depth images from the D435i supported collision avoidance, together with a horizontally mounted Velodyne Puck 32MR LiDAR. An Ouster LiDAR augmented the data collection system, but both LiDARs are not covered by this work. Wheel encoder data and IMU readings were synchronized with detection sensors by an NVidia Jetson computer.

Dataset: The dataset was collected at an active construction site provided by an industrial partner, with a primary focus on MEP assets. Using the Robot Operating System (ROS), our robotic platform was manually teleoperated through the site, capturing data stored as *rosbags*. Images were extracted at a rate of 10 Hz from the rosbags, followed by a filtering process to remove redundant and blurred frames. The remaining images were annotated for the 10 specific MEP components listed in Table 1 using an online annotation tool. The dataset is shared online.

The final annotated dataset was divided into 70% training, 20% validation, and 10% testing splits. Table 1 also presents the number of images and instances per class. Example images are shown in the results section (Figure 3), illustrating the challenges posed by real-world data collection on a mobile platform. These include objects appearing at varying distances, being partially occluded, or blurred due to motion, reflecting the dynamic and unpredictable nature of construction site environments.

Table 1. Dataset details with number of images (and instances) per each class.

No.	Class	Images(instances)
	All	8885(14064)
1	Boiler	768(769)
2	Cable Tray fitting	252(317)
3	Electrical Panel	658(1022)
4	Fire Alarm Detector	872(877)
5	Pipe Fitting	3625(5014)
6	Valve	610(906)
7	Electrical Outlet	786(1023)
8	Generator	124(124)
9	Light	1992(2168)
10	Pump	1536(1844)

Evaluation Procedure: For quantitative evaluation, we assessed the performance of a lightweight object detector (YOLO11 Nano pre-trained on COCO [16]) fine-tuned (200 epochs) on our dataset, comparing its best performing weights (on the validation split) against three pre-trained large open-vocabulary models without fine-tuning: Grounding DINO, Grounded SAM2, and DETIC. The evaluation metrics included precision, recall, and F1 score (All measured on the testing split at 0.5 intersection over union).

3 Results and Discussion

Table 2 depicts the results of F1 score for YOLO11 Nano compared to the open-vocabulary models GSAM2, GDINO, and DETIC, evaluated on the test dataset. YOLO11 achieved a total precision of 0.87, outperforming GSAM2, GDINO, and DETIC, which scored 0.018, 0.02, and 0.015, respectively. Similarly, YOLO11 exhibited superior recall at 0.901, compared to 0.042, 0.084, and 0.018 for the open-vocabulary models.

Table 2. F1 score total and per class for each model evaluated in the testing dataset split

Class	YOLO11	GSAM2	GDINO	DETIC
all	0.89	0.018	0.032	0.014
1	0.94	0.01	0	0.005
2	0.86	0	0	0
3	0.9	0.045	0.059	0.043
4	0.94	0.036	0.026	0.063
5	0.86	0.01	0.013	0.013
6	0.87	0.005	0.007	0
7	0.82	0.005	0.003	0
8	0.92	0	0.006	0
9	0.81	0.038	0.027	0
10	0.93	0.014	0.123	0.013

The results underscore the superior performance of YOLO11 Nano, with improvements exceeding 85% in

precision, 82% in recall, and 86% in F1 score compared to the best-performing open-vocabulary model (GDINO). YOLO11 also achieved consistent detection across all classes, including class 2 (Cable Tray Fitting), which none of the open-vocabulary models detected. The failure of these models in detecting such specialized classes likely stems from their reliance on large, generic datasets like COCO [16] and LVIS [17], which lack adequate representation of MEP-specific components. DETIC exhibited the poorest performance, failing in five out of ten classes, while GSAM2 and GDINO each failed in two classes.

Figure 2 presents the normalized confusion matrix for YOLO11 Nano's predictions, demonstrating robust performance even for less-represented classes, such as generators, with a detection accuracy of 91%. This emphasizes YOLO11's capability to handle imbalanced datasets, a common challenge in domain-specific applications.

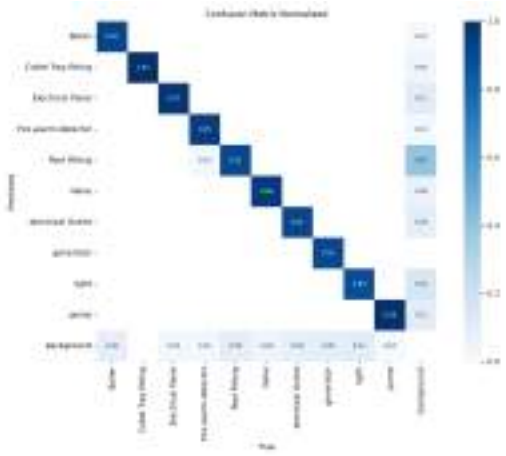


Figure 2. YOLO11 normalized confusion matrix for its predictions on the testing dataset split.

Figure 3 provides qualitative examples of YOLO11 Nano's predictions, displaying predicted classes, bounding boxes, and confidence scores. In some cases, the model outperformed human annotations, identifying overlooked objects like valves in the bottom-right corner of an image. Conversely, in instances of blurred images, the model avoided false positives, aligning with human annotators' limitations. However, false detections occasionally occurred, particularly for reflective surfaces, which resembled target classes like light sources.

These findings highlight the limitations of general-purpose open-vocabulary models in specialized environments like construction sites, where dynamic conditions, specialized components, and unique visual contexts demand fine-tuned, domain-specific approaches.

Real-world Applicability: Assessing the applicability of open-vocabulary models for real-world robotic systems compared to lightweight object detectors like YOLO11



Figure 3. Qualitative examples of YOLO11 Nano's performance, displaying predicted classes, bounding boxes, and confidence scores.

Nano requires careful consideration of model size and memory requirements. Model size, expressed in terms of parameters, is a critical factor influencing memory usage and computational efficiency, especially on resource-constrained platforms. Among the evaluated models, GSAM2 exhibited the largest memory footprint with 910 million parameters—350 times larger than YOLO11 Nano, which has just 2.6 million. GDINO followed, with a parameter count 67 times larger than YOLO11 Nano, while DETIC, the smallest open-vocabulary model, required 164 million parameters, making it 63 times larger than YOLO11 Nano.

Computational efficiency is equally crucial for deployment on real-time robotic systems. YOLO11 Nano was specifically evaluated on an NVIDIA Jetson Orin Nano, processing image samples from the testing dataset at a speed of 23.36 frames per second. This performance meets the requirements for real-time applications, demonstrating that YOLO11 Nano is highly suited for resource-constrained environments, unlike larger open-vocabulary models whose significant computational demands make them less practical for such applications.

4 Conclusion

This study compared the performance of pre-trained open-vocabulary detection models and fine-tuned lightweight detectors for robotic MEP detection in construction sites. Using a dataset collected with a tele-operated robotic platform, we demonstrated that fine-tuned models like YOLO11 Nano significantly outperformed open-vocabulary models in detection accuracy, computational efficiency, and real-time applicability. While open-vocabulary models offer promising generalization capabilities, their performance on domain-specific tasks remains

limited.

With the huge advancements in vision-language models, we hope this work encourages research on narrowing this gap, possibly, by fine-tuning open-vocabulary models on specialized datasets, potentially combining their adaptability with the precision of supervised models. Another direction could be to design construction-related prompts to adapt the open-vocabulary detectors.

References

[1] Guglielmo Carra, Alfredo Argiolas, Alessandro Bellissima, Marta Niccolini, and Matteo Ragaglia. Robotics in the construction industry: State of the art and future opportunities. In *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, volume 35, pages 1–8. IAARC Publications, 2018.

[2] Rafael Gomes Braga, Muhammad Owais Tahir, Ivanka Iordanova, and David St-Onge. Robotic deployment on construction sites: considerations for safety and productivity impact. *arXiv preprint arXiv:2404.13143*, 2024.

[3] Chun Wang, Peng Ouyang, Xiaodong Liu, Zhihua Zou, Yuanping Wang, and Ruiling Wang. Future prospects of the assembly model for mep systems in chinese buildings: A whole life cycle perspective. *Applied Sciences*, 14(15):6818, 2024.

[4] Seungjun Roh, Zeeshan Aziz, and Feniosky Pena-Mora. An object-based 3d walk-through model for interior construction progress monitoring. *Automation in Construction*, 20(1):66–75, 2011.

[5] Minsoo Park, Jinyeong Bak, Seunghee Park, et al. Small and overlapping worker detection at construction sites. *Automation in Construction*, 151:104856, 2023.

[6] Fei Hao, Tianqi Zhang, Guanghong He, Ruofei Dou, and Chao Meng. Casnli-yolo: construction site multi-target detection method based on improved yolov5s. *Measurement Science and Technology*, 35 (8):085202, 2024.

[7] Tausif Diwan, G Anirudh, and Jitendra V Tembhurne. Object detection using yolo: Challenges, architectural successors, datasets and applications. *multimedia Tools and Applications*, 82(6):9243–9275, 2023.

[8] Tianheng Cheng, Lin Song, Yixiao Ge, Wenyu Liu, Xinggang Wang, and Ying Shan. Yolo-world: Real-time open-vocabulary object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16901–16911, 2024.

[9] Hongxu Pu, Xincong Yang, Jing Li, and Runhao Guo. Autorepo: A general framework for multimodal llm-based automated construction reporting. *Expert Systems with Applications*, 255:124601, 2024.

[10] Boyu Wang, Zhengyi Chen, Mingkai Li, Qian Wang, Chao Yin, and Jack CP Cheng. Omni-scan2bim: A ready-to-use scan2bim approach based on vision foundation models for mep scenes. *Automation in Construction*, 162:105384, 2024.

[11] Maghsood Salimi, Mohammad Loni, Sara Afshar, Antonio Cicchetti, and Marjan Sirjani. Constscene: Dataset and model for advancing robust semantic segmentation in construction environments. *arXiv preprint arXiv:2312.16516*, 2023.

[12] R Khanam and M Hussain. Yolov11: An overview of the key architectural enhancements. arxiv 2024. *arXiv preprint arXiv:2410.17725*.

[13] Tianhe Ren, Shilong Liu, Ailing Zeng, Jing Lin, Kunchang Li, He Cao, Jiayu Chen, Xinyu Huang, Yukang Chen, Feng Yan, et al. Grounded sam: Assembling open-world models for diverse visual tasks. *arXiv preprint arXiv:2401.14159*, 2024.

[14] Shilong Liu, Zhaoyang Zeng, Tianhe Ren, Feng Li, Hao Zhang, Jie Yang, Qing Jiang, Chunyuan Li, Jianwei Yang, Hang Su, et al. Grounding dino: Marrying dino with grounded pre-training for open-set object detection. In *European Conference on Computer Vision*, pages 38–55. Springer, 2025.

[15] Xingyi Zhou, Rohit Girdhar, Armand Joulin, Philipp Krähenbühl, and Ishan Misra. Detecting twenty-thousand classes using image-level supervision. In *European Conference on Computer Vision*, pages 350–368. Springer, 2022.

[16] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pages 740–755. Springer, 2014.

[17] Agrim Gupta, Piotr Dollar, and Ross Girshick. LVIS: A dataset for large vocabulary instance segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019.

Deep Point Cloud Building Envelope Segmentation (DeeP-CuBES) using Deep Learning

Balaji Selvakumar, Yifang Liu, Nolan W. Hayes, Diana Hun, Bryan P. Maldonado

Buildings and Transportation Science Division, Oak Ridge National Laboratory*, United States of America

selvakumarb@ornl.gov, liuy5@ornl.gov, hayesnw@ornl.gov, hunde@ornl.gov, maldonadopbp@ornl.gov

Abstract -

Building Information Modeling (BIM) plays an important role in building design and construction, particularly for achieving energy-efficient retrofits. Building envelope retrofits using panelized prefabricated system, such as those popularized by the Energiesprong program, need accurate as-built dimensions of facade features (windows, doors, etc.) to achieve the desired thermal and air tightness. Traditionally, building surveying is done manually, resulting in a time-consuming and labor-intensive process. Recently, 3D point clouds from terrestrial LiDAR have been used to automate the generation of as-built dimensions of existing buildings. However, automated BIM using LiDAR relies on solving the point cloud semantic segmentation (PCSS) problem. In this work, we propose a robust pipeline for solving the PCSS problem using deep neural networks, focusing on overcoming challenges posed by imbalanced datasets and complex architectural features. We introduce the first high-density, labeled, and validated building envelope point cloud dataset derived from multiple building scans, specifically curated to tackle challenges in facade-level segmentation. Results from the trained neural networks show that advanced attention-based architectures and incorporating radiometry (light intensity and RGB) features significantly boost segmentation accuracy for windows and doors.

Keywords -

point cloud, semantic segmentation, supervised learning

1 Introduction

More than 40% of the buildings in the United States were constructed prior to 1980, when energy regulations were in place [1]. Building envelope retrofits using prefabricated overclad panels offer a promising solution to increase the energy efficiency of these older buildings. Moreover, this process minimizes occupant disruption and has a high potential for automation, shortening assembly time at the job site and increasing overall construction quality. However, this type of retrofit requires precise as-built dimensions of the building to effectively design and manufacture the overclad panels, highlighting the need for Building Information Modeling (BIM) to create a digital twin. Creating an as-built retrofit BIM involves measuring the geometry of building facade features, such as windows and doors, to produce a dimensional digital twin that facilitates the design and installation of prefabricated overclad panels. However, this process is often time-consuming, labor intensive, and susceptible to errors. Therefore, automating this process has substantial benefits when trying to achieve building envelope retrofits at a large scale [2].

Recent advancements in 3D LiDAR scanning technologies have enabled the acquisition of high-density point cloud data, providing detailed geometric representations of scanned structures. However, point clouds are inherently unstructured, making the extraction of semantic information—such as identifying facades, windows, and doors—a complex but essential task for creating BIM-ready datasets [3]. This process can be automated by solving the Point Cloud Semantic Segmentation (PCSS) problem, revolutionizing building surveying by eliminating the need for manual annotation and generated dimensional twins in real time. Despite its promise, several challenges remain, particularly for datasets with imbalanced class distributions and complex architectural features [4]. This study focuses on improving facade segmentation for retrofitting applications by introducing a labeled and validated high-density facade dataset, curated from multiple building scans. This dataset bridges a critical gap in resources required for evaluating advanced neural network for solving the PCSS problem.

Two deep learning models were evaluated: PointNet++, based on a multi-layer perceptron (MLP) architecture [5], and the Superpoint Transformer (SPT), a transformer-inspired model leveraging hierarchical feature learning and attention mechanisms [6]. These models were fine-tuned by training them on additional radiometry features, enabling richer feature representations and achieving superior segmentation accuracy. Furthermore, to address the

*Notice: This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (https://www.energy.gov/doe-public-access-plan).

inherent challenges of imbalanced datasets, where walls dominate and classes such as windows and doors are underrepresented, we employ a combination of sampling techniques and weighted loss functions.

The structure of the paper is as follows. Section 2 reviews related work on PCSS in the construction industry, highlighting the need for labeled data. Section 3 details the dataset collection, labeling validation, post-processing procedures, and the class distribution within the dataset. Section 4 discusses the deep neural network architectures used and their evaluation metrics. Finally, Section 5 presents the results, while Section 6 presents the conclusions and future work.

2 Related Works

2.1 PCSS applications in construction

Recent breakthroughs in deep neural networks have led to the development of powerful deep learning models which can capture and learn texture, shape, and contextual information from images, accelerating the use of computer vision in various applications. Semantic segmentation is a widely used technique in computer vision, which is the process of assigning a label to every pixel in an image, indicating which category the corresponding pixel belongs to. This helps in better scene understanding of 2D images. Among various scene understanding problems, 3D semantic segmentation is helpful in determining accurate object boundaries with similar patterns along with their labels [7]. Unlike 2D images, 3D data generated by laser scanners are unordered and invariant to permutations, making it difficult for the deep neural networks to learn the desired features [3]. Due to its inherent nature, point clouds are subject to outliers and noise. In addition, non-uniform point densities in various regions affect point cloud feature learning. Deep learning (DL)-based segmentation for point clouds has been used in various applications such as robotics, industrial automation, autonomous driving [4] and 3D scene understanding, enabling automated decision making in three-dimensional space[8] . Datasets like KITTI[9] and nuScenes [10] are widely used in benchmarking point cloud segmentation algorithms and their performance in autonomous driving in real world scenarios.

Given that, in general, old buildings lack an as-built, or even an as-designed BIM, point cloud data has become popular within the construction sector for 3D BIM generation. Recently, innovative methods have been developed to efficiently create as-built BIMs for various construction projects, including envelope retrofits, re-cladding, and window replacements. [11]. PCSS can be used to generate accurate as-built BIMs from 3D laser scans of the building, where the accuracy depends mainly on the sensor and range. However, the efficiency of supervised machine learning used in PCSS is heavily dependant on the available labeled training data. Currently, the amount and variety of open source point cloud datasets to train, validate, and compare new neural network architectures are mainly limited to the autonomous driving industry. Beyond autonomous driving, numerous datasets like S3DIS featuring residential and commercial buildings focus on indoor room and office scenes captured using RGB-D devices [12]. Building exterior envelopes, however, usually correspond to sparse urban datasets like Semantic3D[13] and Toronto3D[14] collected using mobile/aerial laser scanning technology and those featuring [15] and [16] facade-level classes are of sparse density. Therefore, it is essential to address the scarcity of dense, accurately labeled point cloud data of exterior building envelopes, specifically designed for semantic segmentation tasks aimed at enhancing productivity and reducing costs in retrofit and new construction projects.

2.2 Class imbalance in real-world datasets

The performance of DL classification models is heavily impacted by how well the data is distributed across the different classes.[17] To achieve optimal performance, and avoid overfitting on individual classes, the model has to be trained on datasets that are as balanced as possible [7]. In a real-world point cloud scan of a building exterior envelope, as much as 80% of elements can be classified as walls points, while the rest correspond to features such as windows and doors.

Standard learning algorithms assume balanced class distributions and severe imbalances like our application, can lead to biased decision boundaries on under-represented minority classes. This study[18] systematically analyzed how imbalanced data skews model predictions towards majority classes, leading to biased decision boundaries. Due to the large presence of the wall points in real world facade dataset like ours, the model can be heavily skewed and can learn to over classify the majority group. Furthermore, class imbalance can distort the performance evaluation when using metrics such as accuracy, as correctly classifying the major classes may yield high scores that incorrectly indicate good performance [19].

3 Dataset Collection and Labeling

For data collection, we used the Leica MultiStation MS60 to scan buildings across the ORNL campus. The MS60 is a terrestrial laser scanner (TLS) that allows users to customize scan density, enabling the collection of dense scans with high accuracy. The lidar scanner returns the Cartesian coordinates x, y, z and the light intensity I of a point on a surface measured by the laser. In addition, the TLS is equipped with a 5 MPx RGB camera fea-



Figure 1. Sample facades used for DeeP-CuBES

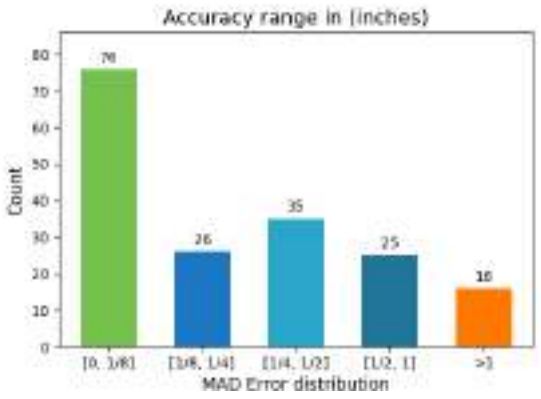


Figure 2. Error distribution for window dimensions

turing a CMOS sensor which takes captures images of the scanned area to later interpolate the points with the pixel RGB values. Therefore, the final raw point cloud belongs to a seven-dimensional space with coordinates $\begin{bmatrix} x & y & z & I & r & g & b \end{bmatrix}$. A total of 25 building facades across the ORNL campus were scanned and the point clouds were initially cleaned to remove outliers such as trees and other occlusions using CloudCompare [20]. A subset of the facades collected is shown in Figure 1.

With the cleaned, pre-processed data of different individual facades, we used the ORNL-developed Automatic point Cloud Building Envelope Segmentation (Auto-CuBES) algorithm to accelerate data labeling [21]. Auto-CuBES uses a sequence of unsupervised machine learning methods, such as k-means and DBSCAN, to segment the building envelope into wall points, windows points, door points, and other categories (building extrusions) with an accuracy of 3mm. Although Auto-CuBES is significantly faster than manual segmentation, it requires several segmentation steps which involve calibrating various parameter values to segment window and door openings from the wall points. The segmentation results from all building scans were grouped into three classes: windows+doors, walls, and others. The results from Auto-CuBES were validated by comparing them to manual measurements of the window dimensions (length and height) using the TLS laser. Figure 2 shows the distribution of Mean Absolute Deviation (MAD) observed in our dataset. Note that most of the error distribution is heavily skewed toward lower values (ranging between 0 to 1/8 in), as indicated by the high count in the first bin. Also note that the manual measurements are not the actual ground truth, as it may be subject to human inaccuracies. This result suggests that the segmentation is accurate for most cases, with less than 10% of the dataset having a MAD error greater than 1 inch. After labeling and validation, we created the Oak Ridge National Laboratory Building Envelope Library (ORNOBEL) for exterior building envelopes.

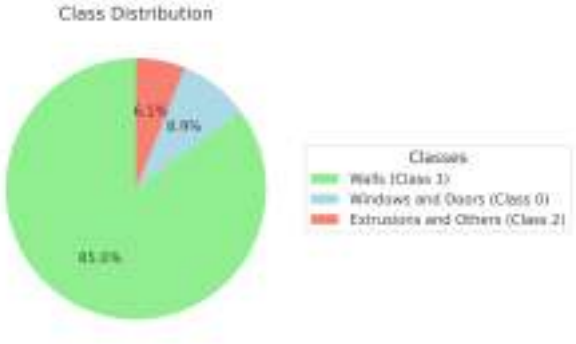


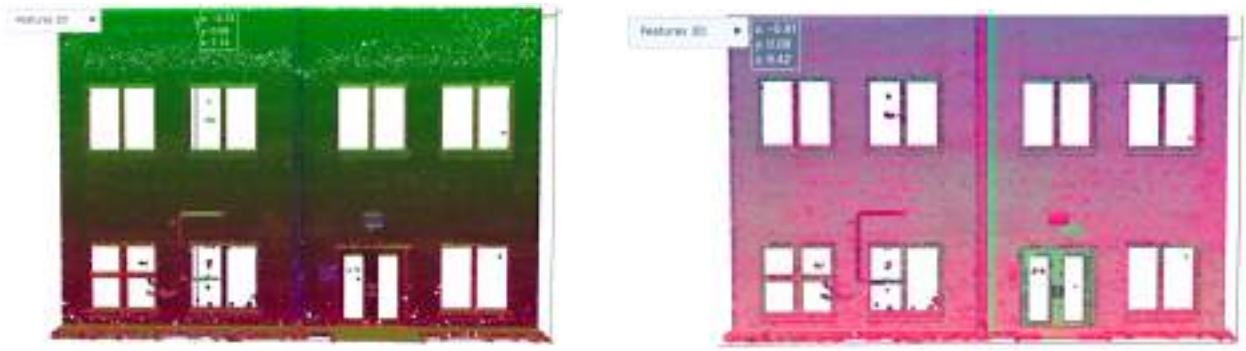
Figure 3. Distribution of 223 million points within the Oak Ridge National Laboratory Building Envelope Library (ORNOBEL)

The ORNOBEL dataset consists of dense facade scans of 25 different multistory buildings with 3mm accuracy and more than 220 million points of labeled data. Figure 3 shows the distribution of the three classes used, namely windows+doors, walls, and others. These categories can be further subdivided in the future. Note that 85% of the data consists of wall points, creating a significant imbalance that can affect the performance of any trained model.

3.1 Post processing and preparing the data

To ensure the point cloud data is standardized, aligned, and ready for neural network model training, we implemented a series of pre-processing steps, as outlined below:

- **Spatial Centering & Intensity Normalization:** The origin of each facade point cloud was moved to the geometric center of its minimum bounding box. This ensures consistency of the spatial positioning of all walls and openings (windows and doors) facilitating easier model interpretation. Light intensity was normalized to avoid large values observed in the raw data.



(a) Default feature map without radiometry features

(b) Fine tuned feature map with added radiometry features

Figure 4. Comparison of fine tuned feature partition map for SPT.

To handle outliers in intensity values, we employed the *Robust Z-Score* normalization method:

$$I_{\text{robust}} = 0.6745 \cdot \frac{I - \text{Median}(I)}{\text{MAD}} \tag{1}$$

where I is the raw intensity value, Median(I) is the median intensity, and MAD is the Median Absolute Deviation. This ensures that the intensity values are scaled robustly, particularly across facade surfaces and doesn't skew the model training.

- **PCA-Based Alignment:** For facade-specific data preparation, Principal Component Analysis (PCA) was applied to the wall points using scikit-learn to establish a consistent reference axis [22]. Subsequently, point clouds representing windows and doors were aligned to this PCA-derived axis to standardize their orientations. This alignment step is crucial for facade datasets collected without a common TLS orientation, as it ensures uniformity, allowing models to concentrate on geometric features such as coplanarity without being influenced by variations in orientation.

- **Output Format:** RGB color values were left unchanged, as they can serve as additional radiometric features for identifying facade components. The preprocessed point clouds, with spatial centering, normalized intensity, PCA alignment, and original RGB values, were exported in standard h5 formats. This approach ensures that the data is directly usable for neural network training.

4 Deep Neural Networks applied to PCSS

PointNet++ was a groundbreaking model introduced in 2017 to solve the Point Cloud Semantic Segmentation(PCSS) problem [5]. The algorithm pioneered the hierarchical learning of both local and global features directly from unordered point clouds. Given its historical significance, we used PointNet++ as a baseline to evaluate the PCSS accuracy on our dataset. PointNet++ uses a Multi Layer Perceptron (MLP) architecture and processes point clouds hierarchically, using set abstraction implemented by multiple MLP layers to capture local and global features. However, it only uses the xyz coordinates for feature learning, ignoring other informative attributes like light intensity and RGB. When trained and tested on the ORNOBEL dataset, the model performed poorly in capturing spatial relationships for minority classes such as windows and doors, highlighting the difficulty of dealing in imbalanced classes.

We have implemented Superpoint Transformer (SPT) neural network architecture [6], an advanced model which uses additional feature sets and super point partitions to improvise feature learning. This model draws inspiration from the U-Net convolutional architecture [23], which has been groundbreaking in 2D image segmentation. By integrating transformer-based self-attention mechanisms, which has seen breakthroughs in language understanding, SPT effectively captures the spatial dependencies between unordered points, while its U-Net-like encoder-decoder structure allows for multi-scale feature learning. Additionally, its compact size, with fewer parameters, ensures computational efficiency, making it an ideal choice for dense point cloud segmentation of building envelopes.

4.1 Superpoints features

Superpoints are groups of points clustered together based on their local geometric properties. The Superpoint Transformer outperforms PointNet++ by leveraging a more sophisticated architecture inspired by U-Net and superpoint graphs. Unlike PointNet++, SPT uses a combination of xyz coordinates, light intensity, and RGB values to construct a rich, hierarchical feature map. These features are used to construct a meaningful hierarchical partition with adjacency relationship with neighboring points. Subsequently, SPT organizes the point clouds with feature

maps into superpoints (i.e., clusters of points with adaptive and shared local features) enabling efficient and structured feature aggregation.

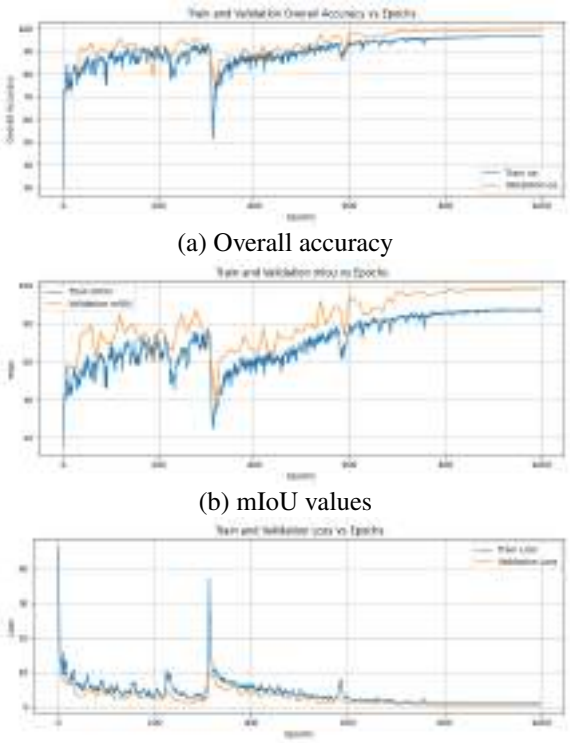
The transformer encoder-decoder architecture with self-attention allows SPT to learn spatial dependencies between points effectively, which is critical for capturing complex facade openings and features. SPT integrates MLP layers to extract point-wise features and project them onto higher dimensions before applying attention. This combines the strengths of both MLPs and transformers. Our fine-tuning process involved customizing SPT to optimize its ability to learn facade-specific features as shown in Figure 4. With the appropriate calibrated parameters in SPT, classification accuracy across all metrics was improved compared with the PointNet++ baseline results, particularly for correctly classifying minority classes.

4.2 Deep neural networks training

To establish a point of comparison for improvements, we generated a baseline models using PointNet++. The original PointNet++ model was trained without implementing any sampling methods or dataset balancing, relying solely on xyz coordinates for feature learning (no radiometry features like intensity or RGB were used). The model performed poorly across all classes due to the dataset's inherent imbalance. To address this, we used a combination of sampling methods: weighted cross-entropy loss as a passive sampling method and downsampling majority class as an active approach. While these techniques significantly improved the model's performance, the generalization to non-downsampled wall points remained poor. This highlighted the limitations of addressing class imbalance solely by downsampling a single class, which is impractical for real-world applications.

The Superpoint Transformer model was initially trained on the ORNOBEL dataset without fine-tuning or incorporating intensity and RGB features. Moreover, the SPT was trained directly on the class-imbalanced dataset, providing a more robust and effective solution to the challenges posed by the dataset's structure. The model performed extremely well, yielding a significant boost over the Pointnet++ model's results. To effectively train our neural networks, we employed cross-entropy loss functions. Specifically, for the PointNet++ model, we applied a weighted cross-entropy loss to explicitly handle class imbalance within the dataset. For the SuperPoint Transformer model, the standard categorical cross-entropy loss was sufficient enough.

Overall Accuracy: measures the ratio of correctly classified points across the entire dataset. While accuracy was recorded, it is misleading for imbalanced datasets as it disproportionately favors the majority class (wall points in our case)



(a) Overall accuracy

(b) mIoU values

(c) Model Loss

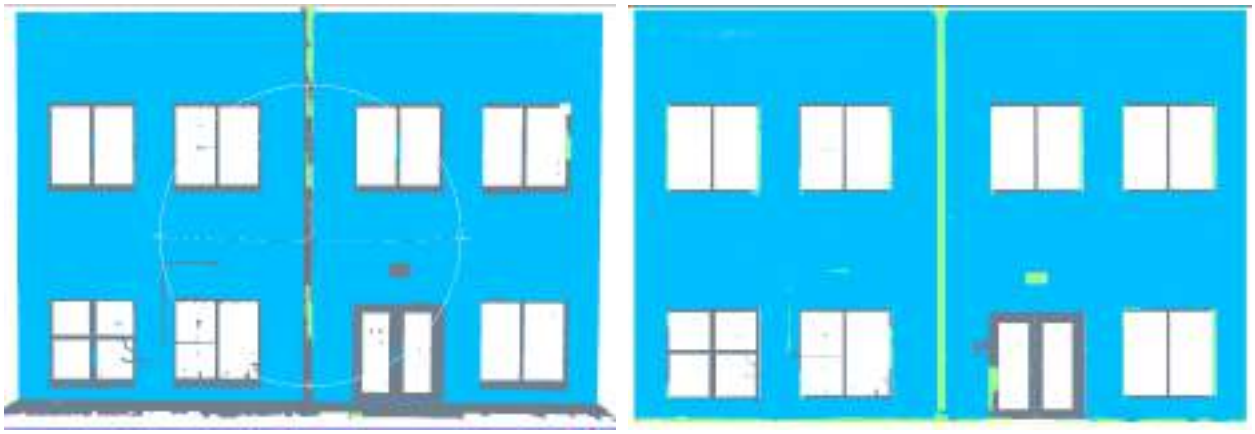
Figure 5. Comparison of training metrics from SPT

mean Intersection over Union (mIoU): this metric provides a class-wise measure of segmentation accuracy. This metric is particularly effective for semantic segmentation, as it evaluates the overlap between predicted and ground-truth points for each class.

The SPT model was trained for up to 1000 epochs with an early stopping criterion to prevent over-fitting. Pointnet++, on the other hand, was trained for 50 epochs. A standard batch size of 32 was utilized for both models, balancing computational efficiency and effective gradient updates. Figure 5 shows the three different aforementioned metrics during the training process of the SPT model for baseline comparisons. It is noteworthy that overall accuracy quickly reaches a high value >80% largely due to the bias towards wall points. Meanwhile, the mIoU values do not reach their maximum until the end of the training phase.

5 Results and Discussion

The training and testing experiments were conducted on a system equipped with an NVIDIA RTX 6000 Ada Generation GPU, Intel Xeon Gold 6442Y CPU, and 128 GB of RAM. Training times varied depending on the model complexity and number of epochs, with PointNet++ com-



(a) Pointnet++ prediction on down-sampled wall points

(b) SPT prediction on actual points

Figure 6. Comparison of model predictions.

pleting in approximately 3 hours per run and Superpoint Transformer taking around 2 hours per run. From our experiments, we observed that well-represented classes with planar geometries (such as walls) are correctly segmented in most of the instances. Figure 6b shows, the segmentation results for the Pointnet++ (left) and SPT (right). The improved segmentation accuracy of SPT model resulted from careful feature fine-tuning, where additional point features (such as RGB, intensity, and elevation) were integrated beyond the default set used in the original implementation.

Note that, in general, the wall points (denoted in blue color) are well identified in both models. However, opening points (grey color) were more accurately segmented using SPT. In the case of Pointnet++, such points were not distinctly separated from the third class (green color)which corresponded to other protrusions, such as gutters or ground platforms. The testing error metrics shown in Table 1 demonstrates that the SPT has a superior performance for solving the PCSS problem for building envelopes. This improved performance is primarily attributed to the SPT's attention-based architecture, which effectively captures detailed spatial relationships and long-range contextual features. Additionally, incorporating radiometric information, such as RGB and intensity values, enables richer feature representations, significantly improving segmentation accuracy for facade elements like windows and doors. It is important to note, however, that RGB features do not inherently resolve class imbalance; instead, the enhanced accuracy from radiometric features improve overall model performance.

After the PCSS problem is solved, the digital twin of the building envelope is created by finding the minimum bounding box of the different classes. Figure 7 shows the bounding box for each opening. Note that the SPT model struggled with accurate predictions along the boundaries of windows and doors, where the transition between classes introduces ambiguities leading to partition errors that cannot be fully corrected with the current model. This boundary defines the transition between classes and lacks a distinct geometric or feature discontinuities where two or more classes intersect. Table 2 compares the dimensions of the digital twin extracted using the SPT model versus the original one defined using the Auto-CuBES algorithm. The opening are number from 1 to 8, going from left to right, and from bottom to top. Note that the bottom openings show the largest errors, while the top openings show a low height error and large width error. Upon close inspection, the errors are mainly due to outliers in the opening classes. One of the last steps in Auto-CuBES involve removing outliers from each identified cluster, drastically increasing its accuracy. The SPT model, on the other hand, its very efficient at identifying the main points belonging to a particular cluster, but cannot completely remove the outliers within a cluster. This exemplifies the limitations of relying solely on a deep learning model, and directs our research towards a more robust PCSS solution that uses both neural networks and unsupervised clustering techniques.

Table 1. Testing error metrics on ORNOBEL Dataset

Metrics	mIOU	OA%
Pointnet++ (downsampled)	93	97%
Superpoint Transformer	97.3	98.7%

6 Conclusion and Future Work

In this work, we present ORNOBEL, a high density, labeled 3D facade segmentation dataset specifically de-

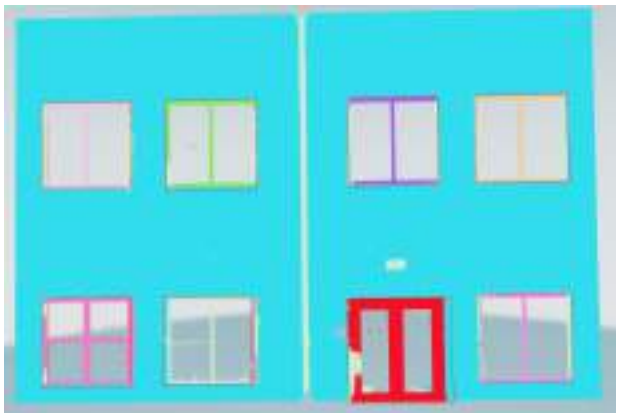


Figure 7. Bounding box applied to SPT results

Table 2. Comparison between digital twin dimensions generated by SPT and Auto-CuBES

Open	SPT (m)		Auto-CuBES		Error (mm)	
	Height	Width	Height	Width	Hght	Wdth
1	2.520	2.211	2.186	1.947	334	264
2	2.141	2.134	1.945	1.864	196	27
3	2.135	1.879	1.943	1.838	192	41
4	1.924	1.866	1.943	1.835	-19	31
5	1.968	1.954	1.956	1.856	12	98
6	1.946	1.940	1.945	1.854	1	86
7	1.960	1.886	1.944	1.843	16	43
8	1.964	1.886	1.952	1.854	16	43

signed for building exterior envelopes, addressing the lack of labeled datasets for building retrofitting applications. Our findings highlight the potential of 3D PCSS in generating precise Building Information Models (BIMs) for retrofitting applications, particularly for designing and deploying prefabricated overclad panels to enhance building energy efficiency. Based on the observations from 2D image-based facade segmentation domain [24, 2], we believe that our 3D facade segmentation dataset will contribute to the further development of 3D facade-oriented methods for construction domain. Results from the deep learning models tested show that a transformer-based architecture has a good potential for solving the semantic segmentation problem in dense building envelope point clouds. However, we plan to extend our work by further training on more state-of-the-art models in combination with various sampling methods and clustering techniques to further address our current limitations. For future work, we aim to expand our dataset by including additional building scans with diverse architectural styles and materials, removing PCA to generalize and address class boundary limitations. Additionally, we will open-source our dataset to address the lack of labeled resources in this domain, fostering collaboration and innovation in 3D facade segmentation. Future exploration will also involve developing methods for faster and improved real-time segmentation capabilities for on-site retrofitting applications through accelerated pipelines.

Acknowledgements

This research was supported by the DOE Office of Energy Efficiency and Renewable Energy (EERE), Building Technologies Office (BTO), under the guidance of Sven Mumme, and used resources at the Building Technologies Research and Integration Center, a DOE-EERE User Facility at Oak Ridge National Laboratory (ORNL). This research was also sponsored by the Laboratory Directed Research and Development Program of ORNL, managed by UT-Battelle, LLC, for the U. S. Department of Energy

References

[1] Iea. Renovation of near 20% of existing building stock to zero-carbon-ready by 2030 is ambitious but necessary – analysis, Sep 2022.

[2] Menglin Dai, Wil O.C. Ward, Gregory Meyers, Danielle Densley Tingley, and Martin Mayfield. Residential building facade segmentation in the urban environment. Building and Environment, 199:107921, 2021. ISSN 0360-1323. doi:https://doi.org/10.1016/j.buildenv.2021.107921. URL https://www.sciencedirect.com/science/article/pii/S0360132321003243.

[3] Charles R. Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation, 2017. URL https://arxiv.org/abs/1612.00593.

[4] Lukas Rauch and Thomas Braml. Semantic point cloud segmentation with deep-learning-based approaches for the construction industry: A survey. Applied Sciences, 13(16), 2023. ISSN 2076-3417. doi:10.3390/app13169146. URL https://www.mdpi.com/2076-3417/13/16/9146.

[5] Charles R Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. arXiv preprint arXiv:1706.02413, 2017.

[6] Damien Robert, Hugo Raguet, and Loic Landrieu. Efficient 3d semantic segmentation with superpoint transformer, 2023. URL https://arxiv.org/abs/2306.08045.

[7] Anh Nguyen and Bac Le. 3d point cloud segmentation: A survey. In 2013 6th

IEEE Conference on Robotics, Automation and Mechatronics (RAM), pages 225–230, 2013. doi:10.1109/RAM.2013.6758588.

[8] Chao Yin, Boyu Wang, Vincent J.L. Gan, Mingzhu Wang, and Jack C.P. Cheng. Automated semantic segmentation of industrial point clouds using respointnet++. Automation in Construction, 130:103874, 2021. ISSN 0926-5805. doi:https://doi.org/10.1016/j.autcon.2021.103874. URL https://www.sciencedirect.com/science/article/pii/S0926580521003253.

[9] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In Conference on Computer Vision and Pattern Recognition (CVPR), 2012.

[10] Whye Kit Fong et al. Panoptic nuscenes: A large-scale benchmark for lidar panoptic segmentation and tracking. 2021.

[11] Xuehan Xiong, Antonio Adan, Burcu Akinci, and Daniel Huber. Automatic creation of semantically rich 3d building models from laser scanner data. Automation in Construction, 31:325–337, 2013. ISSN 0926-5805. doi:https://doi.org/10.1016/j.autcon.2012.10.006. URL https://www.sciencedirect.com/science/article/pii/S0926580512001732.

[12] O. Wysocki, L. Hoegner, and U. Stilla. Tumfacade: Reviewing and enriching point cloud benchmarks for facade segmentation. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVI-2/W1-2022: 529–536, 2022. doi:10.5194/isprs-archives-XLVI-2-W1-2022-529-2022.

[13] Timo Hackel, N. Savinov, L. Ladicky, Jan D. Wegner, K. Schindler, and M. Pollefeys. SEMANTIC3D.NET: A new large-scale point cloud classification benchmark. In ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, volume IV-1-W1, pages 91–98, 2017.

[14] Weikai Tan, Nannan Qin, Lingfei Ma, Ying Li, Jing Du, Guorong Cai, Ke Yang, and Jonathan Li. Toronto-3D: A large-scale mobile lidar dataset for semantic segmentation of urban roadways. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pages 202–203, 2020.

[15] F. Matrone, A. Lingua, R. Pierdicca, E. S. Malinverni, M. Paolanti, E. Grilli, F. Remondino, A. Murtiyoso, and T. Landes. A benchmark for large-scale heritage point cloud semantic segmentation. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII-B2-2020:1419–1426, 2020.

[16] Olaf Wysocki, Jiarui Zhang, and Uwe Stilla. TumfaÇade, 2021. URL https://mediatum.ub.tum.de/1636761.

[17] Kushankur Ghosh, Colin Bellinger, Roberto Corizzo, Paula Branco, Bartosz Krawczyk, and Nathalie Japkowicz. The class imbalance problem in deep learning. Machine Learning, 113 (7):4845–4901, Jul 2024. ISSN 1573-0565. doi:10.1007/s10994-022-06268-8. URL https://doi.org/10.1007/s10994-022-06268-8.

[18] Haibo He and Edwardo A. Garcia. Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering, 21(9):1263–1284, 2009. doi:10.1109/TKDE.2008.239.

[19] Justin M. Johnson and Taghi M. Khoshgoftaar. Survey on deep learning with class imbalance. Journal of Big Data, 6(1), Mar 2019. doi:10.1186/s40537-019-0192-5.

[20] CloudCompare Contributors. CloudCompare (version 2.13.2). http://www.cloudcompare.org/, 2024. GPL software.

[21] Bryan P. Maldonado, Nolan W. Hayes, and Diana Hun. Automatic point cloud building envelope segmentation (auto-cubes) using machine learning. Proceedings of the International Symposium on Automation and Robotics in Construction (IAARC), Jul 2023. doi:10.22260/isarc2023/0009.

[22] F. et al. Pedregosa. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12:2825–2830, 2011.

[23] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. CoRR, abs/1505.04597, 2015. URL http://arxiv.org/abs/1505.04597.

[24] Nariman Habili, Ernest Kwan, Weihao Li, Christfried Webers, Jeremy Oorloff, Mohammad Ali Armin, and Lars Petersson. A hyperspectral and rgb dataset for building facade segmentation, 2022. URL https://arxiv.org/abs/2212.02749.

MAP: A Versatile Platform for Predictive Algorithms in Construction and Beyond

Ramyani Sengupta[1], Emad Elwakil[1] and Yi Jiang[1]

[1]School of Construction Management Technology, Purdue University, USA
sengup11@purdue.edu, eelwakil@purdue.edu, jiang2@purdue.edu

Abstract– The construction industry in North America is a dynamic and rapidly evolving sector characterized by its fast-paced operations and constantly shifting demands. Despite its significance, the industry has been slow to adopt emerging technologies compared to other fields. This lag extends to the integration of artificial intelligence (AI), where vast opportunities for innovation remain underutilized. While large, well-funded firms have begun leveraging AI to develop proprietary applications capable of analyzing historical data to predict patterns and trends, smaller firms often encounter significant barriers, including limited resources and technical expertise, which hinder their ability to adopt such advanced tools.

This study introduces a robust, user-friendly platform that integrates commonly employed machine learning algorithms. The platform empowers smaller firms by enabling them to train, validate, and test their data effectively without requiring extensive technical expertise. Beyond addressing the immediate needs of the construction sector, the platform is intentionally designed for versatility and adaptability, making it applicable across various industries. It supports tasks such as pattern recognition and data visualization, providing a reliable solution for non-experts to evaluate and select the most suitable models for their unique datasets and variables.

Keywords – Modeling, Algorithms, Machine learning; Simple Neural Network, Multi-Layer Perceptron, Support Vector Machine, Random Forest, Regression, Construction

1 Introduction

The construction industry in North America has become a dynamic and rapidly growing sector, with its profits soaring to an impressive $1.8 trillion in 2023 [1] and contributing to 4.4% of the GDP in the United States in 2023 [1]. This remarkable growth highlights the industry's critical role in driving regional economic development and infrastructure advancements. However, in an environment of increasing competition and evolving market demands, every organization within the industry must strive to carve out a distinctive edge that sets it apart from its closest competitors. This differentiation often hinges on how companies embrace and integrate innovative technologies into their operations. Technology adoption has emerged as a key factor in determining which firms can stay ahead of the curve, enhancing overall performance [2]. In an industry where traditional methods and opinions can still hold considerable sway in many areas, the construction industry has often fallen behind in this area [3], and today, the ability to leverage cutting-edge tools and systems can be the most defining characteristic of successful organizations.

Alongside automation, robotics, and drone technology, to name a few, artificial intelligence or 'AI' algorithms has become a quickly growing technology across all industries. Adopting it within the construction industry can help improve efficiency, safety, and overall project outcomes by streamlining project planning, resource management, and real-time decision-making. Currently, available AI tools can improve performance, make firms more competitive, and thus give them an 'edge' [4]. As more and more US construction companies invest in in-house AI algorithms for their historical data to predict project timelines and future issues, this paper aims to streamline the process by creating a platform with different ML algorithms. As it can be challenging to determine which algorithm works best for different types of data, the 'Modelling and Prediction Platform ('MAP' platform) application can help determine the best-fit model for the data. Users can run their data through MAP and easily choose among six different machine learning algorithms (a simple neural network, multi-layer perceptron, support vector machine with five different kernel options, random forest, ridge regression, and lasso regression) to determine which model – with the least amount of error – best fits the data and predicts a pattern between the variables. Certain algorithms have additional parameters that can be used to customize the users' choices. This application can be used for datasets beyond the construction industry, as its visualization, modeling,

and predictive techniques are not limited to construction datasets or project management.

2 AI and Construction

Although challenging, multiple attempts have been made to define AI and its function. "*A being concerned with methods of achieving goals in situations in which the information available has a certain complex character*" was used by McCarthy in a 1988 textbook as an attempt to define artificial intelligence. Since then, its definition has constantly evolved, and its functions have surpassed expectations. Although it can still be termed a comparatively modern form of technology [5], the number of companies using AI technologies has increased by 270% between 2014 and 2018, almost a three-fold increase [6]. This begs the obvious question today: *if everyone is using it, how does it lend a competitive edge*? Mikalef and others, in their 2021 study, argue that in-house tools, i.e., AI capabilities developed by firms themselves, can enhance performance and give companies an edge [4]. Such engineered solutions can allow project managers to track project progress, resolve conflicts and issues, and make informed decisions from any location. A good example of such 'in-house' tools is using historical data for predictions. Some of the heavyweight construction firms in the country, such as Turner Construction, Suffolk Construction, DPR Construction, and AECOM, among others, routinely train AI algorithms nowadays on their proprietary company data from previous decades to predict future scheduling timelines, delay causes, project performance, etc. The idea is that such algorithms can quickly be trained to analyze historical data to detect current issues a*nd predict potential issues* in construction projects. Many of the top 10 construction companies in the US (big names in the industry such as Turner, AECOM, Suffolk, etc.) have investing 'wings' within their firm that heavily research these topics and then infuse capital into startups focusing on specific tools that can improve their projects. This can help enable proactive planning and resource allocation within the company *before* such issues arise. Due to budget constraints and lack of manpower, smaller names often fall behind, and the tool described in this study may help them compete.

2.1 Background on AI Adoption

Case studies can significantly help strengthen the argument that incorporating AI techniques and tools improve company performance and can reiterate the importance of more compact firms having access to such technology. A 2020 study including 150 firms of various sizes investigated the use of eleven types of AI capabilities (generic AI such as task and decision automation, machine learning algorithms, deep learning algorithms, cognitive, cognitive cyber security, natural language processing, robotic personal assistant, pattern recognition, chatbots, neural networks, and virtual companions). The study found that AI use within firms had an incredibly positive result with increased organizational performance at the company level, and process-related variables also improved the company [7]. Professional readiness and AI technologies correlate as well [8]. Another 2022 study investigated 299 Scandinavian countries and found that AI platforms strongly impact organizational performance, whether in the social aspect or financial performance [2]. A 2021 study among 530 employees across Indian telecommunication companies (Vodafone, Idea Telecom, Bharti Airtel, Reliance Jio, etc.) reported that using AI applications and algorithms helps managerial employees get real-time data on employee performance quite effectively and that automated processes and other AI tools can influence performance. Companies can "overachieve", "deliver sustainable results", "optimize resources better", and "do the right thing" far more often if they use AI tools to drive performance [9]. Another empirical study from 2022 among 380 employees of Dubai-based companies found that AI use had positive relationships with organizational structure, process innovation, and employee expertise/performance.

When combined with onsite tools/field work, AI technologies can use technologies like tracking and computer vision algorithms for remote inspection while also maintaining quality assurance/quality control (QA/QC) [10]. Construction robotics and automation also transform how resources are used on construction sites [11]. Robotic systems can perform repetitive or dangerous tasks, such as bricklaying, concrete pouring, or welding, with greater precision and speed than human workers [10, 12, 13]. This increases productivity and reduces the risk of human error and accidents. Today, a combination of sensors and algorithms can be used for cranes (e.g., versatile.ai), which connect to a local IoT system for easy tracking by all stakeholders in a project. Currently, versatile.ai is almost unanimously favored by the top construction companies in the US for site work and remote supervision.

AI technologies, thus, show promise when used for a myriad of onsite and offsite work in construction. It is no surprise that large firms across North America are spending millions of dollars on AI tools for all kinds of construction-related variables – from office productivity to onsite safety and PPI use. As discussed earlier, not all firms are large enough to fund decades of AI research, and the platform discussed in this paper can help smaller firms overcome the labor of building complex tools for their historical project data – a platform like MAP can help them determine which algorithms can help predict and highlight any existing patterns in their datasets.

3 The 'MAP' Platform

Six algorithms were chosen for this platform due to their already well-established status in data analysis – they are well-documented in literature and can work for varying data types. The application, named 'Modelling & Prediction Platform' or 'MAP,' displays an error histogram and a graph between the actual data and predicted data alongside the Mean Squared Error and R-squared coefficient values. This can be used to predict future technology adoption within the industry. The findings can help offer a new appreciation for the challenges and opportunities that may be analogous to adopting AI in construction. The application is currently in its testing phase, and a case study conducted with MAP and real construction data is ongoing and thus is beyond the scope of the current discussion.

3.1 Algorithms in MAP

Keeping in mind the dynamic needs of the construction industry, the following six algorithms have since been encoded into the MAP UI (written in Python ver3.12.6, with sci-kit-learn and Matplotlib – discussed later in this section):

- Simple Neural Network
- Multi-layer Perceptron
- Support Vector Machine (with five kernel functions the user can choose from – linear, polynomial, sigmoid, precomputed, and radial basis function or rbf)
- Random Forest
- Ridge Regression
- Lasso Regression.

A Simple Neural Network is great for learning patterns in data that traditional algorithms might miss. Its main advantage is flexibility—it can adapt to many types of data, like images, text, or numerical data [14] (*in the current version of the MAP UI, however, only numerical data can be used as input and/or training and/or test variables*). It is also quite robust to noise, meaning it can handle data that is not perfectly clean (*again, it is advised to use cleaned data for the MAP platform for good visualizations*). With the right training, it can generalize well to predict new, unseen data. The Multi-Layer Perceptron can be considered a powerful extension of simple neural networks. Its biggest benefit is its ability to model complex relationships by using multiple layers of interconnected nodes. MLPs can solve problems that are not linearly separable, making them ideal for tasks like image classification or natural language processing [15]. They also perform well with large and diverse datasets, which can be helpful for construction companies.

Support Vector Machines, or SVMs, are highly effective for classification and regression tasks. SVMs are particularly good with small to medium-sized datasets and are less prone to overfitting. One of their strengths is the flexibility provided by kernel functions, which lets them work well with both linear and non-linear data [16, 17]. Kernel functions help to convert input data into the required hyperplane/ multidimensional space, and the five kernels that are included in the current version of the MAP model (that can be chosen with ease by the user, depending on their dataset, variable types and/or any modeling requirements) are:

- The linear kernel is an ideal option for simple, linearly separable data,
- The Polynomial kernel excels at capturing complex relationships,
- The sigmoid kernel mimics the activation function in neural networks,
- The precomputed kernel allows customization based on prior knowledge and
- The Radial Basis Function (RBF) is versatile for non-linear patterns.

Random Forest is a versatile and reliable algorithm with extensive uses in classification and regression. Its primary benefit is robustness – it combines the predictions of many decision trees to improve the overall accuracy of the output and reduce overfitting. It also does well at handling missing data and can work with numerical and categorical variables [17]. Additionally, Random Forest provides feature importance scores, helping identify which features matter most [17, 18]. Ridge Regression improves upon linear regression by adding a penalty to large coefficients, and its key benefit is reducing overfitting, which makes it ideal for datasets with many correlated features [19]. This regularization ensures that the model generalizes new data better. Ridge Regression can be particularly useful when all input features contribute to the outcome but need their influence 'moderated' to a certain extent [20]. The final algorithm, called 'Lasso Regression,' is similar to Ridge Regression but has an additional benefit: it can select features by shrinking some coefficients to zero. This makes it a great choice for simplifying models and improving interpretability, especially when dealing with datasets that have many irrelevant features [21]. Thus, the model can be computationally efficient while excelling in identifying the most important variables. All machine learning models in MAP were encoded into the platform using 'sci-kit-learn' (also referred to as 'sklearn'), a freely available and easy-to-use open-source machine learning library kit for the Python programming language, and all visualization techniques were incorporated into the platform by using 'Matplotlib' (a blend of 'MATLAB' – the popular programming environment, 'plot', and 'library') which is another commonly used free, open-source plotting library for the Python programming

language. The user interface was built entirely with Python (the code editor was Visual Studio Code). The user interface (or 'UI') was built to import a comma-separated values (csv) file, read and identify each variable from each column, and then automatically display them in a window for the user to choose input and output variables. The following section discusses the UI in detail.

3.2 Methodology of MAP

The following sections outline the processes behind the user interface design, each option in the application window, the algorithms and their testing alongside the function of each output window.

3.2.1 The UI

The MAP platform is designed to take in a .csv file as an entry (toggle *1. Select .csv file*), parsing the variable list and automatically identifying the variables. Once the file is chosen, the window beneath this option will display a complete list of all the variables. The user can then choose which variables to choose as 'input' options and which elements are to be displayed as 'output' options, and they appear in the next two terminals (underneath the buttons '*2. Choose input variables'* and '*3. Choose output variables'*). A list of algorithms is provided on the right (*4. Choose a model*) from which the user can select any option to train and validate data from the .csv file of their choosing. The figure below shows what the interface of the modeling platform looks like:



Figure 1. The MAP Platform

As discussed earlier, certain algorithms require the user to make further choices. Additional parameters appear in windows like the one below (Figure 2 below shows an example if the user chooses the Support Vector Machine option):



Figure 2. Supplementary parameters for some algorithms

The platform also provides a terminal window (Figure 1) that displays the mean squared error and R-squared values to the user. This can be particularly useful to quickly check if the chosen model suits the user's data or if the parameters need to be customized to fit their needs. Since values for both the training and validation datasets are displayed, it can help the user understand if a different modeling algorithm should be tested.

3.2.2 Model testing and validation

As seen in Figure 1, the platform allows each dataset to be split into two sets for training and validation. Depending on the user and the type of model they choose, one can split the data into any ratios (*they must add up to 1.0 to utilize the entire dataset – in the example figure (Figure 1), the ratios are 0.8 and 0.2*). For example, historical project data can be used for training, while current data can be used for the output.

3.2.3 Output windows

The final output is divided into two graphs, one for the error histogram of the output variable and another for the actual data vs predicted data for the chosen output variable. In the actual output, the error histogram is displayed toward the left of the final output window instead of the top, and the graph of actual vs. predicted values is displayed toward the right side of the output window. Output images from a test training and validation set (using a synthetic dataset for preliminary work) are displayed below in Figures 3 and 4 for a fictional variable from a test dataset:

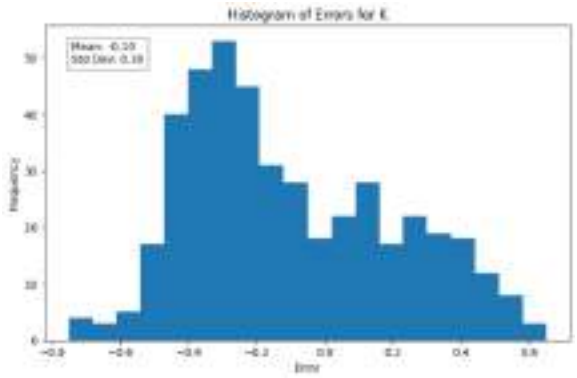


Figure 3. Test output showing an error histogram for a fictional variable '*K*' from a test dataset

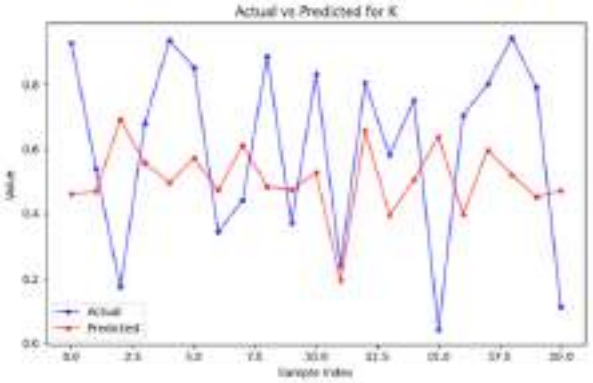


Figure 4. Test output showing the actual vs. predicted data plot for a fictional variable '*K*' from a test dataset

The test dataset shows a weak resemblance between the actual and predicted values, as the dummy dataset does not have identifiable relationships between the variables and were used purely for testing each algorithm. Data from the error histogram was used as one of many identifiers of whether the algorithm is a good choice for the dataset. The values of standard deviation and mean squared and R-squared error (pictured in the top left corner of the error histogram – Figure 3 alongside the terminal window – in Figure 1) also contribute to the user being able to determine of the model is a good fit for their chosen dataset.

4 Conclusion

In stark contrast to other industries, despite such clear positives, stakeholders and investors in this industry are still quite unsure of the potential benefits and overall financial worth of the construction and AI team-up [22]. As the industry changes and adjusts, investigating and investing in AI technologies will be crucial to maintaining competitiveness and meeting the increasing demand for smarter and greener projects. The future of construction lies in the successful adoption of AI, making this an essential area of exploration for the industry. Accurate information from real employees in the industry is crucial for researchers to gain more perspective on the usefulness of AI in construction.

While addressing the current needs of the construction industry, the platform's design is also simple, versatile, customizable, and adaptable, allowing it to be used in a wide range of other industries for tasks such as pattern recognition and data visualization. The platform's customization options also enable users to tailor its functionalities to suit their specific datasets and variables. This flexibility makes it particularly appealing for users who are not overly familiar with ML algorithms, as it provides an accessible way to experiment with different models and determine which one performs best for their unique business needs. The simplicity of the interface also ensures that using it requires little training and can be mostly intuitive. By bridging the gap between AI capabilities and practical usability, this platform offers smaller firms an opportunity to compete on a level playing field with their larger counterparts while encouraging broader AI adoption across various industries. Such algorithms hold immense potential to revolutionize small construction companies, as they can provide firms with evidence-based, data-driven insights that were previously only accessible to larger firms. MAP can help more compact companies optimize their project timelines by using their historical data, similar to the larger firms, and then predict any probable delays or other issues. It can also help to improve estimation by identifying patterns (if any) in past projects, helping employees build more accurate and competitive bids. Most importantly, platforms like MAP need not be limited to any singular industry; the application can be used to predict patterns in other firms looking for a simple application to test out predictive algorithms for their data. Future and ongoing work on MAP includes testing and validating it on real-world construction data from construction companies across the continent.

References

[1] Statista. Construction - statistics and facts. Available at: https://www.statista.com/topics/974/construction/#topicOverview (accessed November 19, 2024).

[2] K. Bley, S. F. B. Fredriksen, M. E. Skjærvik, and I. O. Pappas. The role of organizational culture on artificial intelligence capabilities and organizational performance. In *The Role of Digital Technologies in Shaping the Post-Pandemic World*, pages 13–24. Springer International Publishing, 2022.

[3] A. Sawhney, M. Riley, and J. Irizarry. Construction 4.0: An Innovation Platform for the Built Environment. Routledge, 1st edition, 2020.

[4] P. Mikalef and M. Gupta. Artificial intelligence capability: conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. *Information Management*, 58(3), 103434, 2021.

[5] H. Chen, L. Li, and Y. Chen. Explore success factors that impact artificial intelligence adoption on telecom industry in China. *Journal of Management Analytics*, 8(1), 36–68, 2021.

[6] I. O. Pappas, P. Mikalef, M. N. Giannakos, J. Krogstie, and G. Lekakos. Big data and business analytics ecosystems: paving the way towards digital transformation and sustainable societies. *Information Systems and E-Business Management*, 16(3), 479–491, 2018.

[7] S.-L. Wamba-Taguimdje, S. F. Wamba, J. R. K. Kamdjoug, and C. E. T. Wanko. Impact of artificial intelligence on firm performance: Exploring the mediating effect of process-oriented dynamic capabilities. In *Digital Business Transformation*, volume 38, pages 3–18. Springer International Publishing, 2020.

[8] J. Choudrie and E. D. Zamani. Understanding individual user resistance and workarounds of enterprise social networks: The case of Service Ltd. *Journal of Information Technology*, 31(2, SI), 130–151, 2016.

[9] K. Bley, S. F. B. Fredriksen, M. E. Skjærvik, and I. O. Pappas. The role of organizational culture on artificial intelligence capabilities and organizational performance. In *The Role of Digital Technologies in Shaping the Post-Pandemic World*, pages 13–24. Springer International Publishing, 2022.

[10] U. Sahay and G. Kaur. Leveraging AI to drive organizational performance through effective performance. In *2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics, and Cloud) (I-SMAC)*, pages 584–588, Palladam, India, 2021.

[11] Y. Pan and L. Zhang. A BIM-data mining integrated digital twin framework for advanced project management. *Automation in Construction*, 124, 103564, 2021.

[12] N. Emaminejad and R. Akhavian. Trustworthy AI and robotics: Implications for the AEC industry. *Automation in Construction*, 139, 104298, 2022.

[13] C. J. Liang, T. H. Le, Y. Ham, B. R. Mantha, M. H. Cheng, and J. J. Lin. Ethics of artificial intelligence and robotics in the architecture, engineering, and construction industry. *Automation in Construction*, 162, 105369, 2024.

[14] Y. Pan and L. Zhang. Roles of artificial intelligence in construction engineering and management: A critical review and future trends. *Automation in Construction*, 122, 103517, 2021.

[15] E. Brown, J. Gao, P. Holmes, R. Bogacz, M. Gilzenrat, and J. D. Cohen. Simple neural networks that optimize decisions. *International Journal of Bifurcation and Chaos*, 15(3), 803–826, 2005.

[16] M. C. Popescu, V. E. Balas, L. Perescu-Popescu, and N. Mastorakis. Multilayer perceptron and neural networks. *WSEAS Transactions on Circuits and Systems*, 8(7), 579–588, 2009.

[17] S. Suthaharan. Support vector machine. In Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning, pages 207–235, 2016.

[18] S. J. Rigatti. Random forest. *Journal of Insurance Medicine*, 47(1), 31–39, 2017.

[19] L. Breiman. Random forests. *Machine Learning*, 45, 5–32, 2001.

[20] G. C. McDonald. Ridge regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, 1(1), 93–100, 2009.

[21] W. N. van Wieringen. Lecture notes on ridge regression. *arXiv preprint arXiv:1509.09169*, 2015.

[22] J. Ranstam and J. A. Cook. LASSO regression. *Journal of British Surgery*, 105(10), 1348–1348, 2018.

[23] C. Merschbrock and B. E. Munkvold. Effective digital collaboration in the construction industry: A case study of BIM deployment in a hospital construction project. *Computers in Industry*, 73, 1–7, 2015.

BIM-integrated Path Planning for Semi-autonomous Robots: Remote Real-time Ambient Data Collection in Indoor Spaces

Fahad Iqbal[1] and Shayan Mirzabeigi[1,2*]

[1]Department of Sustainable Resources Management, State University of New York College of Environmental Science and Forestry, Syracuse, NY, United States
[2]Department of Mechanical and Aerospace Engineering, Syracuse University, Syracuse, NY, United States
fiqbal@esf.edu, smirzabeigi@esf.edu

Abstract

Rising global temperatures and more frequent extreme weather events have made air conditioning systems essential for maintaining indoor thermal comfort. Building engineers often rely on real-time data to optimize energy retrofits that can also help improve building thermal resilience against these extreme climate events. However, older buildings often lack the automation commonly found in new structures with wired/wireless sensory networks, making retrofitting both costly and labor-intensive. Therefore, this paper introduces a novel framework that combines Building Information Modeling (BIM) with semi-autonomous mobile robots for efficient real-time indoor ambient data collection. BIM serves as a centralized knowledge base for spatial maps and navigation plans for unmanned ground mobile robot (UGR) equipped with sensors. These robots use advanced navigation and drift-correction algorithms to efficiently collect indoor environmental data, including temperature, humidity, and CO_2 levels, in complex indoor environments. The collected data is visualized on a web-based platform, enabling facility managers to make informed decisions and remotely control HVAC systems for greater operational flexibility. This framework was implemented in a case study demonstrating its capability for collecting spatial and temporal ambient data and visualizing them to support human-centered energy-efficient building operations. The results emphasize the framework's scalability, cost-effectiveness, and use case in various building types.

Keywords:
BIM; IoT; Mobile Robot; UGR; Indoor Environmental Quality

1 Introduction

The growing global demand for energy-efficient indoor environments highlights the critical need for innovative approaches to managing Heating, Ventilation, and Air Conditioning (HVAC) systems, which are among the most energy-intensive components of buildings. With individuals spending nearly 80% of their time indoors [1], buildings account for 40% of total global electricity use [2]. According to the International Energy Agency (IEA), 3.6 billion cooling appliances are currently in use worldwide. This number is projected to surge to 14 billion by 2050 if equitable access to cooling is provided [3]. Additionally, the frequency of extreme weather events is expected to rise and there is a need to address thermal resilience in buildings against these climate events through new construction and retrofitting. These highlight the urgent need for more innovative approaches in the built environment and more efficient and adaptive active and passive building systems. In this context, there is an opportunity to explore technologies that can optimize building energy performance and occupant thermal comfort through advanced monitoring and data-driven control strategies.

The effective optimization of building energy use and thermal comfort requires a holistic understanding of various interconnected systems, including the building's mechanical and electrical components, occupant behaviors, as well as external factors like weather patterns. To achieve this, a significant amount of data must be continuously gathered, analyzed, and managed at different scales (e.g., granular levels such as individual rooms and floors). Prior research has shown that whether in real-time, near real-time, or through batch processing, data-driven insights can enhance energy efficiency and occupant comfort [4]. However, the challenge lies in the efficient collection of high-quality spatial and temporal data, particularly in existing buildings where constraints

on system retrofitting can limit the implementation of modern energy management strategies.

The use of wired and wireless sensor technologies has revolutionized data collection in building [5]. Traditionally, inspectors manually collected data through time-intensive processes that yielded limited information. With the integration of wired and wireless sensors into Building Automation Systems (BAS), the need for manual inspections has been greatly reduced. In newly constructed buildings, sensors are often pre-installed and integrated into the operational infrastructure. However, for older buildings, retrofitting such systems is challenging due to high installation costs, labor-intensive processes, and limitations on building closure for such retrofitting [6]. Wired systems, commonly used in applications like video surveillance, are particularly costly and less feasible for larger structures such as dormitories and office complexes, where extensive wiring and installation efforts are required [7]. These challenges necessitate innovative and technological solutions for cost-effective and flexible data collection. UGR provides a promising alternative, offering a versatile and cost-effective approach to collect real-time data without the need for extensive building updates in existing buildings to understand the existing conditions to make informed decisions. By navigating through building interiors autonomously, these robots can gather critical environmental data, overcoming the constraints of traditional wired systems and manual inspections.

2 Background

The use of sensors for indoor environmental data collection is a fundamental approach to improving energy efficiency and thermal comfort in buildings, particularly in regions with extreme climates. In tropical regions, where the adoption of air conditioning systems is on the rise due to the increase in extreme heat events, sensor networks are essential for optimizing HVAC operations. Sensors such as temperature, humidity, Particulate Matter (PM), Carbon Dioxide (CO_2), and occupancy detectors enable real-time monitoring of indoor environmental conditions, allowing for more precise control of air conditioning systems [8], [9]. Research has demonstrated that sensor-based systems can significantly reduce energy consumption by providing actionable data for adjusting HVAC settings based on actual environmental conditions rather than static, user-determined settings [10]. However, in tropical climates, a common issue is that users tend to "over-control" temperature settings, leading to inefficiencies such as excessive compressor operation, which increases power consumption without necessarily improving comfort [11]. The integration of more intelligent sensor systems can help mitigate such over-control by providing data-driven insights and optimizing temperature regulation.

Building Information Modeling (BIM) combined with robotics is an emerging approach for real-time ambient data collection in buildings, especially in complex or older environments where traditional sensor installation might be challenging due to cost, labor, or installation time requirement [12]. UGR, equipped with sensors, can autonomously navigate indoor spaces to collect data on temperature, humidity, and air quality, feeding this information into BIM and Building Energy Modeling (BEM) tools for analysis and optimization. This integration allows for the creation of detailed, dynamic simulation models that provide accurate spatial and environmental insights, which can then be used to tune building operations (e.g., operation of HVAC systems) [13]. Additionally, UGR can access areas that are difficult or impractical for human intervention, making data collection more efficient and comprehensive. Studies have shown that this approach not only improves thermal comfort but also supports energy-efficient building management, particularly in older structures that lack required levels of automation.

The integration of BIM, sensors, and robotics has the potential to revolutionize building management by offering a holistic solution to the challenges of ambient data collection and energy optimization. Recent studies highlight the interactions between these technologies, where BIM provides a comprehensive spatial framework, sensors deliver real-time environmental data, and robotics ensure efficient and flexible data collection in difficult-to-reach areas [14]. Hu et al. developed "AirScope," a robotic system using distributed deep Q-learning to monitor indoor air quality, effectively reducing data latency and optimizing cooperative route planning [15]. Cao et al. proposed an intelligent construction robot system using BIM-based navigation maps, camera calibration, and A* and TEB algorithms for autonomous navigation and obstacle avoidance, demonstrating its feasibility for advancing construction automation [16].

This paper presents a novel framework integrating BIM with UGR for real-time indoor environmental data collection in indoor spaces, particularly addressing the challenges of retrofitting in older buildings that do not have modern Building Management System (BMS). In this study, spatial maps and navigation plans are generated from the building's BIM model for UGR equipped with lightweight sensors. These UGRs efficiently navigate complex environments to collect critical data such as temperature, humidity, PM and CO_2 levels. The collected data is processed and visualized on a web-based platform, providing facility managers with actionable insights for optimizing HVAC systems. An experimental case study demonstrates the framework's

effectiveness, showcasing its scalability, cost-efficiency, and suitability for diverse building types. The results highlight the framework's potential to transform building operation management practices in existing buildings with minimal infrastructure investments, contributing to global sustainability goals.

3 Research Methodology

This section outlines the conceptual framework and methodology of the study, detailing different phases and system implementation. The study integrates BIM, Internet of Things (IoT) sensors, and semi-autonomous robotics in a case study. BIM semantic data is used for navigation, allowing the UGR to move autonomously. The robot, equipped with IoT sensors, collects environmental data, which is stored in a cloud database and visualized on a web-based BIM platform.

3.1 Framework

The framework begins with the extraction of data from an Industry Foundation Classes (IFC) file, which represents the building's geometry and structure within BIM. This data is processed to create a spatial map of the building, defining key building components such as walls, doors, and rooms. The spatial map enables a semi-autonomous robot equipped with IoT sensors to navigate through the building and collect environmental data, including temperature, humidity, PM, and CO_2 levels. This information is stored in a cloud-based database and accessed via a REST API for visualization on the BIM web platform, allowing facility managers to make real-time adjustments to HVAC systems in order to improve energy efficiency, occupant thermal comfort, and indoor air quality. Figure 1 provides an overview of the framework.

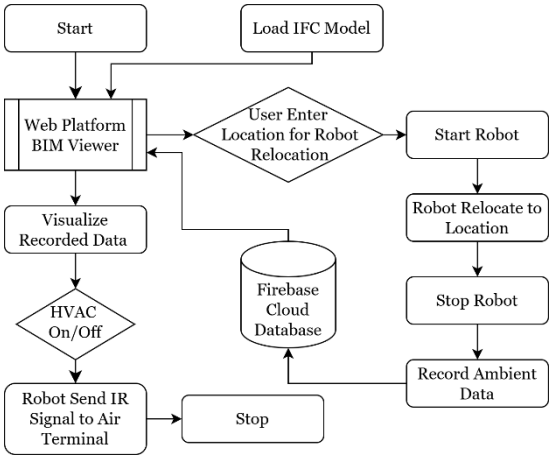


Figure 1. Overview of the proposed framework.

The process begins with loading an IFC file onto the web platform. The user then specifies the robot's starting location, from which it autonomously moves through the building, collecting and uploading data to the Firebase Cloud database. The retrieved data is displayed on the platform in a user-friendly interface, enabling informed, data-driven decisions to enhance building operations and indoor environmental quality.

3.2 Web Platform

The web platform for data visualization and control is built with IFC.js, a JavaScript library that enables the visualization of IFC data in web applications. It displays building geometry, sensor data, and system performance in an intuitive and user-friendly interface. The backend is powered by a Node.js server that runs Python scripts for processing navigation maps and retrieving sensor data from the cloud database to the web platform. For cloud data storage, the platform uses Firebase's Database, a cloud-based NoSQL solution. Firebase's real-time data syncing capabilities ensure instant updates across all connected devices, critical for real-time monitoring applications. Using a combination of TypeScript and Python, the platform offers a scalable solution for monitoring building conditions, allowing facility managers to optimize HVAC operations and remotely control the robot. Figure 2 shows a snapshot of the web platform with two windows: one for visualizing sensor data and controlling HVAC operations, and the other one for robot control and navigation plan.

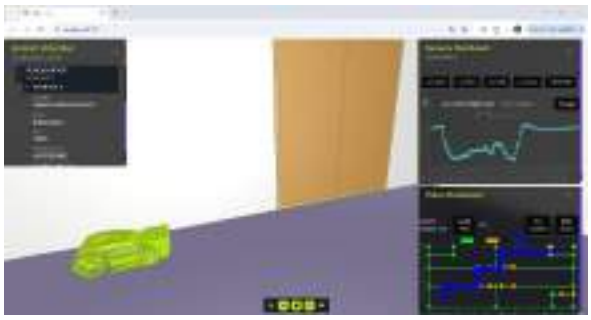


Figure 2. User interface for developed web platform for robot control and ambient data visualization.

3.3 UGR and Sensor Module Development

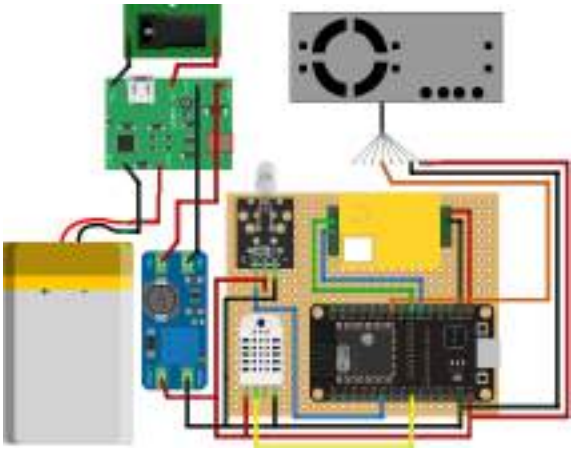
A. UGR Robot Development

The UGR is a compact and efficient platform for autonomous ambient data collection. It features a 256 × 150 × 2 mm chassis with a 500 g load capacity, ensuring a lightweight and durable build. Equipped with 65 mm wheels, it offers smooth mobility across indoor surfaces. The robot is powered by a 3.7V, 5000mAh lithium-ion battery with a step-up converter to supply the required

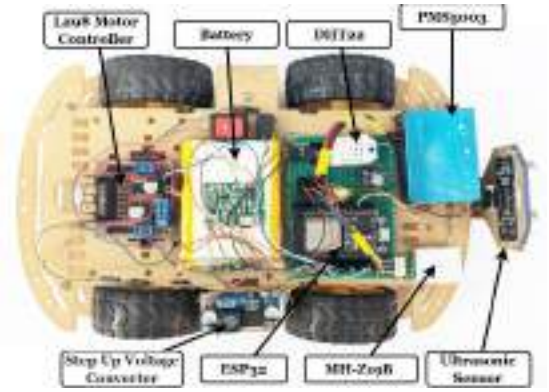
voltage for motors and sensors, providing a runtime of up to 4 hours on a full charge. ESP32-WROOM microcontroller enables wireless communication and real-time sensor data processing, making it an effective solution for environmental monitoring and smart automation applications.

B. Sensing Module Development

The framework includes IoT-based sensors installed on a UGR for ambient data collection. It includes a DHT22 for temperature and humidity recording, PMS5003 for particulate matter measurement, and the MH-Z19B sensor for CO_2 measurement. The DHT22 sensor provides high accuracy for temperature and humidity measurements, with a temperature accuracy range of ±0.5°C and a humidity accuracy of ±2-5%. This makes it ideal for indoor climate monitoring [17]. The PMS5003 sensor delivers real-time particulate matter (PM2.5 and PM10) data with an accuracy of ±10 µg/m³, ensuring a healthy indoor environment [18]. The MH-Z19B is a reliable sensor for measuring CO_2 concentrations, offering an accuracy of ±50 ppm in the range of 400-5000 ppm, contributing to comprehensive indoor air quality monitoring [19]. These sensors are connected to an ESP32-WROOM microcontroller, which interfaces with a Wi-Fi network. Data recorded by the sensors is first stored locally on the ESP32 and then transmitted to a Firebase database via HTTP POST requests, ensuring efficient and reliable data management. Figure 3 illustrates the circuit diagram of the sensor setup and their installation locations on the UGR, providing a clear overview of the system's hardware integration.



(a)



(b)

Figure 3a. Circuit diagram of the sensors installed on UGR; b. Sensors configuration on UGR.

3.4 Navigation Algorithms

The navigation of the UGR adopted in this study is based on the study by Chen et al. [20], which employs the A* algorithm, a widely applied pathfinding method in robotics. A* combines the strengths of Dijkstra's Algorithm and Greedy Best-First Search, enabling efficient navigation in complex environments [21]. This approach allows the robot to balance exploring the shortest paths while considering a heuristic to direct movement toward a target location. The process begins when a user uploads an IFC file to the platform, where different components (e.g., walls, doors, and furnishing) are extracted from the building model. Walls and furnishing are treated as obstacles, while doors are identified as access points, crucial for allowing the robot to navigate between different rooms and spaces.

When the user inputs the target room or location on the developed web platform, the target coordinates are sent to a Python script via a HTTP POST request. The algorithm then calculates the optimal path from the robot's current location to the target destination, ensuring that obstacles are avoided and that the robot can efficiently move to the target point. Once the path is identified, the Python code computes the distances and angles required for navigation. The path is displayed on a 2D graph for the user's review. Upon user acceptance of the path, the path data is uploaded to the cloud database. This data includes the distance the UGR needs to traverse and the angles at which it must turn. Once the path data is successfully stored, the robot initiates its movement towards the targeted location, which may be a space, room, corridor, or other predefined area. Upon arrival at the target location, the robot begins recording relevant data, which is subsequently stored in the cloud database for further analysis and processing. Figure 4. provided shows a building plan where walls are represented as obstacles and doors as access points. The robot's path is illustrated with blue lines, highlighting the

optimal route calculated by the A* algorithm from the starting location to the target room. The figure highlights how the robot navigates around obstacles and through access points, offering a clear representation of its motion path within the building layout.

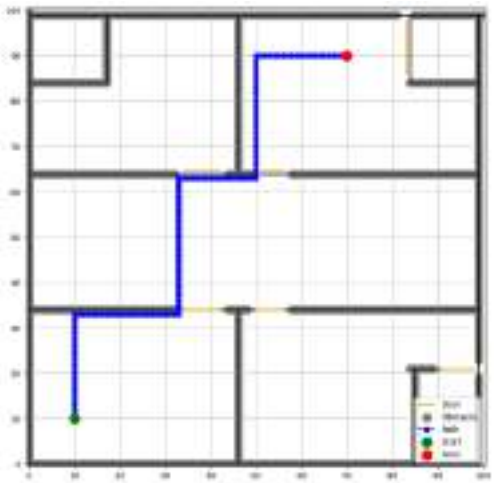
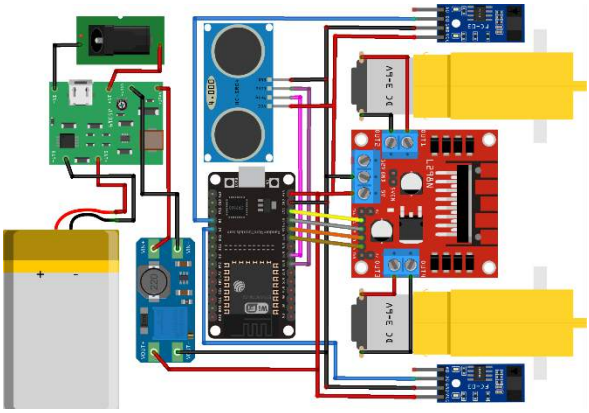


Figure 4. UGR Navigation Path in a 2D Building Layout with Obstacles and Access Points.

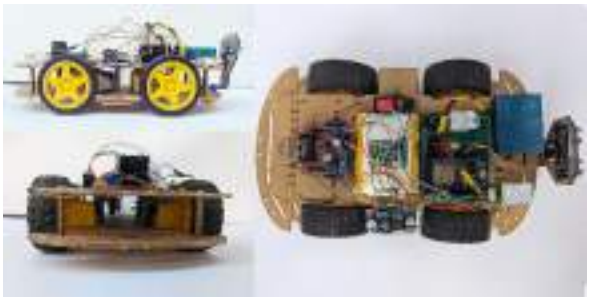
3.5 Robot Working

The UGR used for data collection is a four-wheel-drive vehicle powered by gear motors with speed encoders and is controlled via an LM298 motor driver [22], [23]. The robot's movements are managed by the motor driver and an ESP microcontroller. In addition to motor control, an ultrasonic sensor is employed in front of the robot for obstacle detection, ensuring the robot can avoid unexpected objects and navigate safely in dynamic environments [24]. Given the low-cost nature of the prototype, the dynamic obstacle avoidance approach is simple yet effective. When an obstacle appears in the robot's path, it initially pauses and waits for the obstruction to clear. If the obstacle remains stationary, the UGV adjusts its path by turning left or right to navigate around it, ensuring continued movement without complex real-time mapping. Once the A* algorithm sets the path, the distance and angle the robot needs to cover are stored in a cloud database. When the user clicks the "Start Robot" button on the platform, the ESP retrieves this data and directs the robot to travel the specified distance and angle. The speed encoder counts the wheel revolutions to ensure the robot stops after covering the required distance. The LM298 motor driver enables bidirectional control, allowing the robot to move forward, backward, or rotate, while the speed controller ensures smooth and precise movement.

This cost-effective setup allows the robot to navigate complex indoor environments and collect environmental data efficiently, with minimal disruption. Figure 5a shows the circuit diagram of the developed UGR, illustrating the connections between the ESP, LM298 motor driver, motors, speed encoders, and the power circuit, while Figure 5b displays images of the developed robot, highlighting the configuration of all sensors and hardware components in the actual setup.



(a)



(b)

Figure 5a. Circuit diagram of the UGR; b. Images of the developed robot and hardware configuration.

4 Result and Discussion

4.1 Experimental Setup of Case Study

To test the accuracy and efficiency of the developed framework, an experimental study was conducted in a 9x12m indoor space comprising four rooms. The robot successfully navigated through the test environment, using its integration with BIM to access accurate spatial maps. This integration allowed the robot to traverse complex indoor layouts with ease. The A* algorithm, combined with drift-correction mechanisms, ensured optimized movement, achieving precise path planning and time-efficient navigation. Upon reaching designated locations, the robot began collecting real-time environmental data on temperature, humidity, PM2.5,

PM10, and CO_2 levels. The results demonstrated the framework's reliability, as the robot's trajectory closely followed the planned navigation path in dynamic indoor conditions. However, since the current system relies on non-autonomous navigation, it lacks real-time path adaptation and localization. This odometry-based approach is susceptible to cumulative drift and relies on predefined paths, limiting its ability to handle unexpected obstacles or environmental changes. While effective in structured environments, its performance can be affected in dynamic settings, where real-time localization and adaptive path planning would improve overall navigation accuracy.

4.2 Results and Interpretations

The environmental data collected by the robot was continuously logged over a 24-hour period, with measurements taken every 5 minutes. As shown in Figure 6, the temperature and humidity levels are plotted on a dual-axis graph, clearly illustrating the inverse relationship between these two parameters. The general trend in the data can also be observed with 3rd-degree polynomial illustrated in the graph. The temperature showed moderate fluctuations, ranging from a minimum of 10.1°C to a maximum of 14.0°C, with an average of 12.2°C. In contrast, the humidity remained relatively stable, fluctuating between 50.0% and 72.5%, with an average of 67.6%. This automated data collection ensured accuracy and consistency, eliminating the need for manual measurements and providing a comprehensive dataset.

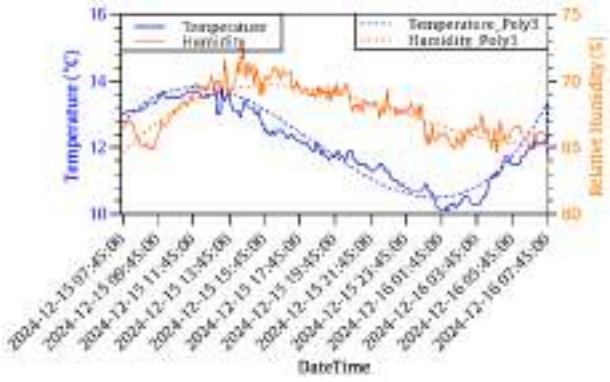


Figure 6. Recorded temperature (°C) and humidity (%) data over 24 hours.

Air quality data, including PM2.5 and PM10, are depicted in Figure 7. The graph demonstrates distinct spikes in particulate matter concentrations at specific times, likely corresponding to activities such as cleaning or movement within the test environment. Despite these occasional fluctuations, the overall air quality remained within acceptable indoor thresholds based on WHO standards. PM2.5 levels for the recorded data ranged from 13.9 to 41.6 µg/m³, with a 24-hour average of 28.7 µg/m³, slightly exceeding the WHO interim target of 25 µg/m³. PM10 levels varied between 15.6 and 47.8 µg/m³, with a 24-hour mean of 33.0 µg/m³, well within the WHO guideline limit of 50 µg/m³. The generalized trend observed in the data is shown using a 3rd-degree polynomial on the graph. These values reflect indoor air quality for the test space, where particulate matter concentrations can vary slightly throughout the day. The use of a dual-axis scale in this graph effectively highlights the differences in magnitude and trends between PM2.5 and PM10 levels, providing a clear representation of the data.

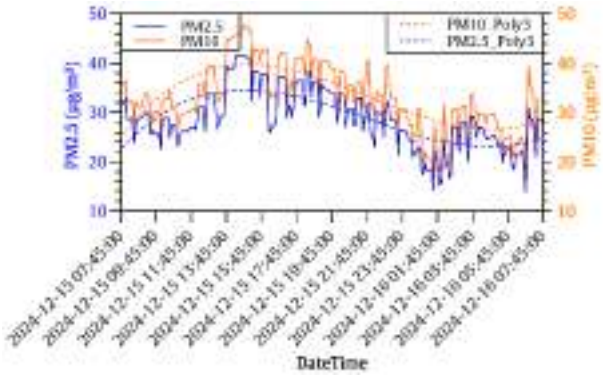


Figure 7. Recorded PM2.5 and PM10 (µg/m³) concentrations over 24 hours.

Figure 8 shows the CO_2 levels, revealing moderate fluctuation, peaking at 1022 ppm with 75% value at 719 ppm. These trends emphasize the importance of adequate ventilation in maintaining healthy indoor air quality. The CO_2 levels in collected data ranged from 446 to 1022 ppm, with an average of 666 ppm, indicating moderate fluctuations throughout the day. Though there are higher fluctuation in data; the overall trend shown on the graph confirms that CO_2 levels remained below critical thresholds, reflecting good ventilation in the test space. The observed indoor environmental data trends not only showcase the robot's ability to collect accurate and real-time data but also provide valuable insights into environmental dynamics over time.

These results highlight the potential of using a UGV for ambient data collection, particularly in retrofit building applications where installing stationary sensors can be costly and labour-intensive. A mobile robot offers a more flexible and cost-effective solution by measuring the spatial distribution of air quality, temperature and humidity across different locations. However, unlike fixed sensors that provide continuous data at a single point, the UGV collects data at intervals as it moves, which may lead to gaps in capturing sudden environmental changes. To mitigate this limitation, the sampling approach should be optimized to balance

spatial coverage and data collection frequency, ensuring a more reliable analysis.

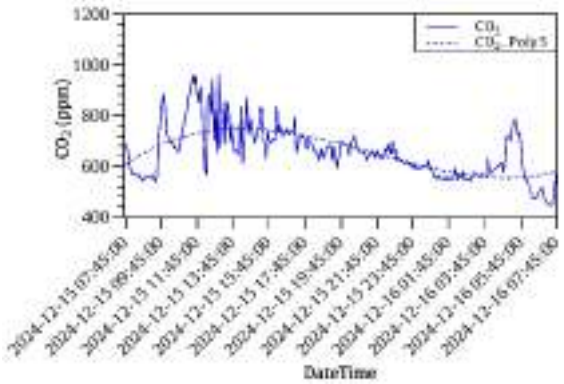


Figure 8. Recorded CO_2 levels (ppm) over 24 hours.

4.3 Limitations

While the framework exhibited promising results, several limitations exist. One notable issue is the robot's reliance on pre-uploaded updated BIM data, which can lead to navigation or data collection inaccuracies if the models are outdated or incomplete. This is certainly an important aspect to be addressed in future work for buildings without any prior BIM documentation. Additionally, the system's performance in highly dynamic or crowded environments has not been fully tested, and testing in such environments may shed light on further adaption to rapid changes or irregular conditions by improving the state-of-the-art algorithms. Another critical limitation is the need for continuous wireless internet access, as the robot depends on real-time data transmission and communication with the server for data processing and navigation updates.

The current sensor setup is limited to a small set of environmental parameters such as temperature, humidity, PM2.5, PM10, and CO_2. Expanding the range of measurable parameters by incorporating additional sensors comparing the data with that from stationary sensors installed in various spaces could provide a more holistic environmental analysis. Furthermore, the accuracy of the robot may vary, as odometry is prone to surface slippage and wheel malfunction, leading to cumulative errors over time. Future research is needed to focus on enhancing navigation algorithms, expanding sensor capabilities, and addressing the requirement for continuous internet access to ensure seamless operation in diverse and dynamic environments.

5 Conclusion

The study presents a novel framework that integrates BIM and semi-autonomous mobile robots to address the challenges of real-time ambient data collection in older buildings with limited automation infrastructure. By using BIM for spatial mapping and navigation, the proposed system enables efficient robot movement and accurate environmental monitoring. The framework demonstrated its ability to collect minimum critical indoor environmental data parameters, including temperature, humidity, particulate matter, and CO_2 levels, in a scalable and cost-effective manner. The real-time data visualization on a web platform further empowers facility managers to optimize HVAC operations, improving both energy efficiency and occupant thermal comfort. Results from the case study confirmed the system's reliability, with consistent data accuracy and efficient robot navigation, even in complex indoor layouts.

While the findings highlight the potential of this framework to transform building management practices, certain limitations must be addressed. The reliance on updated BIM models and consistent wireless connectivity poses challenges for dynamic or crowded environments. Additionally, the limited range of sensors restricts the scope of environmental analysis. Future research should focus on enhancing navigation algorithms for greater adaptability, expanding the sensor suite to cover more environmental parameters, and developing offline capabilities to mitigate connectivity issues. Despite these limitations, this study provides a solid foundation for scalable, real-time ambient data collection, contributing to sustainable and resilient building operations.

References

[1] P. H. V. Reddy, S. M. Hussain, R. S. Reddy, and R. Vinay, "Iot based Indoor Environment Monitoring and Controlling with Plant care," in *2024 IEEE 5th India Council International Subsections Conference (INDISCON)*, IEEE, Aug. 2024, pp. 1–5. doi: 10.1109/INDISCON62179.2024.10744357.

[2] T. Leader *et al.*, "Annual energy outlook 2014," *U.S. Energy Information Addministration*, 2014.

[3] Maciej Serda, F. G. Becker, M. Cleary, and Al. Et, "Cooling Emissions and Policy Synthesis Report," 2020.

[4] C. C. Menassa, N. Taylor, and J. Nelson, "A framework for automated control and commissioning of hybrid ventilation systems in complex buildings," *Autom Constr*, vol. 30, 2013, doi: 10.1016/j.autcon.2012.11.022.

[5] S. Raza, M. Faheem, and M. Guenes, "Industrial wireless sensor and actuator networks in industry 4.0: Exploring requirements, protocols, and

challenges—A MAC survey," *International Journal of Communication Systems*, vol. 32, no. 15, 2019, doi: 10.1002/dac.4074.

[6] M. Demirbas, "Wireless Sensor Networks for Monitoring of Large Public Buildings," *COMPUTER NETWORKS*, vol. 46, 2005.

[7] Y. Yu, M. Gola, G. Settimo, M. Buffoli, and S. Capolongo, "Feasibility and Affordability of Low-Cost Air Sensors with Internet of Things for Indoor Air Quality Monitoring in Residential Buildings: Systematic Review on Sensor Information and Residential Applications, with Experience-Based Discussions," *Atmosphere (Basel)*, vol. 15, no. 10, p. 1170, Sep. 2024, doi: 10.3390/atmos15101170.

[8] D. Robinson, D. Sanders, and E. Mazharsolook, "Sensor-based ambient intelligence for optimal energy efficiency," *Sensor Review*, vol. 34, no. 2, 2014, doi: 10.1108/SR-10-2012-667.

[9] S. Mirzabeigi, S. Soltanian-Zadeh, B. Carter, and J. Zhang, "Towards Occupant-Centric Building Retrofit Assessments: An Integrated Protocol for Real-Time Monitoring of Building Energy Use, Indoor Environmental Quality, and Enclosure Performance in Relationship to Occupant Behaviors," *ASHRAE Trans*, vol. 130, pp. 70–78, 2024.

[10] Y. Bae *et al.*, "Sensor impacts on building and HVAC controls: A critical review for building energy performance," 2021. doi: 10.1016/j.adapen.2021.100068.

[11] A. M. Ali, S. A. A. Shukor, N. A. Rahim, Z. M. Razlan, Z. A. Z. Jamal, and K. Kohlhof, "IoT-Based Smart Air Conditioning Control for Thermal Comfort," in *2019 IEEE International Conference on Automatic Control and Intelligent Systems, I2CACIS 2019 - Proceedings*, 2019. doi: 10.1109/I2CACIS.2019.8825079.

[12] B. R. K. Mantha, C. C. Menassa, and V. R. Kamat, "Robotic data collection and simulation for evaluation of building retrofit performance," *Autom Constr*, vol. 92, 2018, doi: 10.1016/j.autcon.2018.03.026.

[13] Y. Geng *et al.*, "Robot-based mobile sensing system for high-resolution indoor temperature monitoring," *Autom Constr*, vol. 142, 2022, doi: 10.1016/j.autcon.2022.104477.

[14] A. Hamieh, A. Ben Makhlouf, B. Louhichi, and D. Deneux, "A BIM-based method to plan indoor paths," *Autom Constr*, vol. 113, 2020, doi: 10.1016/j.autcon.2020.103120.

[15] Z. Hu, S. Cong, T. Song, K. Bian, and L. Song, "AirScope: Mobile Robots-Assisted Cooperative Indoor Air Quality Sensing by Distributed Deep Reinforcement Learning," *IEEE Internet Things J*, vol. 7, no. 9, 2020, doi: 10.1109/JIOT.2020.3004339.

[16] F. Cao and X. Liu, "BIM-based Global Path Planning and Obstacle Avoidance Research for Building Inspection Robots," in *2024 IEEE 4th International Conference on Electronic Technology, Communication and Information (ICETCI)*, IEEE, May 2024, pp. 1392–1397. doi: 10.1109/ICETCI61221.2024.10594429.

[17] I. Udrea, V. I. Gheorghe, C. G. Alionte, A. M. Dogeanu, and L. A. Cartal, "IoT solution for thermal comfort and air quality monitoring," in *15th International Conference on Electronics, Computers and Artificial Intelligence, ECAI 2023 - Proceedings*, 2023. doi: 10.1109/ECAI58194.2023.10194230.

[18] D. Suriano and M. Prato, "An Investigation on the Possible Application Areas of Low-Cost PM Sensors for Air Quality Monitoring," *Sensors*, vol. 23, no. 8, 2023, doi: 10.3390/s23083976.

[19] G. Coulby, A. K. Clear, O. Jones, and A. Godfrey, "Low-cost, multimodal environmental monitoring based on the Internet of Things," *Build Environ*, vol. 203, 2021, doi: 10.1016/j.buildenv.2021.108014.

[20] Z. Chen, K. Chen, and J. C. P. Cheng, "A BIM-integrated Indoor Path Planning Framework for Unmanned Ground Robot," in *Proceedings of the International Symposium on Automation and Robotics in Construction*, 2021. doi: 10.22260/isarc2021/0105.

[21] J. A. Subramanian *et al.*, "Integrating Computer Vision and Photogrammetry for Autonomous Aerial Vehicle Landing in Static Environment," *IEEE Access*, vol. 12, 2024, doi: 10.1109/ACCESS.2024.3349419.

[22] B. Debnath, A. Ghosh, and S. Deb, "Designing a Rapid Optical Response Trigger(RORT) for self-navigating and path explorer robot," in *Procedia Computer Science*, 2020. doi: 10.1016/j.procs.2020.03.365.

[23] F. Iqbal, S. Ahmed, F. Amin, S. Qayyum, and F. Ullah, "Integrating BIM–IoT and Autonomous Mobile Robots for Construction Site Layout Printing," *Buildings*, vol. 13, no. 9, 2023, doi: 10.3390/buildings13092212.

[24] R. Chinmayi *et al.*, "Obstacle Detection and Avoidance Robot," in *2018 IEEE International Conference on Computational Intelligence and Computing Research, ICCIC 2018*, 2018. doi: 10.1109/ICCIC.2018.8782344.

Automated Image Dataset Generation Using Web Scraping and Large Multimodal Models for Construction Applications

Shrouk Gharib[1] and Osama Moselhi[2]

[1] Ph.D. Candidate, Department of Building, Civil and Environmental Engineering, Concordia University, Montréal, QC, Canada.
[2] Professor and Director, Centre for Innovation in Construction and Infrastructure Engineering and Management (CICIEM), Department of Building, Civil and Environmental Engineering, Concordia University, Montréal, QC, Canada.
shrouk.gharib@mail.concordia.ca, moselhi@encs.concordia.ca

Abstract –

The construction industry is undergoing a rapid transformation through the integration of digital technologies and construction. These advancements drive methods like object detection, enabling applications such as progress monitoring and material reuse to support sustainability goals. By fostering innovation, reducing material waste, and promoting circular economy principles, these technologies hold immense potential to revolutionize construction practices. However, the scarcity of dataset availability significantly hinders the development of effective machine learning (ML) applications in construction. Manual data collection remains resource-intensive, error-prone, and time-consuming, creating further barriers. This study introduces an automated methodology that combines web scraping and Large Multimodal Models (LMMs) to address these limitations. Using the Gemini Vision model, over 55,000 images were collected, including 2,691 identified as containing structural steel components—such as beams, columns, trusses, steel frames, and steel structures. Metadata generation achieved 94.8% accuracy, enabling efficient categorization through a rule-based filtering approach. The proposed methodology bridges significant gaps in the current literature by delivering a scalable, reusable framework for dataset development tailored to construction applications. By automating data acquisition and metadata generation, it enhances dataset usability for methods like object detection and applications such as progress monitoring. This research establishes a foundation for advancing AI-driven solutions in construction automation and sustainability, contributing to the industry's digital transformation.

Keywords –

Data Acquisition; Web Scraping; Construction Automation; Large Multimodal Models (LMMs); Metadata Generation; Dataset Scalability; Structural Steel Components; Automated Dataset Development.

1 Introduction

The construction industry is undergoing a transformation by the integration of digital and automated technologies. The advancements enhance efficiency, accuracy, and sustainability in tasks in tasks such as object detection for progress monitoring and material reuse planning [1–3]. These technologies hold immense potential to improve construction workflows, reduce material waste, and promote circular economy principles [4–7]. However, fully leveraging these technologies is hindered by the scarcity of high-quality domain-specific datasets, i.e. structural steel components. This limitation hampers the development, validation, and deployment of machine learning (ML) models for critical applications, including structural steel reuse, deconstruction progress monitoring, and material tracking [8–10].

Existing studies underscore the importance of datasets for progress monitoring for accurately identifying and categorizing construction materials, such as structural steel components, based on their properties, dimensions, and suitability for reuse [11–13]. However, many available datasets suffer from limited scope, inadequate diversity, and insufficient annotation quality, making them unsuitable for training robust ML models. For instance, datasets for structural steel applications often lack crucial elements such as component geometries, detailed metadata, and standardized classifications—key requirements for ML-driven solutions [14]. Moreover, manual data collection methods frequently used in this domain are labor-intensive, time-consuming, and prone to inconsistencies, further compounding the issue [15–17]. These challenges highlight the vital need for automated, scalable methodologies to develop precise and standardized datasets for ML applications.

Web scraping has emerged as a powerful tool to address these challenges by automating the collection of large datasets from diverse online sources [18] This technique has been successfully applied in domains such as healthcare and

natural disaster monitoring, demonstrating its scalability and efficiency [19,20]. In construction, web scraping has shown promise in aggregating information on materials, components, and project designs, offering a viable alternative to manual data collection [21,22]. By scraping images and metadata of structural steel components from online sources, this approach can create datasets with sufficient volume and diversity to support advanced ML applications. Additionally, web scraping addresses challenges like inconsistent webpage structures and data cleaning, ensuring comprehensive and high-quality datasets [23–25].

Recent advancements in artificial intelligence, particularly in Multimodal Models derived from Large Language Models (LLMs) such as GPT-4 and Gemini Vision, further enhance the potential for dataset generation [26]. Multimodal models combine visual and textual processing capabilities, making them uniquely suited for tasks such as metadata generation and image classification [27]. These models analyze both image content and contextual text to produce descriptive metadata, enabling hierarchical organization and effective classification [28,29]. Unlike manual annotation methods, which are error-prone and inefficient, multimodal models offer unparalleled accuracy and scalability when processing large datasets [20]. Furthermore, their ability to analyze and generate structured descriptions from unstructured data makes them indispensable for research in this domain [20,30].

Despite these advancements, three critical gaps remain: (1) a lack of scalable, automated methodologies for creating domain-specific datasets in construction; (2) underutilization of multimodal models for metadata generation and hierarchical organization; and (3) insufficient exploration of the broader potential of datasets, such as their reuse in deconstruction, material tracking, or progress monitoring to support sustainability goals. Addressing these gaps is essential to advancing the field and enabling meaningful applications in construction automation, precision engineering, and sustainable practices.

This study addresses these challenges by presenting an automated methodology that integrates web scraping, multimodal models, and data organization techniques to create a scalable and reusable dataset for structural steel components. By automating the entire process—from data collection to organization—this methodology establishes a robust foundation for training ML models tailored to construction applications. Furthermore, the structured and annotated dataset produced by this study directly supports circular economy principles by facilitating material reuse, reducing waste, and optimizing resource utilization. These contributions align with sustainability goals by minimizing the environmental impact of construction practices and promoting efficient workflows, offering valuable solutions for the industry's transition toward more sustainable operations[31].

2 Objectives

This study presents an automated methodology that integrates web scraping and Large Multimodal Models (LMMs) to generate standardized, structured datasets for construction. The methodology addresses critical challenges of dataset scalability, usability, and annotation quality, with a focus on consistent and accurate component type labeling. The key objectives include: (1) develop an automated web scraping pipeline to efficiently collect construction-related images from diverse online sources; (2) utilize multimodal LMMs to generate descriptive metadata for each image, enabling precise identification and classification of construction components; (3) implement a keyword rule-based categorization system to filter and organize images into a hierarchical folder structure that represents unique construction components, improving the usability and accessibility of the dataset; and (4) preparing a dataset with unique identifiers for each image to ensure consistency, standardization, and usability for downstream applications, such as training object detection models and supporting progress monitoring and sustainability-driven practices.

3 Methodology

This study employed a systematic, multi-phase methodology to acquire, process, and categorize data related to structural steel components. Each phase was carefully designed to ensure quality, reliability, and scalability while addressing the challenges posed by diverse and large-scale datasets. The workflow depicted in Figure 1 comprises four main steps: (1) data acquisition, (2) metadata generation, (3) dataset categorization, and (4) standardization.

3.1 Phase 1: Data Acquisition

The first phase aimed to automate the acquisition of structural steel images and metadata from online repositories and websites. This required developing a robust web scraping pipeline capable of navigating diverse and often unstructured web sources.

3.1.1 Web Scraping Implementation

The web scraping pipeline was implemented in Python, incorporating the following core libraries: *BeautifulSoup* for parsing *HTML* and *XML* documents, and *Requests* for sending *HTTP* requests and retrieving webpage content, and *Urllib* for *URL* parsing and management. Figure 2 illustrates the pseudocode for the web scraping pipeline, detailing its core components; URL sanitization, dynamic crawling, image downloading, and data organization. The pipeline included:

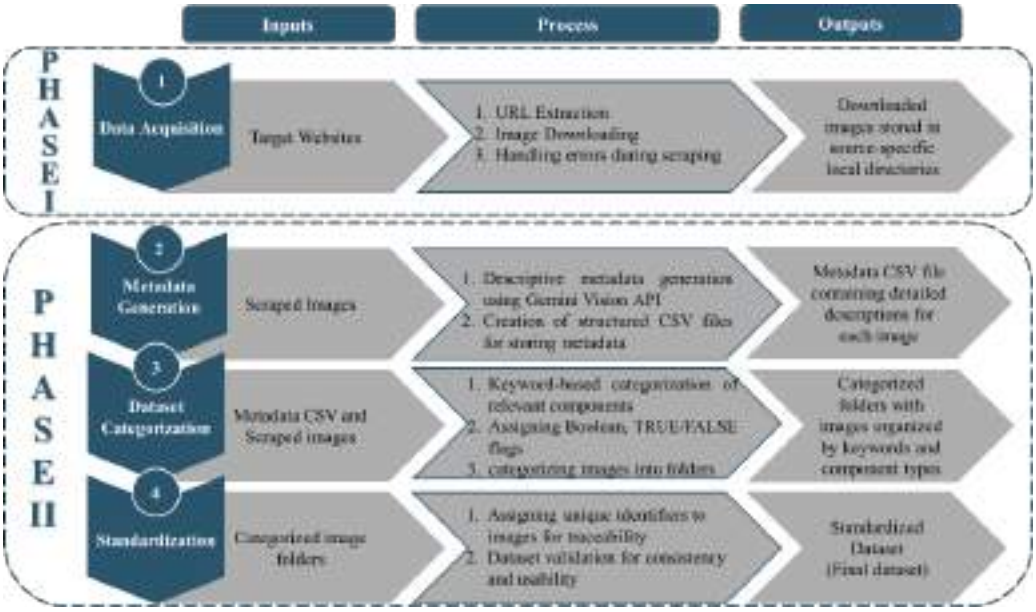


Figure 1. Workflow for Automated Dataset Generation Using Web Scraping and Large Multimodal Models

1. Sanitizing URLs and Filenames*:* A custom function, *sanitize_filename*, was designed to remove invalid characters from filenames and URLs, ensuring compatibility with file storage systems. This was achieved using Python's *re* module to apply regular expressions.
2. Dynamic Website Crawling*:* The *crawl_website* function recursively traversed web pages to retrieve images. *BeautifulSoup* was used to extract <*img*> tags, and a custom user-agent header mimicked a browser request, reducing IP blocking risks. A retry mechanism handled timeouts or rate-limit errors (HTTP 429) with randomized delays.
3. Downloading Images*:* image collection process relied on a targeted web scraping pipeline designed to extract domain-specific images related to structural steel. Relevant websites, including construction-focused repositories, engineering resources, and online catalogs featuring structural steel components (e.g., trusses, beams, columns, and frames), were identified and prioritized. Specific keywords such as "structural steel components," "steel trusses," "steel beams," and "steel frames" were used to guide the search and refine the scraping process. The *download_image* function saved images from extracted URLs into source-specific directories to facilitate later validation and traceability.
4. Handling Large-Scale Data**:** The pipeline could handle up to 100 pages per website and included delays between requests to avoid server flags. The *os* library was used to create a structured directory system for efficient data management.

```
IMPORT required libraries: requests, BeautifulSoup, os, re, random
  SET base_url and target folder for image downloads
  DEFINE User-Agent for headers

DEFINE FUNCTION sanitize_filename(filename):
    REMOVE invalid characters using regex
    RETURN cleaned filename

DEFINE FUNCTION download_image(url, folder):
    SEND GET request with headers
    IF successful:
      SAVE image to the target folder with sanitized filename

DEFINE FUNCTION scrape_images(page_content, folder):
    PARSE page content using BeautifulSoup
    EXTRACT image URLs and call download_image()

DEFINE FUNCTION crawl_website(base_url, folder):
    FETCH page content
    CALL scrape_images()
    RECURSE for linked pages

DEFINE MAIN FUNCTION():
    SET base_url and target folder
    CALL crawl_website(base_url, target folder)

CALL MAIN FUNCTION() to execute the web scraping process
```

Figure 2. Pseudocode illustrating the automated web scraping pipeline.

3.1.2 Data Volume and Storage

The scraping process collected 55,926 images, organized into source-specific folders. Each image's metadata (e.g., source URL and filename) was logged for validation, ensuring efficient management and traceability in subsequent phases. Figure 1 provides a high-level overview of this process.

3.2 Phase 2: Metadata Generation, Categorization, and Standardization

The second phase aimed to enhance the collected images with detailed metadata, categorize the data, and prepare a standardized dataset for downstream applications. This phase covered Steps 2–4 of the methodology.

3.2.1 Metadata Generation

Metadata for the dataset was generated using Gemini Vision, a state-of-the-art Large Multimodal Model (LMM) designed for analyzing visual content and generating textual descriptions. The metadata generation process was implemented in Python and executed on Google Colab, leveraging its computational capabilities for efficient large-scale processing. The Gemini Vision API was used for descriptive text generation, as illustrated in Figure 3.

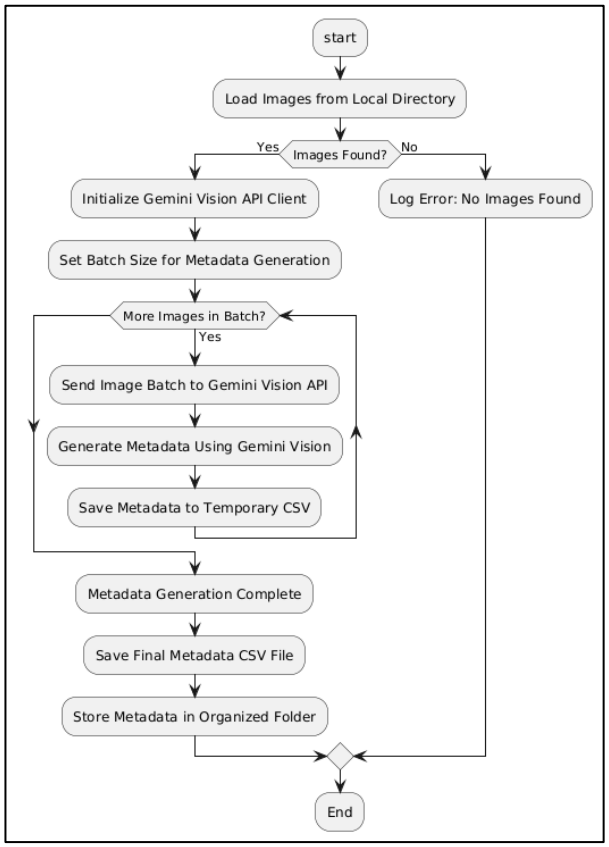


Figure 3 Workflow for metadata generation using Gemini Vision

1. Parallel Processing: To handle the large-scale dataset effectively, the *process_images_in_folder* function utilized Python's *ThreadPoolExecutor* for parallel processing. This approach distributed the workload across multiple system cores, significantly reducing processing time while ensuring scalability.
2. Error Handling: The system incorporated robust error-handling mechanisms. Issues such as corrupted image files or API timeouts were intercepted and logged with detailed error messages. Failed images were either skipped or re-queued for reprocessing, ensuring minimal data loss.
3. Metadata Validation: he metadata descriptions generated by Gemini Vision were stored in a structured Pandas DataFrame, paired with unique image IDs for traceability. To ensure accuracy, random samples (5% of the dataset) were manually reviewed, verifying their consistency with the actual visual content.

3.2.2 Metadata Example and Storage

Each metadata record included two essential fields; image ID: a unique identifier for each image, ensuring traceability, and description: a textual description generated by Gemini Vision, detailing the image's content. The final metadata was exported to a structured CSV file (*image_dataset_preparation.csv*) to facilitate seamless integration into downstream processes. This standardized file format ensured seamless integration into subsequent phases, enabling efficient accessibility and usability.

3.2.3 Rule-Based Categorization

A Python function, *extract_steel_structures*, scanned metadata descriptions for keywords indicating structural steel components (e.g., "beam," "column," "truss," "steel frame"). The categorization steps were (as illustrated in Figure 4):

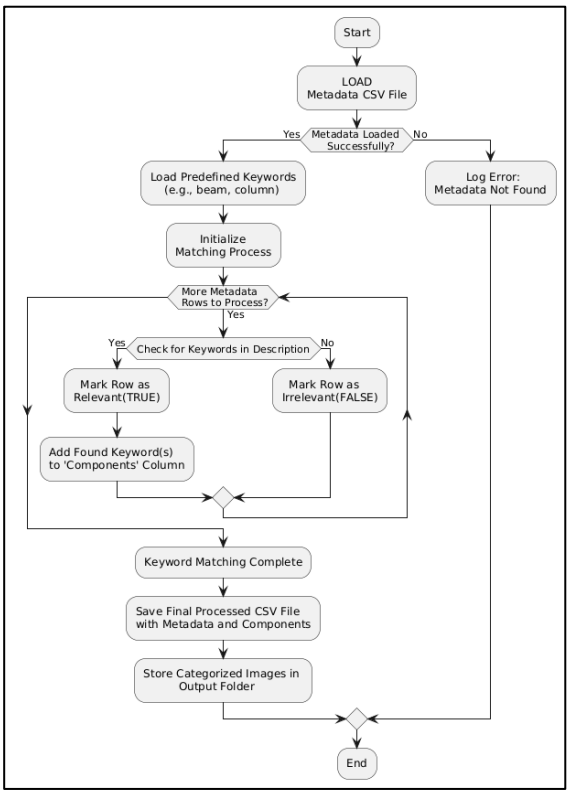


Figure 4 Keyword Matching Flowchart

1. Keyword Matching: Regular expressions were employed to identify variations in phrasing, such as "steel trusses" and "trusses of steel."
2. Field Assignment: Two fields were populated for each image: "*steel_structure"* a Boolean field *(TRUE/FALSE)* indicating the presence of structural steel components and "*components"* which is a list of matched keywords representing specific structural elements.

3.2.4 Validation and Refinement

To ensure the dataset's accuracy and reliability, a validation and refinement process was conducted. A diverse 5% sample of the dataset was manually inspected for categorization accuracy. Discrepancies, such as false positives and ambiguous descriptions, were analyzed to refine keyword matching rules. This iterative process improved classification precision and enhanced the overall quality of the dataset.

3.2.5 Final Dataset

The final dataset included 2,691 images, each labeled as TRUE for containing structural steel components. These images were categorized into the following components: beams: 1,716 images; columns: 554 images; steel frames: 169 images; trusses: 116 images; and steel structures: 506 images.

4 Results and Discussions

The proposed methodology successfully generated a comprehensive dataset of 2,691 structural steel images from an initial pool of 55,926 images retrieved from various online sources. This process demonstrated the robustness and effectiveness of the methodology in automating large-scale data acquisition, metadata generation, and dataset standardization, specifically tailored for construction applications. Each stage of the workflow addressed critical challenges, including data scarcity, inconsistent metadata, and the need for scalable solutions to support machine learning applications in the construction sector.

The web scraping process retrieved 55,926 images from diverse online sources using a pipeline optimized for dynamic website structures and large-scale data handling. Following keyword-based filtering and metadata validation, 53,235 images were excluded due to irrelevance or insufficient metadata descriptions, resulting in 2,691 images deemed relevant to structural steel components. These components were categorized into key structural elements: beams (63.8%), columns (20.6%), steel structures (18.8%), steel frames (6.3%), and trusses (4.3%). This categorization, summarized in Table 1, reflects the prevalence of these components in real-world structural steel usage, ensuring the dataset's practical applicability for construction applications.

Table 1. Categorization of Structural Steel Components in the dataset

Component type	Count	Percentage of Dataset
Beams	1,716	63.8%
Columns	554	20.6%
Steel Frames	168	6.3%.
Trusses	116	4.3%
Steel Structures	506	18.8%
Total number of images in the dataset		2,691

The metadata generation phase utilized Gemini Vision's Large Multimodal Model (LMM) to produce detailed descriptions for each image. A sample of 5% of the dataset, approximately 135 images, was manually inspected to verify metadata accuracy and categorization reliability. Observed discrepancies, including 7 false positives caused by ambiguous descriptions, informed iterative refinements of keyword matching rules and metadata prompts. These adjustments ensured alignment between generated metadata and visual content, resulting in a high accuracy rate of 94.8%. Approximately 5.2% of the metadata required manual correction due to ambiguity or incomplete details. Examples of metadata descriptions and their associated categorization are presented in Figure 5.

The metadata generation process, as summarized in Table 2, was completed in 11 hours and 42 minutes. This time refers exclusively to the metadata generation process, including image annotation and description. The workflow was implemented using Python on Google Colab, leveraging Python's *ThreadPoolExecutor* for parallel processing to efficiently handle the large-scale dataset. The computing environment included a Google Colab Pro+ subscription, providing access to an NVIDIA Tesla P100 GPU and 16 GB of RAM, which ensured sufficient computational power for metadata generation. This setup facilitated concurrent processing of multiple images, significantly reducing runtime. Dynamic crawling and retry mechanisms effectively mitigated errors caused by website structures and rate limits. This dataset provides valuable insights into structural steel components, including the unique geometries of trusses and steel frames frequently used for load distribution in roofing systems and other applications. It supports downstream machine learning applications, such as object detection models for automated progress monitoring on construction sites. Additionally, the categorized dataset holds potential for circular economy practices, enabling the identification and reuse of structural steel components during deconstruction and other sustainability-focused activities.

1	Image ID	Description	steel_structure	components
125	ecf40e32-07a5-4f44-ae42-599650b07b7e	The image shows a steel structure of a building under construction.	TRUE	steel structure
126	f935743a-4731-4c00-b64a-c19de3f474c3	The image shows a metal framework of a building under construction. There are many steel beams and girders, and the ceiling is made of corrugated metal.	TRUE	beam
127	d8e96038-154c-4474-9549-de60d88b449e	The image shows a steel structure of a building under construction. The structure consists of columns, beams, and trusses. The columns are vertical members that support the beams. The beams are horizontal members that span between the columns. The trusses are triangular members that support the roof.	TRUE	beam, column, truss
128	8e37c727-c9c4-4a5b-b422-233621258bb6	A steel frame of a container, likely used for shipping or storage.	FALSE	
129	670de234-a2c0-41dd-9a05-3374b1984fb4	The image shows a steel structure of a building under construction. There are a lot of steel beams and columns. The beams are connected to the columns with bolts and nuts. The columns are supported by the foundation. The roof is made of metal sheets.	TRUE	beam, column
130	798a0f3a-5469-40f9-b33d-11db971a0f60	The image shows a metal frame that could be used as the base for a table, bench, or other piece of furniture.	FALSE	

Figure 5 Metadata and Categorization of Structural Steel Images

Table 2 Key performance Metrics of dataset Preparation

Metric	Value	Description
Total Images Collected	55,926	Number of images retrieved from diverse online sources.
Filtered Images	2,691	Number of images retained after keyword-based filtering.
Excluded Images	53,235	Images excluded due to irrelevance or insufficient metadata.
Metadata Accuracy (For 5% of data)	94.8%	Percentage of descriptions accurately matching image content.
Processing Speed	1.33 images/ second	Average speed achieved during scraping and metadata generation.
Time Taken	11 hours, 42 minutes	Total time taken for web scraping and dataset preparation.

The methodology demonstrated significant scalability and flexibility, overcoming challenges such as dynamic webpage structures, redundant data, and linguistic inconsistencies in metadata descriptions. Iterative refinements to scraping algorithms, metadata prompts, and keyword-based categorization rules ensured a high-quality dataset, ready for diverse construction applications. These results establish a robust framework for generating high-quality datasets and pave the way for future advancements, such as integrating real-time data acquisition and expanding to other construction materials like concrete and timber. This scalable approach sets a benchmark for automated dataset generation, enhancing the adoption of AI in construction practices.

5 Summary and Concluding Remarks

This study demonstrates how web scraping and Large Multimodal Models (LMMs) can effectively create a scalable, structured dataset tailored for construction applications. By automating data collection, metadata generation, and dataset organization, the methodology overcomes inefficiencies in manual data gathering, limited dataset availability, and insufficient annotation quality. Using Gemini Vision, descriptive metadata for 2,691 structural steel images were generated with 94.8% accuracy, selected from an initial pool of over 55,000 images. These images were categorized into beams, columns, trusses, steel frames, and steel structures, enabling streamlined tasks like object detection for progress monitoring and construction automation.

Key contributions of this study include: (1) developing an automated web scraping pipeline to

efficiently collect construction-related images from diverse online sources; (2) utilizing Gemini Vision for high-quality metadata generation to enable precise identification and classification of construction components; (3) implementing a keyword-based categorization system to organize and filter the dataset, using a rule-based approach with Boolean labels and hierarchical folder structures; (4) delivering a reusable dataset with standardized records and unique image identifiers to support downstream applications like progress monitoring and circular economy initiatives; and (5) establishing a scalable framework adaptable to construction materials beyond structural steel, including concrete, timber, and composites, to address broader dataset challenges.

While this study focuses on structural steel, the methodology is adaptable to other construction materials, offering potential for broader applications. However, the selection of websites might favor commonly available components, and keyword choices could introduce bias, influencing dataset composition. To minimize this bias, careful source selection and validation were implemented, along with a comprehensive keyword strategy designed to capture diverse naming conventions, ensuring balanced representation.

To ensure the robustness of the dataset. Future work should address discrepancies, such as false positives and overlooked false negatives, and explore cost-effective LMMs and open-source alternatives for wider adoption. Additional enhancements, such as real-time data acquisition from active construction sites and expanding the dataset to include diverse materials, would improve dataset usability and diversity. These developments would further AI-driven advancements in construction automation and sustainability initiatives. Overall, this study establishes a robust methodology for automated data generation and AI applications in construction. By addressing scalability, usability, and annotation challenges, it accelerates digital transformation, promotes sustainable practices, and lays the groundwork for future innovations in the built environment.

References

[1] M.A. Musarat, N. Hameed, M. Altaf, W.S. Alaloul, M. Al Salaheen, A.M. Alawag, Digital Transformation of the Construction Industry: A Review, in: 2021 International Conference on Decision Aid Sciences and Application (DASA), IEEE, 2021: pp. 897–902. https://doi.org/10.1109/DASA53625.2021.9682303.

[2] W.S. Alaloul, K.M. Alzubi, A.B. Malkawi, M. Al Salaheen, M.A. Musarat, Productivity monitoring in building construction projects: a systematic review, Engineering, Construction and Architectural Management 29 (2022) 2760–2785. https://doi.org/10.1108/ECAM-03-2021-0211.

[3] S. Gharib, O. Moselhi, A review of Computer Vision-Based Techniques for Construction Progress Monitoring, in: 2023. https://doi.org/10.22260/ISARC2023/0071.

[4] N. Zhang, J. Ye, Y. Zhong, Z. Chen, Digital Transformation in the Chinese Construction Industry: Status, Barriers, and Impact, Buildings 13 (2023) 1092. https://doi.org/10.3390/buildings13041092.

[5] C. Toșa, C.P. Paneru, A. Joudavi, A.K.M. Tarigan, Digital transformation, incentives, and pro-environmental behaviour: Assessing the uptake of sustainability in companies' transition towards circular economy, Sustain Prod Consum 47 (2024) 632–643. https://doi.org/10.1016/j.spc.2024.04.032.

[6] O. Samuelson, L. Stehn, DIGITAL TRANSFORMATION IN CONSTRUCTION – A REVIEW, Journal of Information Technology in Construction 28 (2023) 385–404. https://doi.org/10.36680/j.itcon.2023.020.

[7] H.R. Seeberg, S.M. Haakonsen, M. Luczkowski, Systematic Mapping of Circular Economy in Structural Engineering, Buildings 14 (2024) 1165. https://doi.org/10.3390/buildings14041165.

[8] P. Hradil, L. Fülöp, M. Wahlström, C. Del Castillo, The new construction products regulation: Opportunity or barrier for reused constructional steel?, in: Life-Cycle of Structures and Infrastructure Systems - Proceedings of the 8th International Symposium on Life-Cycle Civil Engineering, IALCCE 2023, 2023. https://doi.org/10.1201/9781003323020-90.

[9] D.D. Tingley, J. Allwood, Reuse of structural steel: The opportunities and challenges, European Steel Environment & Energy Congress (2014).

[10] D. Ness, J. Swift, D.C. Ranasinghe, K. Xing, V. Soebarto, Smart steel: New paradigms for the reuse of steel enabled by digital tracking and modelling, J Clean Prod 98 (2015) 292–303. https://doi.org/10.1016/j.jclepro.2014.08.055.

[11] S. Bunrit, N. Kerdprasop, K. Kerdprasop, Evaluating on the Transfer Learning of CNN Architectures to a Construction Material Image Classification Task, Int J Mach Learn Comput 9 (2019) 201–207. https://doi.org/10.18178/ijmlc.2019.9.2.787.

[12] M. Soleymani, M. Bonyani, H. Mahami, F. Nasirzadeh, Construction material classification

on imbalanced datasets using Vision Transformer architecture (ViT), (2022). https://doi.org/10.21203/rs.3.rs-1948162/v1.

[13] A. Pal, J.J. Lin, S.H. Hsieh, M. Golparvar-Fard, Automated vision-based construction progress monitoring in built environment through digital twin, Developments in the Built Environment 16 (2023). https://doi.org/10.1016/j.dibe.2023.100247.

[14] S. Kitayama, O. Iuorio, Disassembly and Reuse of Structural Members in Steel-Framed Buildings: State-of-the-Art Review of Connection Systems and Future Research Trends, Journal of Architectural Engineering 29 (2023). https://doi.org/10.1061/jaeied.aeeng-1615.

[15] Z. Liu, A Review of Research on Digital Transformation in Construction Enterprises, Academic Journal of Science and Technology 9 (2024) 300–305. https://doi.org/10.54097/00gkv045.

[16] Y. Zhang, R. Chang, W. Mao, J. Zuo, L. Liu, Y. Han, Challenges of Automating Interior Construction Progress Monitoring, J Constr Eng Manag 150 (2024). https://doi.org/10.1061/jcemd4.coeng-14637.

[17] V.K. Reja, K. Varghese, Q.P. Ha, Computer vision-based construction progress monitoring, Autom Constr 138 (2022). https://doi.org/10.1016/j.autcon.2022.104245.

[18] S. Upadhyay, V. Pant, S. Bhasin, M.K. Pattanshetti, Articulating the construction of a web scraper for massive data extraction, in: 2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT), IEEE, 2017: pp. 1–4. https://doi.org/10.1109/ICECCT.2017.8117827.

[19] S. Ganesh, A.P. Celestina, R. Jayashree, K.V. HariPriya, Web Automation in Health Care, in: 2019 IEEE International Conference on Innovations in Communication, Computing and Instrumentation (ICCI), IEEE, 2019: pp. 28–35. https://doi.org/10.1109/ICCI46240.2019.9404388.

[20] S. Rennie, M. Buchbinder, E. Juengst, L. Brinkley-Rubinstein, C. Blue, D.L. Rosen, Scraping the Web for Public Health Gains: Ethical Considerations from a 'Big Data' Research Project on HIV and Incarceration, Public Health Ethics 13 (2020) 111–121. https://doi.org/10.1093/phe/phaa006.

[21] S.L. Zulu, A.M. Saad, S.O. Ajayi, M. Dulaimi, M. Unuigbe, Digital leadership enactment in the construction industry: barriers undermining effective transformation, Engineering, Construction and Architectural Management 31 (2024) 4062–4078. https://doi.org/10.1108/ECAM-05-2022-0491.

[22] L. Jacques de Sousa, J. Poças Martins, L. Sanhudo, Tackling the Data Sourcing Problem in Construction Procurement Using File-Scraping Algorithms, in: The 1st International Online Conference on Buildings, MDPI, Basel Switzerland, 2023: p. 34. https://doi.org/10.3390/IOCBD2023-15190.

[23] V. Koscheyev, V. Rapgof, V. Vinogradova, Digital transformation of construction organizations, IOP Conf Ser Mater Sci Eng 497 (2019) 012010. https://doi.org/10.1088/1757-899X/497/1/012010.

[24] R.K. Sichkariuk, N. V. Kornilovska, I.A. Lurie, M.A. Voronenko, Development and implementation of effective Web Scraping methods for automated data collection and processing using Python, Informatics. Culture. Technology 1 (2024) 62–68. https://doi.org/10.15276/ict.01.2024.08.

[25] A. Omar, O. Moselhi, An Integrated Approach for Automated Acquisition of Bridge Data and Deficiency Evaluation, in: 2023. https://doi.org/10.22260/ISARC2023/0047.

[26] S. Agrawal, J. Trenkle, J. Kawale, Beyond Labels: Leveraging Deep Learning and LLMs for Content Metadata, (2023). https://doi.org/10.1145/3604915.3608883.

[27] B. Dikmans, D. Kang, A Brief Survey into the Field of Automatic Image Dataset Generation through Web Scraping and Query Expansion, Journal of Information Processing Systems 19 (2023) 602–613. https://doi.org/10.3745/JIPS.04.0288.

[28] E. Uzun, Scraping Relevant Images from Web Pages without Download, ACM Transactions on the Web 18 (2024) 1–27. https://doi.org/10.1145/3616849.

[29] E. Uzun, A Novel Web Scraping Approach Using the Additional Information Obtained From Web Pages, IEEE Access 8 (2020) 61726–61740. https://doi.org/10.1109/ACCESS.2020.2984503.

[30] A. Karpathy, L. Fei-Fei, Deep Visual-Semantic Alignments for Generating Image Descriptions, IEEE Trans Pattern Anal Mach Intell 39 (2017) 664–676. https://doi.org/10.1109/TPAMI.2016.2598339.

[31] V. Ilin, I. Kalinov, P. Karpyshev, D. Tsetserukou, DeepScanner: a Robotic System for Automated 2D Object Dataset Collection with Annotations, (2021). http://arxiv.org/abs/2108.02555.

Next-Generation Stretchable EDA Sensors with Ag-NP Fractals for Motion Artifact-Free Physiological Sensing in Dynamic Environments

Yogesh Gautam,[1] and Houtan Jebelli, Ph.D.[2,*]

[1] Department of Civil and Environmental Engineering, University of Illinois Urbana-Champaign, Champaign, IL.
[2]Department of Civil and Environmental Engineering, University of Illinois Urbana-Champaign, Champaign, IL. (corresponding author)

ygautam2@illinois.edu , hjebelli@illinois.edu

Abstract

Construction workers frequently encounter substantial mental stress due to the high demands of their roles, including tight deadlines and challenging environmental conditions. Although electrodermal activity (EDA) sensors offer a promising solution for monitoring stress levels, traditional designs are often inflexible, leading to issues such as poor skin contact, elevated contact impedance, and motion artifacts that compromise the quality of the collected signals. To overcome these challenges, this paper introduces a stretchable EDA sensor featuring a Peano-Gosper fractal design aimed at enhancing adaptability and minimizing motion-related interference. Through a finite element analysis, several design variations were assessed, with the fractal pattern demonstrating superior skin contact and reduced mechanical strain, making it highly suitable for real-world applications. Experimental results indicated significant advancements in signal-to-noise ratio (SNR), with the fractal-designed sensor achieving nearly 8 dB improvement compared to conventional EDA sensors during experimental construction activities. This development enhances the reliability of EDA measurements and corresponding mental stress monitoring, ultimately contributing to improved health and safety outcomes for construction workers.

Keywords –

EMG ; Flexible Sensor; EDA; Motion Artifact; Physiological Monitoring

1 Introduction

The construction industry has long grappled with the pervasive issue of mental stress among its workforce, a problem that has far-reaching consequences for both individual health and overall project outcomes [1]. Recent studies have indicated that construction workers experience some of the highest rates of stress-related mental health issues across all industries, with an estimated 35% reporting symptoms of anxiety or depression. This chronic exposure to occupational stress not only impacts worker well-being but also significantly increases the risk of serious health complications, including cardiovascular diseases and musculoskeletal disorders [2]. Moreover, stress-induced cognitive impairments can lead to decreased situational awareness and decision-making abilities, potentially compromising worksite safety and productivity [3,4].

Electrodermal activity (EDA) has emerged as a promising biomarker for quantifying sympathetic nervous system activation, which is closely linked to stress responses [5]. EDA measurements, which capture changes in skin conductance resulting from sweat gland activity, have shown a robust correlation ($r = 0.78$) with self-reported stress levels in controlled laboratory settings. In the context of construction, preliminary field studies utilizing commercial EDA sensors have demonstrated the potential for real-time stress monitoring [6]. However, the dynamic and physically demanding nature of construction environments presents unique challenges that often result in degraded sensor performance, with reduced SNR compared to stationary conditions.

The widespread adoption of wearable health monitoring technologies in the construction sector has been hindered by limitations in current sensor designs [7]. While significant advancements have been made in data processing algorithms and machine learning models for stress detection, the underlying hardware often fails to meet the specific demands of construction tasks [8]. Commercial EDA sensors, typically designed for general-purpose physiological monitoring, employ rigid materials and standardized form factors that are ill-suited for the diverse and strenuous activities inherent to construction work. These design shortcomings manifest in several critical issues, including increased contact

impedance due to poor skin conformity, motion artifacts from sensor displacement, and user discomfort during prolonged wear [9].

To address these challenges, this paper proposes a novel solution: a fractal-based, patch-type flexible EDA sensor specifically engineered for use in construction environments. The innovative fractal design offers enhanced skin conformity, potentially reducing contact impedance and increased SNR. This approach not only minimizes motion artifacts but also improves overall signal quality and user comfort. By leveraging advanced materials and optimized geometries, the proposed sensor aims to maintain consistent performance across a wide range of construction activities, from repetitive tasks to sudden, high-intensity movements. This tailored solution has the potential to revolutionize real-time stress monitoring in construction workers, providing more accurate and reliable data for both immediate intervention and long-term occupational health strategies.

The paper is structured as follows: In Section 2, a brief literature review is provided, focusing on flexible EDA sensors and health monitoring applications of EDA sensors. Section 3 describes the proposed methodology, including the technical details of the design and fabrication techniques. In Section 4, a case study is presented, which includes details of the experiments conducted to develop and evaluate the developed sensor. The results are analyzed and discussed in Sections 5 and 6. Finally, Section 7 concludes the paper.

2 Background and Related Work

2.1 Stress, the Sympathetic Nervous System, and EDA

Stress triggers a complex cascade of physiological responses, primarily mediated by the autonomic nervous system (ANS). The ANS comprises two main branches: the sympathetic nervous system (SNS), which initiates the "fight-or-flight" response, and the parasympathetic nervous system, responsible for "rest-and-digest" functions [10]. During stress, the SNS becomes dominant, leading to a range of physiological changes aimed at mobilizing energy resources and preparing the body for action [11].

The skin, particularly its sweat glands and vasculature, is exclusively innervated by the SNS, making it an ideal target for assessing sympathetic activation. Electrodermal activity (EDA), which measures changes in skin electrical properties due to sweat gland activity, provides a direct and relatively unambiguous index of SNS arousal [12]. This contrasts with other physiological measures, such as heart rate variability, which reflect the complex interplay between both SNS and PNS influences.

EDA is typically quantified through skin conductance measurements, which are proportional to sweat secretion. The palmar surfaces of hands and feet, with their high density of eccrine sweat glands , are preferred measurement sites. The EDA signal comprises two main components: tonic skin conductance level (SCL) and phasic skin conductance responses (SCRs). SCL represents the slowly changing baseline level of electrical conductivity, reflecting overall psychophysiological activation [13]. SCRs are rapid, transient increases in conductance, occurring either in response to specific stimuli (specific SCRs) or spontaneously (non-specific SCRs or NS.SCRs).

Specific SCRs typically manifest 1 to 6 seconds post-stimulus and are characterized by several key parameters: amplitude (magnitude of conductance change), latency (time between stimulus onset and SCR initiation), rise time (duration from SCR onset to peak), and half-recovery time (time for amplitude to decrease by 50% from peak). NS.SCRs, while morphologically similar to specific SCRs, occur in the absence of identifiable external stimuli. The frequency and mean amplitude of NS.SCRs serve as additional indicators of psychophysiological arousal [14].

Importantly, both SCL and SCR patterns exhibit significant inter-individual variability, necessitating careful baseline establishment and within-subject comparisons in experimental designs. Numerous studies have demonstrated the superior sensitivity of EDA measures, particularly SCL and NS.SCR frequency, in tracking stress responses compared to other physiological indicators. This sensitivity, combined with the relatively straightforward interpretation of EDA in terms of SNS activity, makes it a valuable tool for real-time stress monitoring in dynamic environments such as construction sites.

2.2 Flexible EDA sensors

Flexible electrodermal activity (EDA) sensors represent a significant advancement in wearable biomonitoring technology, offering potential solutions to the longstanding challenges associated with traditional rigid sensor designs [15]. The development of these sensors is driven by the need for more reliable and comfortable continuous monitoring systems, particularly in dynamic environments such as construction sites [9,16,17].

One of the primary advantages of flexible EDA sensors is their ability to maintain consistent skin contact during movement, potentially reducing motion artifacts that plague inflexible commercial sensors. These artifacts, which can significantly distort physiological signals, are particularly problematic in occupational settings where workers engage in varied physical activities. While algorithmic approaches such as adaptive filtering and

machine learning techniques have shown promise in mitigating motion artifacts, they are inherently limited by the quality of the raw signal input [7,8]. Flexible sensors address this issue at the hardware level, potentially improving signal fidelity before any software processing occurs.

Recent advancements in materials science have enabled the development of novel flexible sensor substrates and electrode materials. Conductive polymers, carbon nanotubes, and metallic nanowires have been successfully incorporated into stretchable matrices, creating sensors capable of withstanding substantial deformation while maintaining electrical performance [18]. These materials often exhibit unique electromechanical properties, such as increased conductivity under strain, which can be leveraged to enhance sensor sensitivity and resilience. The design of flexible EDA sensors also allows for innovative electrode patterns and geometries that were previously unfeasible with rigid substrates. Fractal-inspired designs, for instance, have shown promise in optimizing the tradeoff between mechanical flexibility and electrical performance [19]. These intricate patterns can potentially increase the effective surface area for EDA measurement while allowing the sensor to conform to the complex topography of human skin, particularly in areas with high curvature or frequent movement.

3 Materials and Methodology

The overview of the proposed methodology is presented in Figure 1, In the first step, we 3D model 8 different variations of fractal curves with varying order and shapes to fit within a square patch of length 5 mm. Here we model Hilbert, Moore, z-space and Peano-Gosper curves. In the second stage we perform static analysis subjecting the 3D models to varying tensile forces and evaluate stress/strain characteristics. We calculate von-mises stress, strain values, and geometric characteristics such as overage area. Based on the results, we choose Peano-Gosper order 4 curve for fabrication. During fabrication, we use Silver Nanoparticle modified ink to print in into a flexible polyimide surface using a piezoelectric printing mechanism. After printing the Ag-NP surface is UV-treated using sintering. Then, the developed electrode is evaluated and compared with the commercially available electrodes for noise reduction.

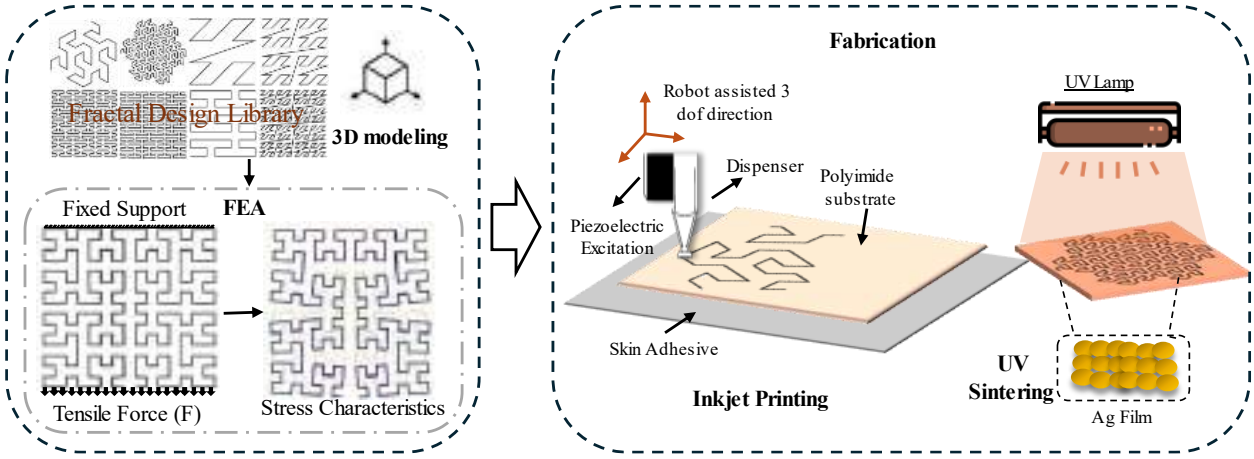


Figure 1: Overview of the methodology. Multiple Fractal designs are 3D modeled and then analyzed for mechanical properties using finite element analysis. The best one is selected based on a mixture of von Mises stress, strain, and geometric properties like coverage to surface area ratio. The selected design is printed using AgNP ink into a flexible polyimide substrate followed by polyimide

3.1 Design and FEM evaluation of Flexible Electrodes

Eight different fractal curve variations were modeled in 3D, including Hilbert, Moore, Z-space, and Peano curves of varying orders, each confined within a 5 mm square patch. The Peano-Gosper curve, originally termed the flow-snake, represents a sophisticated space-filling pattern that exhibits unique hexagonal characteristics. When implemented in second and third orders, this curve generates intricate self-replicating structures that optimize both spatial utilization and mechanical adaptability.

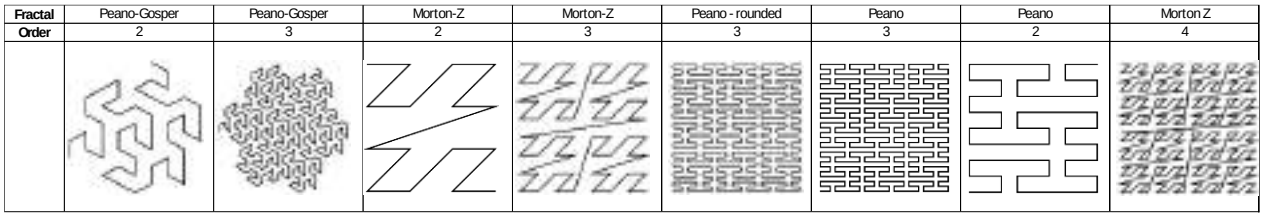


Figure 2: Fractal curves with respective order analysed for this study.

The Morton Z-curve, alternatively known as the Z-order curve, employs an innovative approach to dimensional mapping, effectively transforming two-dimensional spatial relationships into linear sequences while maintaining proximity correlations. Implemented across orders 2, 3, and 4, this pattern creates a hierarchical arrangement that achieves an optimal equilibrium between connectivity and coverage. The Z-curve's architecture proves particularly advantageous in applications requiring efficient spatial data organization and multidimensional query optimization.

Among these patterns, the Peano curve demonstrates the mathematical possibility of continuous coverage across a complete unit square. Its modified variant, the Peano-rounded design, incorporates strategic curvature elements to enhance mechanical performance. These modifications serve multiple purposes: they minimize stress concentration at junction points, preserve continuous surface coverage, and significantly improve the structure's overall flexibility and stretchability characteristics.

These models were subjected to static analysis simulating tensile forces ranging from 0.5 N to 2 N, representing the spectrum of mechanical stresses encountered during typical construction activities. The FEA simulation utilized a non-linear large deformation model to account for the elastic behavior of the flexible substrate. The material properties of the silver nanoparticle (AgNP) conductive ink and the polyimide substrate were incorporated into the model, with particular attention to the interface between these materials.

$$\sigma_v = \sqrt{\frac{1}{2}\left[\sum_{x,y,z}\left(\sigma_x - \sigma_y\right)^2\right] + 3\left(\tau_{xy}^2 + \tau_{yz}^2 + \tau_{zx}^2\right)} \quad (1)$$

$$Ku = F, \qquad \in = \frac{\Delta l}{l_0} \quad (2)$$

where, $\sigma_x, \sigma_y, and\ \sigma_z$ are the normal stresses along the x,y, and z axes, respectively and $\tau_{xy}, \tau_{yz}, and\ \tau_{zx}$ are the shear stresses in the respective planes.

3.2 Sensor Fabrication and Sintering

The sensor electrodes were designed to fit comfortably on human skin, optimized for flexibility and durability in field environments. Polyimide, selected as the substrate, provides a balance of lightweight structure and flexibility similar to the tensile qualities of human skin, allowing the sensor to conform to body contours with minimal discomfort. This Kapton HN polyimide film offers high elongation capacity, low thermal expansion, and favourable tensile strength, making it an ideal foundation for wearable sensors that experience frequent motion. The selected thickness (0.125 mm) of this film supports the sensor's resilience while maintaining a minimal profile for improved adhesion and reduced interference during movement.

The electrodes were created using Ag nano-ink, chosen for its excellent conductivity and mechanical robustness, essential for clear and stable signal transmission. This ink, consisting of silver nanoparticles (AgNPs), combines a high-weight percentage of Ag with properties suited for high-resolution printing. The small particle size allows smooth deposition without clogging, while the moderate viscosity ensures steady flow through the printer's micro-dispensing system. These factors together enhance the ink's adherence to the polyimide surface, resulting in conductive pathways with low resistance after sintering. This choice of material ensures the electrodes can perform reliably under various physical strains while keeping the overall structure lightweight and thin $V(t) = V_0\ sin(\omega t) = V_0\ sin(2\pi f t)$. The printing process employed a SonoPlot Microplotter proto, utilizing an ultrasonic fluid injection mechanism to achieve precise electrode patterning. This technique relies on a piezoelectric-controlled glass dispenser, where AC-induced vibrations allow controlled dispensing of the Ag nano-ink onto the substrate. By adjusting the frequency of the AC signal, the fluid's flow rate is precisely managed, creating a serpentine pattern optimized for flexibility and durability. Following the printing phase, UV sintering with a pulsed light system (X-1100:XENON) was applied to solidify the AgNPs without overheating the polyimide film. This sintering process removed solvents and bound the particles into a stable, conductive layer, completing the sensor's

formation with minimal thermal impact on the flexible substrate.

4 Experimental Case Study

To study the artifact reduction in the developed EDA sensor roofing construction task was used (Figure 2). The EDA signals were recorded between the index and ring finger. In one scenario, the subjects were asked to stay static without moving their fingers to record the base-line signals and analyze the noise properties in the static position. To induce motion, a hammering-based construction task was used. During the hammering task, the subjects hammered shingles into a prebuilt roof structure. This task was performed for both the commercial and developed electrodes. The detailed results are presented in Section 5.

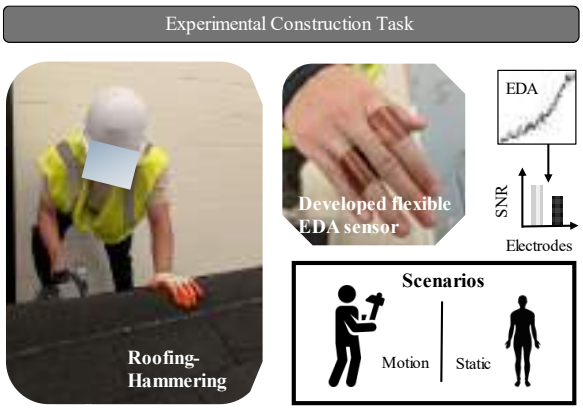


Figure 3: Experimental setup comprises two scenarios with and without motion. For inducing motion hammering task is used where the subject hammers shingles into a prebuilt roof structure. The signals are acquired for developed and commercial electrodes.

The signals acquired from the EDA sensors were amplified using bio-amplifier from AD-Instruments. Signals were acquired at the range of 100 Hz.

5 Results and Discussion

The comparative analysis of the various fractal-based electrode designs revealed significant performance variations across mechanical properties and signal quality metrics. The results demonstrate the superior characteristics of higher-order fractal patterns, particularly in terms of stress distribution and motion artifact reduction.

5.1 Mechanical Performance Analysis

The finite element analysis revealed distinctive mechanical behaviors across the various fractal patterns tested. The Peano-Gosper curve of order 2 initially showed the highest von Mises stress at 5.5E+10 N/m², while its order 3 counterpart demonstrated a dramatic improvement, with stress levels dropping to 3.9E+08 N/m². This significant reduction in stress indicates that higher-order Peano-Gosper patterns effectively distribute mechanical loads throughout their structure.

The Morton Z-curve patterns exhibited consistent mechanical performance across different orders, with stress levels stabilizing around 2.0E+10 N/m². The order 3 and 4 variations maintained relatively uniform stress distribution, suggesting reliable mechanical stability. These patterns demonstrated better structural integrity while accommodating deformation, with strain values ranging between 0.1 and 0.2.

The Peano and Peano-rounded variations demonstrated notably different characteristics. The Peano-rounded design of order 3 achieved moderate stress levels of approximately 1.5E+10 N/m² and maintained relatively low strain values of around 0.15. The rounded corners incorporated into this design proved effective in mitigating stress concentration points, a common challenge in angular fractal patterns. The standard Peano curve showed similar stress characteristics but with slightly higher strain values, indicating that the rounded modifications successfully improved the overall mechanical performance.

Strain analysis across all patterns revealed a consistent trend where higher-order patterns generally exhibited better strain management. The Peano-Gosper order 2 design showed the highest strain at 0.55, while its order 3 variant demonstrated remarkably lower strain values (0.004). This pattern was consistent across all designs, with higher-order variations consistently outperforming their lower-order counterparts in strain management. The Morton Z-curve patterns, particularly in orders 3 and 4, maintained moderate strain levels, suggesting good deformation characteristics under applied loads.

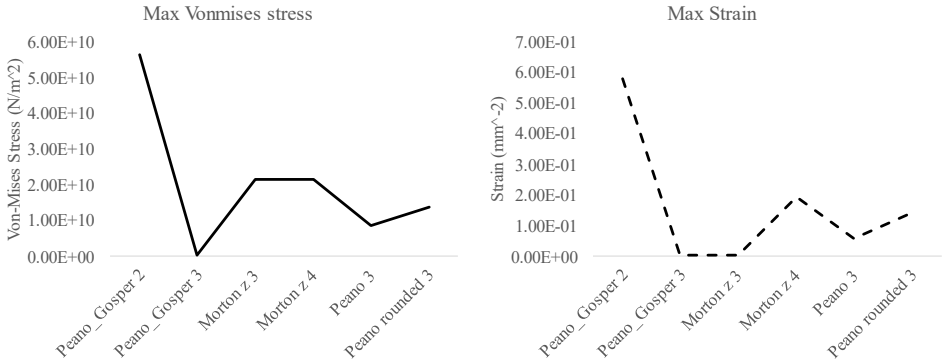


Figure 4: Max Von-mises stress and maximum strain for the given loading conditions.

The comprehensive analysis of both stress and strain characteristics indicates that while higher-order patterns generally offer superior mechanical properties, the optimal design must balance performance with manufacturing feasibility. The Peano-gosper order 3 pattern emerged as a particularly promising candidate, offering an excellent compromise between stress distribution, strain management, and practical implementation considerations. Based on this comparison, Peano-Gosper 3 was selected for the fabrication and further analysis.

5.2 Motion Artifact Reduction

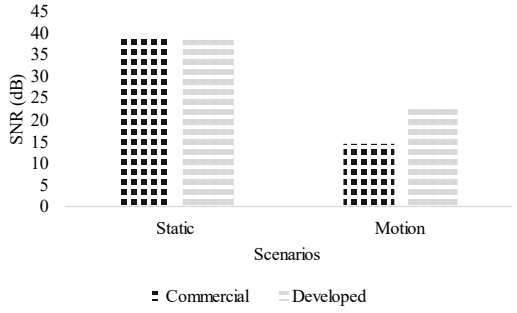


Figure 5: The developed electrodes provide significant improvement in SNR for the hammering task which induces significant motion.

To compare the performance of the developed electrode with the existing commercial modules we used bipolar finger electrodes made from brightly polished stainless steel plates (ADInstruments). As shown in Figure 5, the designed serpentine electrode exhibits a significantly higher SNR of approximately 23 dB compared to the commercial electrode, which has an SNR of around 15 dB. This difference in SNR performance can be attributed to the optimized electrode design and manufacturing process employed for the serpentine structure. We also compared the SNR both during static position (rest) and dynamic position (hammering with the sensing hand) to assess the difference during dynamic activities. While both electrodes provide similar performance during rest ($|SNR| < 1db$) the performance during dynamic activity (hammering) the SNR improvement is significant (8 db). The enhanced SNR indicates that the designed electrode is capable of providing cleaner, more reliable EDA signals, which is a critical requirement for accurate physiological monitoring in wearable health applications.

6 Conclusions

This research develops and analyses flexible sensor technology through the development of fractal-based electrodes for EDA monitoring applications. The comprehensive mechanical analysis of various fractal patterns has yielded significant insights into their performance characteristics and potential for wearable sensing applications. The finite element analysis results demonstrate remarkable improvements in mechanical stability and stress distribution across different fractal designs.

The investigation revealed that higher-order fractal patterns consistently outperform their lower-order counterparts. The Peano-Gosper pattern showed the most dramatic improvement, with von Mises stress reduction from 5.5E+10 N/m² in order 2 to 3.9E+8 N/m² in order 3. Similarly, the Morton Z-curve patterns maintained consistent stress levels around 2.0E+10 N/m², demonstrating reliable mechanical stability across orders 3 and 4. The fabricated sensor of the Peano-Gosper order 3 demonstrated an 8db of SNR improvement when compared to commercial sensors while performing construction tasks.

These advancements address critical limitations in current wearable sensor technology, particularly in maintaining stable electrode-skin contact during dynamic activities. However, several challenges remain for future

investigation. The integration of flexible amplification circuits and wireless communication systems would enhance the sensor's practical applicability. Additionally, further research into surface modification techniques could improve long-term durability and signal stability.

Future work should focus on expanding this fractal-based design methodology to other physiological monitoring applications and developing automated manufacturing processes for complex patterns. The potential applications extend beyond EDA sensing to include other biomonitoring systems, suggesting broader implications for wearable health monitoring technology. Despite current limitations in system integration and long-term durability, this research establishes a promising foundation for next-generation flexible sensors that combine mechanical reliability with enhanced signal quality.

7 Acknowledgments

The work presented in this paper was supported by a National Science Foundation Award (No. 2401745, 'Future of Construction Workplace Health Monitoring'). Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] Jebelli H, Hwang S, Lee SH. EEG-based workers' stress recognition at construction sites. Autom Constr 2018;93:315–24. https://doi.org/10.1016/j.autcon.2018.05.027.

[2] Gautam Yogesh, Liu Y, Jebelli H. A Multimodal Machine Learning Approach to Fatigue Assessment in Construction Workers Using Wearable Technology. 2024 International Conference on Computing in Civil Engineering, Pittsburgh: 2024.

[3] Ojha A, Shakerian S, Habibnezhad M, Jebelli H. Feasibility Verification of Multimodal Wearable Sensing System for Holistic Health Monitoring of Construction Workers. Lecture Notes in Civil Engineering 2023;239:283–94. https://doi.org/10.1007/978-981-19-0503-2_23/FIGURES/3.

[4] Liu Y, Gautam Y, Ojha A, Jebelli H. Studying the Effects of Back-Support Exoskeletons on Workers' Cognitive Load during Material Handling Tasks. 2024 Construction Research Congress (CRC), 2024.

[5] Gautam Y, Ojha A, Liu Y, Jebelli H. Advancing Realtime Physiological Health Monitoring in Construction: A Feasibility Study of Fatigue Estimation of Construction Workers on Smart Devices. 2024 ASCE International Conference on Computing in Civil Engineering, Pittsburgh: 2024.

[6] Setz C, Arnrich B, Schumm J, La Marca R, Tröster G, Ehlert U. Discriminating stress from cognitive load using a wearable eda device. IEEE Transactions on Information Technology in Biomedicine 2010;14:410–7. https://doi.org/10.1109/TITB.2009.2036164.

[7] Gautam Y, Jebelli H. Autoencoder-based Photoplethysmography (PPG) signal reliability enhancement in construction health monitoring. Autom Constr 2024;165:105537. https://doi.org/10.1016/J.AUTCON.2024.105537.

[8] Liu Y, Habibnezhad M, Shayesteh S, Jebelli H, Lee S. Paving the Way for Future EEG Studies in Construction: Dependent Component Analysis for Automatic Ocular Artifact Removal from Brainwave Signals. J Constr Eng Manag 2021;147:04021087. https://doi.org/10.1061/(ASCE)CO.1943-7862.0002097.

[9] Gautam Y, Jebelli H. Design of flexible polyimide-based serpentine EMG sensor for AI-enabled fatigue detection in construction. Sens Biosensing Res 2024;46:100713. https://doi.org/10.1016/J.SBSR.2024.100713.

[10] Hugdahl K. Cognitive influences on human autonomic nervous system function. Curr Opin Neurobiol 1996;6:252–8. https://doi.org/10.1016/S0959-4388(96)80080-8.

[11] Muscatell KA, Eisenberger NI. A Social Neuroscience Perspective on Stress and Health. Soc Personal Psychol Compass 2012;6:890–904. https://doi.org/10.1111/J.1751-9004.2012.00467.X.

[12] Posada-Quintero HF, Chon KH. Innovations in Electrodermal Activity Data Collection and Signal Processing: A Systematic Review. Sensors 2020, Vol 20, Page 479 2020;20:479. https://doi.org/10.3390/S20020479.

[13] Greco A, Valenza G, Scilingo EP. Modeling for the Analysis of the EDA. Advances in Electrodermal Activity Processing with Applications for Mental Health 2016:19–33. https://doi.org/10.1007/978-3-319-46705-4_2.

[14] Tan PX, Kamioka E, Tan PX, Kamioka E. Collaborative Approach Using Psychophysiology and Psychophysics for Optimal Threshold Determination in HAS Service QoE Management. Journal of Computer and Communications 2018;6:57–81. https://doi.org/10.4236/JCC.2018.68005.

[15] Haddad PA, Servati A, Soltanian S, Ko F, Servati P. Effects of Flexible Dry Electrode Design on Electrodermal Activity Stimulus Response Detection. IEEE Trans Biomed Eng 2017;64:2979–87. https://doi.org/10.1109/TBME.2017.2754220.

[16] Hwang S, Jebelli H, Choi B, Choi M, Lee S. Measuring Workers' Emotional State during Construction Tasks Using Wearable EEG. J Constr Eng Manag 2018;144:04018050. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001506.

[17] Shayesteh S, Ojha A, Jebelli H. Workers' Trust in Collaborative Construction Robots: EEG-Based Trust Recognition in an Immersive Environment. Automation and Robotics in the Architecture, Engineering, and Construction Industry 2022:201–15. https://doi.org/10.1007/978-3-030-77163-8_10.

[18] Chen S, Qi J, Fan S, Qiao Z, Yeo JC, Lim CT. Flexible Wearable Sensors for Cardiovascular Health Monitoring. Adv Healthc Mater 2021;10. https://doi.org/10.1002/ADHM.202100116.

[19] Fan JA, Yeo WH, Su Y, Hattori Y, Lee W, Jung SY, et al. Fractal design concepts for stretchable electronics. Nature Communications 2014 5:1 2014;5:1–8. https://doi.org/10.1038/ncomms4266.

Pole-Shaped Object Extraction from LiDAR Point Clouds of Roads

Danesh Shokri[1], Shirin Malihi[2], Saeid Homayouni[3], Christian Larouche[4], Heidar Rastiveis[5], Ioannis Brilakis[6]

[1]Département des Sciences Géomatiques,Université Laval, Canada
[2]Civil Engineering Department, University of Cambridge, United Kingdom
[3]Centre Eau Terre Environnement, Institut National de la Recherche Scientifique, Canada
[4]Département des Sciences Géomatiques, Université Laval, Canada
[5]Civil and Construction Engineering, Oregon State University, USA
[6]Civil Engineering Department, University of Cambridge, United Kingdom

[1]danesh.shokri.1@ulaval.ca, [2]sm2852@cam.ac.uk, [3]saeid.homayouni@inrs.ca,
[4]christian.larouche@scg.ulaval.ca, [5]heidar.rastiveis@oregonstate.edu, [6]ib340@cam.ac.uk,

Abstract -

Mobile Laser Scanner (MLS) technology enables the acquisition of high-precision three-dimensional lidar point cloud data of roadside infrastructure elements, including traffic signs and street poles. This research presents an innovative and computationally efficient algorithm for the extraction of pole-shaped objects from MLS point clouds. The methodology addresses the computational challenges associated with large-scale point cloud processing through a trajectory-based segmentation and ground point filtering. The algorithm implements geometric descriptors, specifically linearity and verticality to identify potential pole structures. Subsequently, the statistical outlier removal (SOR) filter is applied to refine the candidate pole detection results. The final classification of true pole objects is achieved through the application of ground clearance criteria. The algorithm's effectiveness was validated through empirical testing across three distinct urban and suburban environments, utilizing data acquired from two different MLS systems. The experimental results demonstrated successful performance metrics, achieving mean precision and recall rates of 97.21.

Keywords -

Road, Linearity, Verticality, Trajectory, LiDAR, Poles.

1 Introduction

Poles along roads are really important for new technologies like self-driving cars and creating 3D maps [1]. By knowing exactly where these poles are located, we can better understand road conditions and improve how we design and manage transportation infrastructure [2]. While traditional inspection methods involving manual ground-based assessment are resource-intensive and laborious, there is a critical need for expedited, fully automated inspection protocols for roadside infrastructure evaluation [3].

MLS systems have emerged as powerful tools in this field, offering significant advantages over other remote sensing technologies [2]. Unlike optical imaging, which struggles with illumination variations, laser scanning systems can measure distance information directly without being adversely affected by ambient lighting or shadows [4] , [5]. These systems generate three-dimensional spatial information of object surfaces, eliminating the time-intensive computational processes typically required for deriving 3D coordinates from imagery [6].

Previous research has explored various techniques for extracting pole-shaped objects from point clouds, including voxel-based procedures and adaptive radius cylinder models [7]. However, existing approaches often face limitations such as dependence on complex voxel transformations, suboptimal thresholding parameters, and requirements for extensive training datasets and computational resources.

In response to these challenges, the research proposes a novel and computationally efficient algorithm for pole-shaped object extraction from mobile laser scanning point clouds. The methodology introduces geometric descriptors for pole detection while maintaining computational efficiency through systematic data segmentation. By addressing the limitations of existing approaches, this research aims to advance automated infrastructure monitoring and environmental mapping technologies.

2 Proposed Method

2.1 Pre-processing

This subsection addresses the critical challenge of managing the vast volume of lidar point cloud data. It emphasizes the importance of effectively removing both noise and ground points, which is essential for reducing the overall data size and ensuring that only relevant information

is retained. This process plays a key role in minimizing unnecessary data, thereby enhancing the efficiency of subsequent data processing and analysis.

2.1.1 Sectioning

MLS systems leverages trajectory data to efficiently segment point cloud information by recording the vehicle's precise location throughout the scanning process. This trajectory data, which is inherently aligned with the point cloud coordinate system, allows for seamless segmentation of the scanned environment into uniform sections without requiring complex georeferencing. By dividing the trajectory into fixed-length segments—in this case, 90m—the corresponding point cloud data is simultaneously partitioned into equivalent sections, creating a systematic approach to processing large spatial datasets [2]. The primary advantage of this method lies in its ability to transform continuous, complex point cloud information into manageable, discrete units, significantly reducing computational complexity and enabling more focused spatial analysis. Since the trajectory data is substantially smaller in volume compared to the lidar point cloud data, this segmentation approach not only simplifies data processing but also improves overall implementation efficiency, making it an elegant solution for handling extensive MLS datasets. Figure 1-a illustrates the trajectory data overlaid on the collected LiDAR point clouds. It is important to note that an elevation shift is required due to the absence of elevation data in the trajectory. As shown in Figure 1-b, the sectioning of the collected data has been appropriately performed, with each section represented in a distinct color.

subcaption
xcolor

2.1.2 Noisy Points Removal

Noisy points (Figure 2) in MLS point clouds can arise from various sources, such as sensor imperfections, environmental noise, and surface reflectivity variations. These points typically exhibit abnormal elevations or low point density, which makes them distinguishable from the rest of the point cloud. The SOR method is a robust technique used to detect and eliminate these outliers by analyzing the statistical distribution of neighboring points [8].

The SOR algorithm operates by calculating the mean distance from each point to its neighboring points within a specified radius. In this study, the number of nearest neighbors (N) is set to 10. For each point in the point cloud, the mean distance to its 10 nearest neighbors is computed (Obtained based on trial and error). These mean distances are then used to define a threshold, which is typically based on the global mean and standard deviation of the mean distances across the entire dataset. Points that have a mean distance significantly larger than the average (i.e., those that fall outside a certain multiple of the standard deviation) are considered statistical outliers. These outliers are likely to be noisy or erroneous, as they do not align with the expected spatial distribution of neighboring points. The general formula for the SOR algorithm can be expressed as [8]:

$$D_i = \frac{1}{N} \sum_{j=1}^{N} |P_i - P_j| \tag{1}$$

where D_i is the mean distance from point P_i to its nearest neighbors P_j, and N is the number of neighbors (in this case, 10). For each point P_i, the algorithm calculates the mean distance to its neighbors and compares it to the overall distribution of mean distances in the dataset. Points with a mean distance greater than a threshold defined as $\mu + k\sigma$ (where μ is the mean and σ is the standard deviation of all mean distances, and k is a defined constant) are flagged as outliers.

xcolor

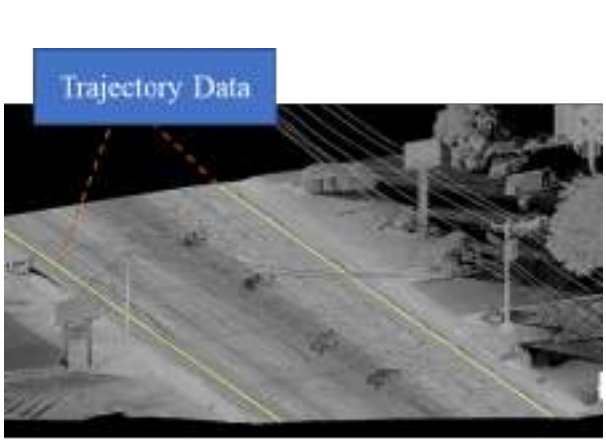
2.1.3 Ground Points Removal

The Cloth Simulation Filter (CSF) method [9] offers an innovative approach to ground point removal in MLS point clouds, addressing the critical challenge of data volume reduction for efficient processing and analysis. By simulating a cloth-like surface that drapes over the point cloud, the algorithm mimics gravitational effects, effectively classifying points as ground or non-ground based on their position relative to this simulated surface. This method demonstrates remarkable versatility, capable of handling diverse terrains including both steep and flat areas, making it particularly valuable for large-scale point cloud datasets.

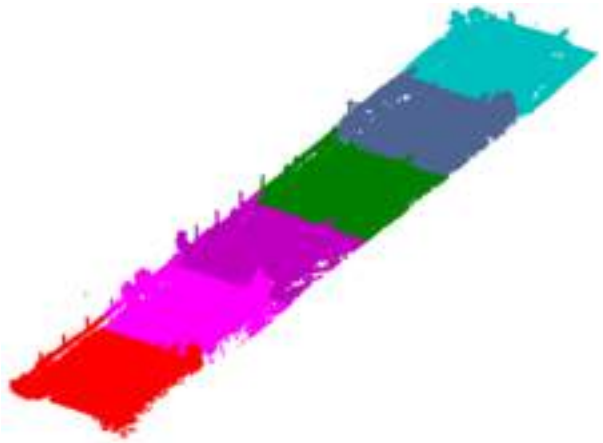
The CSF algorithm's effectiveness hinges on two key parameters: grid size and classification threshold. The grid size, which determines the resolution of the cloth mesh, plays a crucial role in capturing terrain features accurately—in this study, a 0.31m grid size was selected to optimize detail and efficiency. By carefully tuning these parameters, the method can dramatically reduce point cloud volume, as evidenced by its ability to remove approximately 4.5 million ground points from a 5-million-point dataset in just 3 seconds. This computational efficiency, combined with the algorithm's ability to model complex terrain variations, makes the CSF method a powerful tool for preprocessing lidar point cloud data across various environmental contexts.

2.2 Detecting Candidate Poles

This step aims to identify pole-shaped objects among the remaining point clouds, which include various objects



(a) Displaying the trajectory data on the LiDAR point clouds (an elevation shift is needed).



(b) Created sections which partition the point cloud for efficient processing.

Figure 1. (a) Trajectory data overlaid on the LiDAR point clouds, requiring an elevation shift for alignment, and (b) resulting sections dividing the data into manageable units.

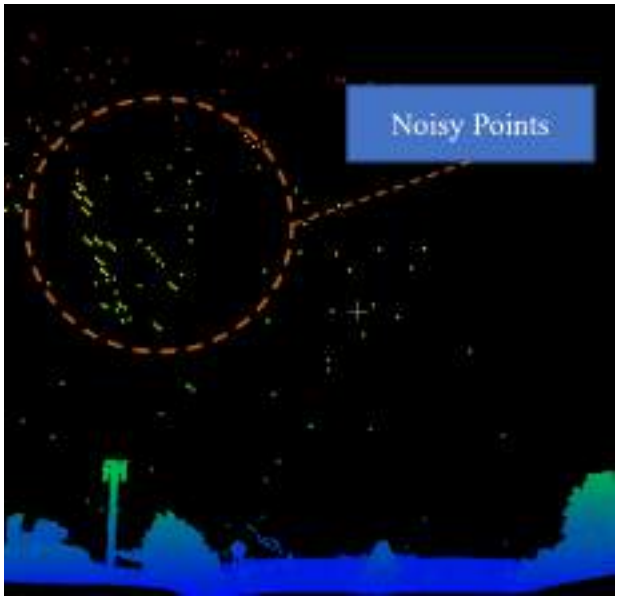


Figure 2. Displaying the noisy points.

that lack meaningful information in isolation. By analyzing the surrounding neighborhood points, we aim to assign meaningful descriptors to each point. For instance, we know that power lines typically exhibit thin, lengthy linear structures, while objects such as poles tend to have a long, vertical shape. To distinguish between these types of objects, rule-based descriptors of linearity and verticality are employed. These two geometric features—verticality and linearity—are particularly effective for characterizing poles. The descriptors are quantified using the eigenvalues obtained from the Principal Component Analysis (PCA) procedure, which captures the directional properties of the point cloud and provides a means to differentiate between vertical and linear structures.

These descriptors are computed for each point in the filtered point cloud dataset after ground point removal. Let $P(i)$ represent the set of points, where i is the index of each point. The descriptors are computed based on eigenvalues derived from PCA of the covariance matrix.

2.2.1 Verticality and Linearity Descriptors

The verticality descriptor $V(i)$ is defined as [2]:

$$V(i) = \frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3} \tag{2}$$

where λ_1, λ_2, λ_3 are the eigenvalues of the covariance matrix, and λ_1 is associated with the dominant vertical direction of the point cloud.

The linearity descriptor $L(i)$ is computed as:

$$L(i) = \frac{\lambda_2}{\lambda_1 + \lambda_2 + \lambda_3} \tag{3}$$

where λ_2 corresponds to the direction of the linearity of the point cloud.

2.2.2 Neighborhood Selection and Thresholding

For each point index i, the neighborhood points are selected based on a spherical neighborhood with radius $r = 0.9$ meters. This can be expressed as:

$$N(i) = \{P(j) \mid \|P(i) - P(j)\| \leq r\} \tag{4}$$

where $N(i)$ is the set of neighboring points for point $P(i)$, and r is the radius for point selection.

2.3 DBSCAN-based Point Cloud Clustering

To detect pole-shaped objects in a point cloud using an unsupervised approach, we employ the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm, which clusters points based on their geometric characteristics [10]. The input to DBSCAN consists of feature vectors for each point, which include the 3D coordinates as well as additional descriptors, such as linearity and verticality. Each point $P(i)$ is represented by a feature vector:

$$\mathbf{x}(i) = \{x_i, y_i, z_i, \text{linearity}_i, \text{verticality}_i\} \tag{5}$$

where x_i, y_i, z_i denote the 3D coordinates of point $P(i)$, and linearity_i and verticality_i are the calculated descriptors based on the local neighborhood and PCA. These descriptors serve as indicators of whether a point is part of a pole-like object or not.

The DBSCAN algorithm requires two parameters: the maximum distance ϵ and the minimum number of points MinPts. The distance threshold ϵ determines the maximum allowed distance between two points for them to be considered as neighbors, and MinPts defines the minimum number of points required in a neighborhood for a point to be classified as a core point. The neighborhood of a point $P(i)$ is defined as the set of points $N_\epsilon(i)$ within a distance of ϵ:

$$N_\epsilon(i) = \{P(j) \mid \|\mathbf{x}(i) - \mathbf{x}(j)\| \leq \epsilon\} \tag{6}$$

where $\|\mathbf{x}(i) - \mathbf{x}(j)\|$ is the Euclidean distance between points $P(i)$ and $P(j)$. If the number of points within this neighborhood is greater than or equal to MinPts, then $P(i)$ is considered a core point, which is a point around which a cluster can form. This is mathematically represented as:

$$|N_\epsilon(i)| \geq \text{MinPts} \tag{7}$$

Points that lie within the neighborhood of a core point but do not themselves have enough points to form a cluster are classified as border points. Border points are included in the same cluster as the core points they are associated with. Any point that is not a core point or a border point is classified as noise.

DBSCAN assigns each core point and its corresponding neighborhood to a cluster, and the final output is a set of clusters $C_1, C_2, \ldots, C_k$, where each cluster contains densely connected points. Points that are classified as noise are not included in any cluster. The overall output of the DBSCAN algorithm is given by:

$$\text{DBSCAN}(\epsilon, \text{MinPts}) = \{C_1, C_2, \ldots, C_k\} \tag{8}$$

where $C_1, C_2, \ldots, C_k$ are the detected clusters corresponding to potential pole-shaped objects. This clustering approach enables the identification of pole-like structures based on their geometric properties without the need for labeled data.

3 Experiment and Results

3.1 Study Area and Results

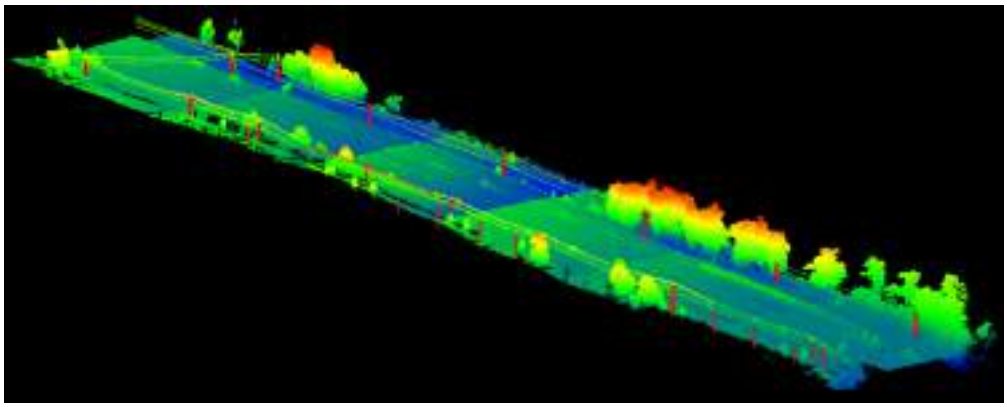
The evaluation of the proposed algorithm is based on datasets from diverse environments, each presenting unique challenges for accurate analysis. The urban region in the USA, with its high object density and reflective surfaces, complicates the distinction of pole-shaped structures due to overlapping vertical objects, occlusions, and varying point densities. The suburban USA region, while less dense, still presents challenges with trees, power lines, and scattered buildings, requiring the algorithm to manage a large dataset with 45 million points and efficiently classify pole-shaped objects. The suburban region in China adds complexity due to geographical differences, sensor calibration variations, and environmental factors like vegetation and local infrastructure. Across all datasets, challenges include point density variations, occlusions, environmental factors, and geographical variability, which demand a robust, adaptive algorithm.

The algorithm demonstrated strong performance (Figure 3) across the test regions, particularly in the suburban USA area where it successfully detected 33,490 pole points with only 796 points undetected and 953 false positives, achieving approximately 97 percent recall accuracy (Table 1). In urban environments, additional challenges arose from phone lines attached to poles and traffic signs, which occasionally led to misclassification. While the algorithm showed reduced accuracy in detecting transmission line pylons and poles with additional equipment such as transformers, its adaptability to various environmental conditions proved effective. The presence of dense trees, cables, and varied building types added complexity to the detection process, yet the algorithm maintained robust performance across different regions, efficiently processing large lidar datasets and adapting to diverse urban and suburban landscapes.

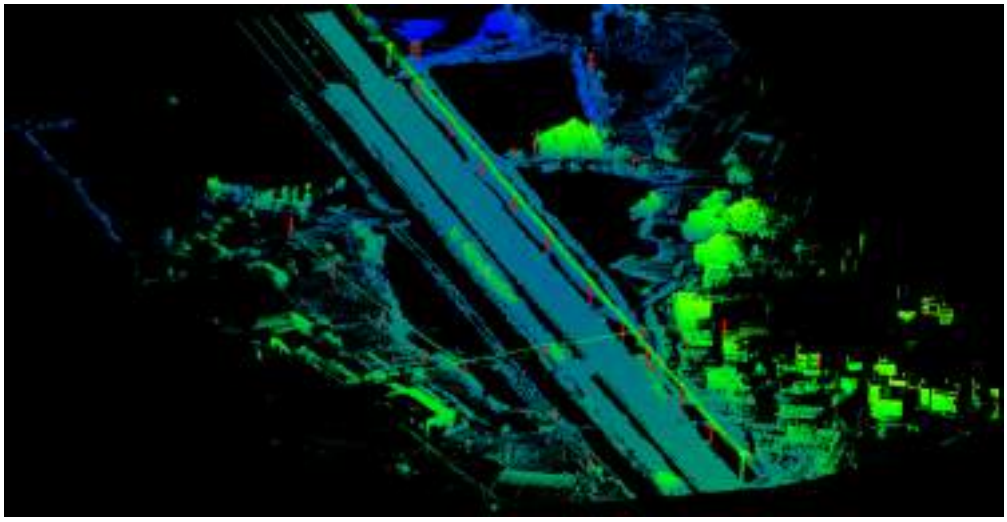
4 Discussion

4.1 Computation System

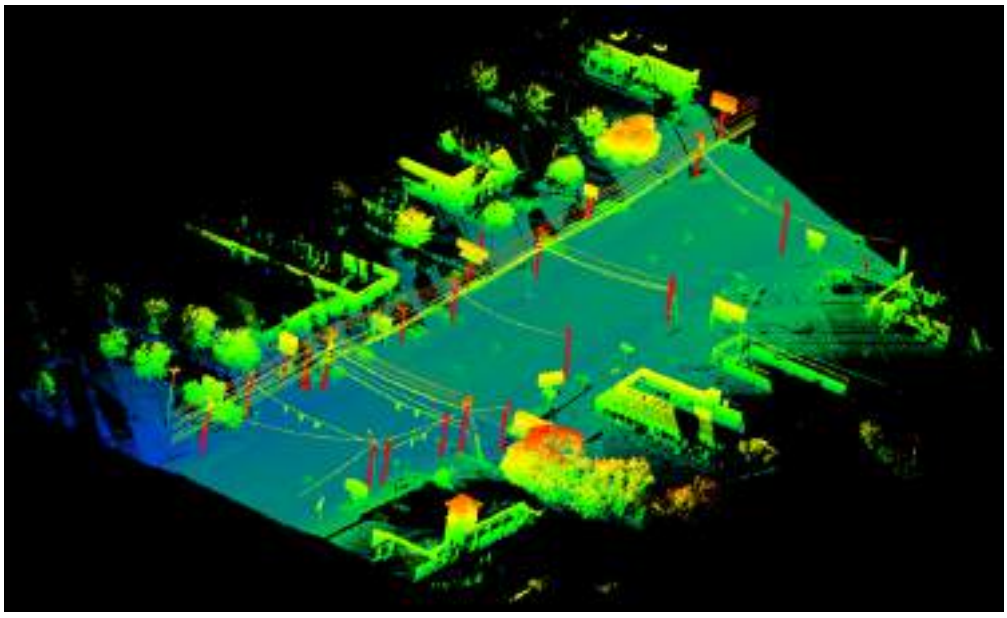
The proposed method presents significant advantages over previous approaches, particularly in terms of system implementation, runtime efficiency, input data requirements, and overall efficiency. Unlike many prior studies that rely on expensive cloud computing systems, the method was executed on an affordable laptop featuring an Intel Core i5 processor, 12 GB of RAM, and an Nvidia GeForce graphics card. This economical setup still



(a) Suburban USA



(b) Suburban China



(c) Urban USA

Figure 3. Visualising the output of the algorithm (the extracted poles are displayed in red color): (a) Suburban USA, (b) Suburban China, (c) Urban USA.

Table 1. Performance metrics of the proposed algorithm across three datasets, with average values.

Dataset	Length (m)	Points (million)	Precision (%)	Recall (%)	F1 Score (%)	Time (s)
Suburban USA	550	45	97.68	97.23	97.45	104
Suburban China	500	19	97.00	95.84	96.42	93
Urban USA	200	15	96.96	94.97	95.95	115
Average	–	–	97.21	96.01	96.61	–

achieved satisfactory performance, making it accessible for practitioners without high-end infrastructure, and particularly suitable for resource-constrained environments like field-based applications or small-scale projects. Additionally, the runtime efficiency of the method is a key strength, processing data in approximately 312 seconds, which is significantly faster than other methods—up to 20 times faster in some cases.

4.2 Datasets

In terms of input data, the method has the advantage of relying solely on 3D lidar point clouds and trajectory data, simplifying the data acquisition process compared to other methods that require additional data sources such as scanlines or labeled training datasets. This makes the algorithm more flexible and easier to implement in situations where such data is difficult or costly to obtain. The method's versatility is further demonstrated by its ability to handle data from multiple MLS platforms and operate effectively in both urban and suburban environments. This adaptability enhances the algorithm's robustness and broadens its applicability across diverse geographic locations and point cloud characteristics. Overall, the proposed algorithm outperforms existing methods by providing high performance on affordable hardware, efficient processing, minimal data requirements, and flexibility across different platforms, making it a practical and valuable tool for pole detection in lidar point clouds.

4.3 Parameter Selection and Optimization

The effectiveness of the proposed algorithm relies on several key parameters, whose values were carefully selected to optimize performance across diverse MLS datasets. While the primary parameters—segmentation length (90 m), SOR nearest neighbors ($N = 10$), and CSF grid size (0.31 m)—were justified in the methodology, additional parameters such as the spherical neighborhood radius ($r = 0.9$ m), DBSCAN clustering parameters ($\epsilon = 0.5$ m, MinPts = 10), and ground clearance threshold (0.5 m) also warrant discussion to elucidate their optimality.

The spherical neighborhood radius of 0.9 meters, used for computing verticality and linearity descriptors, was chosen to encapsulate the local geometry of pole-shaped objects while minimizing interference from adjacent structures. Given that poles in the datasets typically have diameters of 0.2–0.5 meters, a 0.9-meter radius ensures sufficient point inclusion for robust PCA-based eigenvalue computation, capturing vertical extents of 2–3 meters. Empirical trials showed that smaller radii (e.g., 0.5 m) yielded unstable descriptors in sparse regions, whereas larger radii (e.g., 1.5 m) incorporated noise from nearby objects, reducing specificity. This value proved effective across urban and suburban environments, balancing sensitivity and precision.

For DBSCAN clustering, the maximum distance $\epsilon = 0.5$ m and minimum points MinPts = 10 were selected to group points into coherent pole structures. The 0.5-meterϵ aligns with the physical scale of poles and the point density of the MLS data (100–1000 points/m^2), ensuring that points along a single pole are clustered without merging with unrelated objects spaced further apart. A smaller ϵ (e.g., 0.2 m) fragmented poles, while a larger value (e.g., 1.0 m) over-clustered with non-pole features. Similarly,MinPts = 10 reflects the expected point count within a pole's neighborhood, filtering out noise while retaining valid clusters. This combination was iteratively refined to maximize recall and precision, as evidenced by the 97.21% average precision reported in Table 1.

The ground clearance threshold of 0.5 meters, applied as a final classification criterion, distinguishes pole bases from ground-level clutter. This value accommodates terrain variations and sensor height offsets, ensuring that poles are correctly identified above residual ground points post-CSF filtering. Testing revealed that a lower threshold (e.g., 0.2 m) misclassified low vegetation, while a higher threshold (e.g., 1.0 m) missed valid pole bases in uneven areas. The 0.5-meter criterion proved robust across the datasets, enhancing the algorithm's adaptability to diverse landscapes.

These parameter choices were informed by empirical optimization, practical constraints (e.g., computational efficiency on standard hardware), and the physical characteristics of pole-shaped objects in MLS point clouds. While the selected values yielded high performance (average F1 score of 96.61%), future work could explore adaptive parameter tuning based on local point density or environmental context to further enhance generalizability. Nonetheless, the current settings provide a practical and effective baseline for pole detection, as demonstrated by the experimental results.

4.4 Comparison with Baseline Methods

The proposed approach, leveraging trajectory-based segmentation and geometric descriptors, is designed to offer improved efficiency over existing techniques for pole-shaped object extraction from MLS point clouds. While Section 1 highlights limitations of baseline methods—such as voxel-based pole extraction techniques and adaptive radius cylinder models—this subsection provides a qualitative comparison to contextualize our method's advantages.

Unlike voxel-based methods, which rely on computationally intensive grid construction and transformations, our approach uses trajectory data to segment point clouds directly, simplifying preprocessing and reducing computational overhead. Similarly, adaptive radius cylinder models (e.g.,[6]) often require iterative optimization and extensive parameter tuning, whereas our method employs straightforward geometric descriptors (linearity and verticality) to achieve rapid pole detection. A key strength of our algorithm is its independence from labeled training data, a common requirement in many baseline techniques that increases preparation time and resource demands. By contrast, our reliance on raw 3D point clouds and trajectory information enables efficient processing on standard hardware, as demonstrated across diverse urban and suburban datasets (Section 2).

While baseline methods may offer robustness in specific scenarios, such as handling complex occlusions or varying object densities, they typically incur higher computational costs and preparatory complexity. Our approach prioritizes speed and simplicity, achieving high performance (e.g., 97.21% average precision, Table 1 without sacrificing accuracy. This efficiency is particularly valuable for real-time or resource-constrained applications, distinguishing our method from existing techniques that often trade off speed for additional data dependencies or processing steps. Future work could explore quantitative benchmarks against these baselines to further validate these advantages, though the current results underscore the practical benefits of our streamlined design.

4.5 Analysis of Misclassified Cases

While the proposed algorithm demonstrated robust performance across diverse datasets, achieving an average precision of 97.21% and recall of 96.01% Table 1, a small number of misclassified cases warrant critical discussion to understand limitations and guide future improvements. Notably, the algorithm missed two pole-shaped objects in the experimental evaluation, as observed in the suburban USA dataset (Section 3.1), where 796 points were undetected out of 33,490 true pole points.

Several factors may contribute to these misses. First, occlusion by dense vegetation or overlapping infrastructure (e.g., power lines or traffic signs) likely obscured the geometric signatures of these poles. In urban and suburban environments, poles are often situated near trees or cables, which can disrupt the continuity of point cloud data, reducing the effectiveness of linearity and verticality descriptors (Section 2.2). For instance, a pole partially occluded by foliage may exhibit insufficient vertical point density within the 0.9 m spherical neighborhood, failing to meet the DBSCAN clustering criteria ($\epsilon = 0.5$ m, MinPts = 10). Second, poles with non-standard configurations—such as those with attached equipment (e.g., transformers) or irregular bases—may deviate from the expected geometric profile, leading to their exclusion during the ground clearance (0.5 m) or statistical outlier removal stages (Section 2.1.2). The suburban USA dataset, with its mix of trees, power lines, and varied pole designs, presented such challenges, as noted in Section 3.1.

These misclassifications have implications for the algorithm's applicability. In scenarios requiring exhaustive pole detection (e.g., autonomous navigation), missing even a small fraction of poles could impact reliability. However, the low false negative rate (e.g., 796 undetected points versus 33,490 detected) suggests that the algorithm remains highly effective for most practical purposes, such as infrastructure mapping, where near-complete detection suffices. The trade-off between sensitivity and specificity is evident here: stricter geometric and clustering criteria enhance precision (97.21%) by rejecting ambiguous cases but may overlook edge cases like the two missed poles.

To address these limitations, future enhancements could include adaptive neighborhood sizing based on local point density to better handle occlusions, or the integration of contextual features (e.g., pole-top equipment recognition) to improve detection of non-standard poles. Additionally, incorporating multi-pass clustering or post-processing to recover missed candidates could mitigate false negatives, though at the cost of increased computational complexity. These insights, derived from analyzing the misclassified cases, underscore the algorithm's strengths in typical scenarios while highlighting areas for refinement in challenging environments.

5 Conclusion and Future Works

The paper introduces an innovative algorithm for extracting pole-shaped objects from lidar point clouds, addressing the challenges of large-volume spatial data processing in urban and suburban environments. By systematically partitioning point cloud data and employing geometric features like verticality and linearity, the method efficiently identifies and clusters pole-like objects. The algorithm's robustness is demonstrated through extensive testing on complex datasets totaling 1.3 kilometers and 80

million points, achieving an impressive average F1 score of 96.61 percent. Its key advantages include simplicity of implementation, rapid extraction time, independence from training data, and compatibility with various MLS systems. While the results are promising, the authors suggest further validation across different lidar platforms such as airborne or terrestrial laser scanners to comprehensively assess the algorithm's adaptability and performance in diverse environmental conditions.

Acknowledgment

Authors would like to thanks Prof. Haiyan Guan and Wayne A Sarasua for sharing their datasets. Dr. Malihi is supported by the European Union's Horizon 2020 research and innovation programme under the Marie Sk lodowska-Curie grant agreement No. 101034337.

References

[1] Quan Zhang, Yuhang Dai, Tisheng Zhang, Chi Guo, and Xiaoji Niu. Road semantic-enhanced land vehicle integrated navigation in gnss denied environments. *IEEE Transactions on Intelligent Transportation Systems*, 25(12):20889–20899, 2024. doi:10.1109/TITS.2024.3449892.

[2] Danesh Shokri, Heidar Rastiveis, Seyed Mohammad Sheikholeslami, Reza Shahhoseini, and Jonathan Li. Fast extraction of power lines from mobile lidar point clouds based on svm classification in non-urban area. *Earth Observation and Geomatics Engineering*, 5(2):63–73, 2021. ISSN 2588-4352. doi:10.22059/eoge.2022.317348.1092.

[3] Yaowen Pei, Feng Chen, Tao Ma, and Gonghui Gu. A comparative review study on the electrified road structures: Performances, sustainability, and prospects. *Structures*, 62:106185, 2024. ISSN 2352-0124. doi:10.1016/j.istruc.2024.106185. URL https://www.sciencedirect.com/science/article/pii/S2352012424003370.

[4] Inam Ullah, Deepak Adhikari, Habib Khan, M. Shahid Anwar, Shabir Ahmad, and Xiaoshan Bai. Mobile robot localization: Current challenges and future prospective. *Computer Science Review*, 53:100651, 2024. ISSN 1574-0137. doi:10.1016/j.cosrev.2024.100651. URL https://www.sciencedirect.com/science/article/pii/S1574013724000352.

[5] S. Malihi, M. J. Valadan Zoej, M. Hahn, M. Mokhtarzade, and H. Arefi. 3d building reconstruction using dense photogrammetric point cloud. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B3:71–74, 2016. doi:10.5194/isprs-archives-XLI-B3-71-2016. URL https://isprs-archives.copernicus.org/articles/XLI-B3/71/2016/.

[6] Yueqian Shen, Junjun Huang, Jinguo Wang, Jundi Jiang, Junxi Li, and Vagner Ferreira. A review and future directions of techniques for extracting powerlines and pylons from lidar point clouds. *International Journal of Applied Earth Observation and Geoinformation*, 132:104056, 2024. ISSN 1569-8432. doi:10.1016/j.jag.2024.104056. URL https://www.sciencedirect.com/science/article/pii/S1569843224004102.

[7] D. Shokri, H. Rastiveis, A. Shams, and W. A. Sarasua. Utility poles extraction from mobile lidar data in urban area based on density information. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-4/W18:1001–1007, 2019. doi:10.5194/isprs-archives-XLII-4-W18-1001-2019. URL https://isprs-archives.copernicus.org/articles/XLII-4-W18/1001/2019/.

[8] Haris Balta, Jasmin Velagic, Walter Bosschaerts, Geert De Cubber, and Bruno Siciliano. Fast statistical outlier removal based method for large 3D point clouds of outdoor environments. *IFAC-PapersOnLine*, 51(22):348–353, 2018. ISSN 2405-8963. doi:10.1016/j.ifacol.2018.11.566. 12th IFAC Symposium on Robot Control SYROCO 2018.

[9] Wuming Zhang, Jianbo Qi, Peng Wan, Hongtao Wang, Donghui Xie, Xiaoyan Wang, and Guangjian Yan. An easy-to-use airborne lidar data filtering method based on cloth simulation. *Remote Sensing*, 8(6):501, 2016. ISSN 2072-4292. doi:10.3390/rs8060501. URL https://www.mdpi.com/2072-4292/8/6/501.

[10] Kamran Khan, Saif Ur Rehman, Kamran Aziz, Simon Fong, and S. Sarasvady. Dbscan: Past, present and future. In *The Fifth International Conference on the Applications of Digital Information and Web Technologies (ICADIWT 2014)*, pages 232–238, 2014. doi:10.1109/ICADIWT.2014.6814687.

Real-time Multimodal Sensing System for Additive Construction by Extrusion: Integrating thermal, depth and RGB data

Yilin Cai[1] and Julie Ann Hartell[1] and Ashrant Aryal[1]

[1] Department of Construction Science, Texas A&M University, College Station, TX 77843, USA
caiyilin@tamu.edu, julie.hartell@tamu.edu, ashrantaryal@tamu.edu

Abstract

Additive construction by extrusion addresses challenges such as labor shortages and inconsistent quality in traditional construction methods by offering automation, precision, and reduced waste. However, it remains challenging for large scale implementation due to dynamic printing conditions and environmental influences. Real-time control can be greatly beneficial in addressing these variations, which relies on a comprehensive and accurate real-time monitoring system. In this paper, we describe our ongoing work on developing a real-time multimodal sensing system that integrates thermal, depth and RGB data. By combining these sensors, the system can capture surface temperature and geometric data of extruded material, providing detailed feedback on its strength, extrudability and buildability. It provides new insights and information for quality monitoring systems in additive construction by extrusion, achieving a better balance between detailed detection and real-time monitoring. The methodology includes RGB-thermal image alignment, automated sample detection, and structured light-based depth sensing to refine geometric and temperature data. This system lays the groundwork for future advancements in predictive analysis and dynamic real-time printing control.

Keywords –

3D concrete printing; Multimodal sensing; Thermal; Depth; RGB; Sensor fusion;

1 Introduction

Additive construction by extrusion offers solutions that can help alleviate issues such as labor shortages and hard-to-control quality within the construction industry. [1]. It shortens construction time through continuous layer-by-layer deposition, parallel operations, and reduced reliance on curing delays[2]. Specifically, labor demand is reduced by automating processes such as material deposition and eliminating the need for formwork, which traditionally requires intensive manual effort. The precision of additive ensures consistent material distribution, minimizes human error, and enables high-quality results even for complex, customized designs. Furthermore, it optimizes material usage by precisely allocating resources, reduces waste and promotes sustainability by incorporating eco-friendly materials, making it a cost-effective and environmentally conscious solution[3].

Additive construction by extrusion faces a number of challenges when it comes to large scale implementation. First, the printing process is highly dynamic. The complex interplay of parameters such as material rheology, extrusion speed and mechanical motion can lead to unpredictable variations, making it difficult to fully rely on pre-programmed paths and prone to geometric defects. Secondly, printed structures exhibit a high degree of vulnerability to changes in material properties. Even small fluctuations in viscosity or ratios can trigger interlayer destabilization or even lead to structural collapse[4]. In addition, environmental factors such as temperature, humidity, and solar irradiance can affect early-age and final properties of concrete, including strength and durability[5, 6]. These uncertainties highlight the need for a real-time feedback and control system capable of dynamically adjusting printing parameters to efficiently maintain print quality during construction. Such a system requires accurate, fast-response, and highly automated monitoring capabilities, with the ability to capture complex variations, including geometric features and temperature differences.

In this paper, we present the development of a real-time multimodal sensing system designed to address these challenges. The proposed system aims to establish a comprehensive near real-time quality inspection framework to monitor both geometric fidelity and abnormal temperature variations. This approach enables

automatic measurement of geometric deformations and concrete surface temperature shortly after deposition, with the results reported to the operator for informed decision-making. The system lays the foundation for real-time control in additive construction by extrusion. The next section provides a brief review of existing literature on reality capture and non-contact inspection methods in the field of additive construction by extrusion.

2 Related Work

Geometry Monitoring for Real-Time Control: Numerous real-time monitoring and control systems for additive construction by extrusion employ vision-based approaches, utilizing 2D RGB cameras to capture and analyze geometric features. These systems typically position cameras either directly above or at an angle overlooking the printed samples, providing continuous image acquisition of the concrete layers. For example, Kazemian[7] developed a system using a 720p camera to monitor samples with a width of 40 mm, achieving width measurements with an error margin of ±1.7 mm using the OpenCV package. Similarly, Barjuei[8] employed an industrial camera to establish a real-time vision-based control framework for layer-width monitoring demonstrating good adaptability and low error rates. Another study by Jhun[9] utilized a 2D laser and camera module to monitor thickness, width, and cross-sectional area, forming the basis of a real-time geometric quality monitoring system. Additionally, Martin[10] implemented a low-cost depth camera for real-time quality monitoring, focusing on the width and height of layers. The system improved accuracy significantly after applying a statistical filter. While these systems excel in monitoring specific geometric parameters such as width or height, providing detailed feedback on more intricate aspects of concrete, such as collapse or deformation, presents additional challenges.

To enhance the automation of geometric detail monitoring, some researchers have explored the potential of AI technologies and advanced instruments. For instance, Davtalab[11] trained a Convolutional Neural Network (CNN) to detect layer deformation, achieving an accuracy and F1 score greater than 90%. Similarly, Senthilnathan[12] utilized computer vision to analyze images containing detailed deformations by converting them into two-bin linear binary patterns. Deformations were detected using information entropy, demonstrating high sensitivity in identifying defects. Furthermore, Xu[13] employed high-precision structured light scanning to inspect 3D-printed concrete structures, achieving highly accurate geometric monitoring and introducing a novel method for evaluating printing quality. While these methods offer high-precision geometric inspection, their processing times can be relatively extended, which may make them more commonly applied to post-printing quality assessment rather than real-time monitoring.

Exploring Beyond Geometric Features: In addition to geometric features, some researchers have explored other variables as the focus of sensing and monitoring systems for quality control. For instance, Kazemian[14] compared several variables, including power consumption of the agitator motor, extrusion pressure, electrical resistivity, and computer vision, to assess their effectiveness in ensuring printing quality. The results indicated that computer vision-based methods were the most effective for controlling printing parameters and maintaining quality, while electricity consumption was the least effective. Similarly, Rehman[15] developed an integrated sensor-printer system for real-time quality control, which monitors penetration, compression, shear, and tensile tests during the printing process to provide feedback on the performance of freshly extruded concrete. Additionally, Banijamali[16] proposed a novel index combining electrical resistivity and temperature to rapidly evaluate the strength of 3D-printed concrete structures, providing an innovative method for early-age strength monitoring. While these studies provide valuable insights into strength prediction and real-time quality control, their emphasis on mechanical properties leaves opportunities to further explore the role of environmental and material temperatures in influencing extrudability and buildability during the early stages of the printing process

Temperature has been identified as a critical factor affecting the quality of extruded sample. Both environmental temperature and the heat generated during curing influence the structural integrity of the printed concrete[17]. While elevated temperatures can impact strength[18], curing times and heat-induced stresses may lead to deformation and affect bonding characteristics[19, 20]. More critically, temperature directly governs the fresh properties of concrete, including viscosity, yield strength, and flowability, which are fundamental in determining its extrudability and buildability during the 3D printing process[21].

Therefore, real-time monitoring of temperature is essential, as it directly correlates with key printing parameters such as nozzle traveling speed and interval time between layers and/or material dosing. However, the potential of temperature monitoring for real-time quality control in 3D concrete printing has yet to be fully explored.

Summary and Research Gap: Despite significant advancements in geometry monitoring and real-time control for additive construction by extrusion, existing methods, which often rely on single-sensor setups with

relatively focused data sources, may benefit from further enhancements to achieve greater comprehensiveness in observed data dimensions, detail levels, and accuracy. Vision-based approaches effectively monitor parameters like width and height but may have limitations in capturing finer deformations or collapse. While advanced techniques such as AI models and structured light scanning demonstrate high precision, they are often more commonly applied to post-printing quality checks. Additionally, the potential of real-time temperature monitoring, which affects fresh material properties and plays a crucial role in early strength gain, influencing both extrudability and buildability, has yet to be fully explored. To address these shortcomings, this study proposes a multi-modal system that integrates thermal, depth, and RGB data, enabling the automated extraction of detailed, accurate geometric and temperature information to support real-time control.

3 Multimodal sensing system setup

The developed system integrates a FLIR Lepton 3.5 thermal camera and an Astra 2 RGB-depth camera to provide comprehensive monitoring of surface temperature and sample geometry during the concrete printing process. The thermal camera operates with a resolution of 160 x 120 pixels, an operational range of -10°C to 80°C, a dynamic range of -10°C to 400°C, a framerate of 8 fps, and a horizontal field of view of 57º, enabling real-time thermal data acquisition. The Astra 2 RGB-depth camera captures high-resolution RGB images at 1920 x 1080 pixels with a framerate of 30 fps and a horizontal field of view of 74.7º, while its depth function provides depth data with a resolution of 1600 x 1200 pixels at 30 fps. The depth camera employs a structured light sensor with a horizontal and vertical field of view of 58.2° and 45.2°, respectively, operating within a range of 0.6 m to 8 m and achieving a precision of $\leqslant 0.16\%$ at 1 m. Structured light technology projects a patterned field of light onto the surface, capturing deformations caused by surface geometry to reconstruct detailed 3D shapes, which is particularly effective for complex geometries [13].



Figure 1. (a)Setting of the thermal-RGB-depth camera (b)Setting of the system

The thermal camera is securely mounted on top of the RGB-depth camera using a fixed bracket, positioned 5 cm above it, as shown in Figure 1(a). This combined setup is mounted on a movable platform, located 30 cm away from the extruder and angled at 60 degrees relative to the printing platform, providing a comprehensive overview of the printing process and the sample, as illustrated in Figure 1(b). This integrated multi-modal

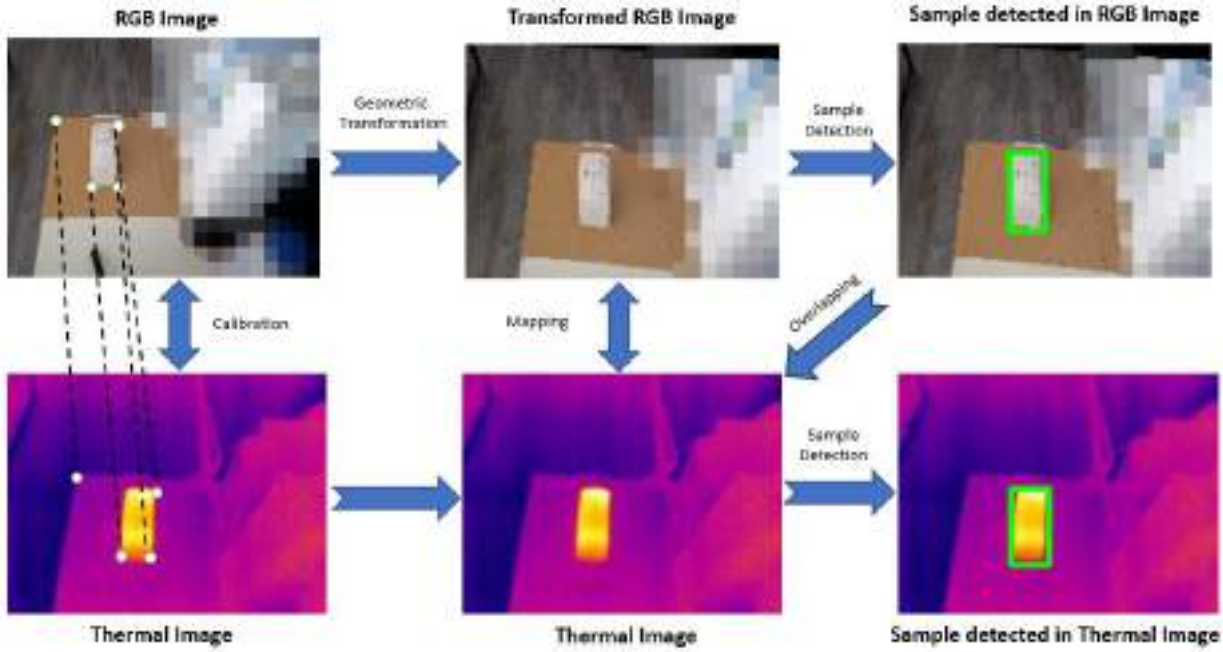


Figure 2. Thermal-RGB data alignment

sensing system enables the automated capture of detailed geometric and temperature data, supporting real-time control and enhancing the precision and effectiveness of the monitoring process during 3D concrete printing.

3.1 Surface temperature capturing

The automated process for capturing the surface temperature of a sample is as follows (shown in Figure 2): First, thermal and RGB images are captured simultaneously at a frame rate of 8 FPS, determined by the lower frame rate of the two cameras. Due to differences in the field of view (FOV) and resolution, the RGB image undergoes a geometric transformation to align its perspective and scale with that of the thermal image. This transformation is guided by a calibration process. Next, a series of image processing steps is applied to the RGB image to identify the sample's contour. The detected sample location in the transformed RGB image is then mapped onto the thermal image, enabling precise localization of the sample within the thermal image. Finally, the temperature data from the selected region in the thermal image is extracted as the sample's surface temperature. The details of each step are explained in the following paragraphs.

Geometric transformation: The RGB images undergoes a calibration process to align with the thermal images. To calibrate it more precisely, we put 4 heating points as calibration points on the printing platform so that they are obvious to identify in both RGB and thermal images. Once the calibration points are identified, the perspective transformation matrix H can be calculated by the OpenCV's package [21]. Then each pixel in the destination image to corresponding pixel in the source image can be calculated using:

$$\begin{bmatrix} x' \\ y' \\ \omega' \end{bmatrix} = H \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (1)$$

Where (x, y) are the coordinates in the source image, (x', y') are the coordinates in the destination image, ω' is the scaling factor for normalization.

The main problems encountered during image transformations involve the extrapolation of non-existing pixels and the interpolation of pixel values. In cases where pixels in the destination image map to out-of-bound coordinates in the source image, extrapolation method which ensures that such pixels remain unmodified in the destination image is employed. Additionally, because the mapped coordinates are often fractional values, bilinear interpolation is used to estimate the pixel values at these fractional coordinates.

Sample detection: to automate the process of sample detection, there are 4 steps as follows: 1) Grayscale conversion, 2) Edge detection, 3) Contour detection, 4) Sample detection. First, the transformed RGB image is converted into a single-channel grayscale image, suitable for edge detection algorithms. The grayscale intensity is given by:

$$I_{gray}(x, y) = 0.2989 \cdot R(x, y) + 0.5870 \cdot G(x, y) + 0.1140 \cdot B(x, y) \quad (2)$$

Where $R(x, y)$, $G(x, y)$, $B(x, y)$ are the red, green and blue intensity values at pixel (x, y). $I_{gray}(x, y)$ is the grayscale intensity.

Then, Canny algorithm[22] is adopted to detect edges. This algorithm identifies regions of rapid intensity change, which often correspond to object boundaries. The gradient of the grayscale image is computed in the x- and y- directions using Sobel filters:

$$G_x = \frac{\partial I_{gray}}{\partial x}, \quad G_y = \frac{\partial I_{gray}}{\partial y} \quad (3)$$

The gradient magnitude and direction are then calculated by:

$$G = \sqrt{G_x^2 + G_y^2} \quad (4)$$

$$\theta = \arctan\left(\frac{G_y}{G_x}\right) \quad (5)$$

Then, Non-Maximum Suppression process eliminates spurious responses by retaining only the strongest gradient magnitudes along the direction of the edge. Through Double Thresholding, pixels with gradient magnitudes above the upper threshold are classified as strong edges, and those below a lower threshold are discarded. Pixels between the thresholds are treated as weak edges. Finally, Edge Tracking by Hysteresis preserves weak edges that are connected to strong edges while suppressing isolated weak edges, producing a binary edge map. Then, contours are



Figure 3. Steps to detect sample in RGB images

detected in the binary edge map. A contour is defined as a curve joining all continuous points along a boundary with the same intensity. The contour with the largest area is selected as the sample's contour. The contour detection example is shown in Figure 3.

Alignment and capturing: after detecting the sample region in the transformed RGB image, a rectangular bounding box is generated and directly applied to the thermal image. The temperature data in each pixel within the bounding box are then extracted and recorded as surface temperature.

3.2 Geometry monitoring

The system for sample geometry monitoring uses the same camera with depth mode.

Point clouds generation: The monitoring process utilizes the in-built point cloud mode, which generates a point cloud based on the depth map and the camera's intrinsic parameters. The (x, y, z) coordinate of each point are computed using the following equations:

$$x = (u - cx) \cdot \frac{z}{f_x} \quad (6)$$

$$y = (v - cy) \cdot \frac{z}{f_y} \quad (7)$$

$$z = depth(u, v) \quad (8)$$

Where u, v are the pixel coordinates in the 2D image, z is the depth value. cx, cy are the center of the optical axis. f_x, $f_x$are the focal length of the camera.

Background filtering: Once the point cloud is generated, an initial coordinate-based filter is applied to remove the background. Given the fixed relative position between the sample and the camera, the filter coordinates can be predefined as constant values. After applying the background filtering, the remaining point cloud consists solely of the platform and the monitored sample.

Plane identification and filtering: The processed point cloud undergoes a plane segmentation and filtering step to isolate the sample. The RANSAC (Random Sample Consensus)[23] algorithm is employed to identify the primary plane, which corresponds to the printing platform. Through iterative random sampling of points, potential planes are generated, and the distances between the points and these planes are calculated. Points within a predefined distance threshold are classified as inliers of a given plane. After 1000 iterations, the plane with the largest number of inliers is identified as the platform plane. To further refine the results, additional points within the threshold distance to the identified plane are included as part of the platform to remove noise. Once the platform points are filtered out, the remaining points in the point cloud are considered to represent the sample.

Outlier filtering: After the removal of the platform plane, some outliers may still remain in the point cloud. To address this, Statistical Outlier Removal (SOR) is employed to filter out these noise points. This process begins by calculating the average distance between each point and its neighboring points (e.g., the 20 nearest neighbors) to establish the global distance distribution, including the mean and standard deviation. Points with distances exceeding two standard deviations from the mean are identified as outliers and subsequently removed from the dataset. This approach, based on detecting local density variations in the point cloud, effectively eliminates noise in sparsely populated regions, resulting in a refined and clean point cloud. The process of the point clouds processing is shown in Figure 4.

4 Discussion

This system emphasizes detailed geometric and temperature inspections, delivering comprehensive and refined data to enhance real-time control in 3D concrete printing. The integration of the thermal and RGB data can extract temperature and enable continuous and automated surface temperature measurement. With its sensor refresh rate capable of handling printing speeds up to 30 cm/s, the system supports most industrial printing conditions[24-26]. Integrated with an established model of material and process parameters, the monitoring system facilitates early identification of conditions that balance the critical aspects of extrudability and buildability.

Early changes in layer geometries during the printing process can be influenced by key process parameters, such as printing speed, extrusion rate, nozzle deposition

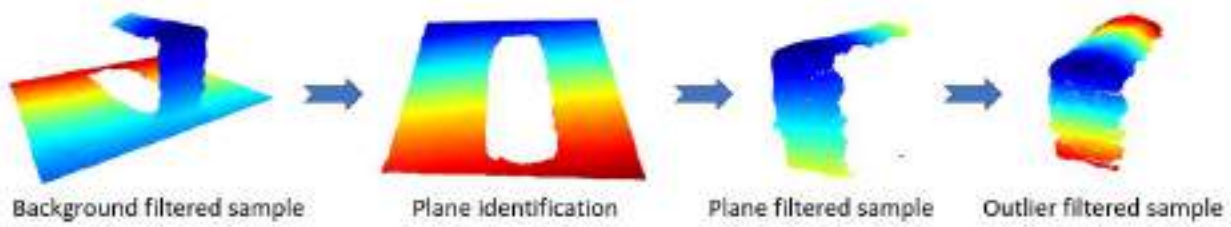


Figure 4. Point clouds processing and filtering

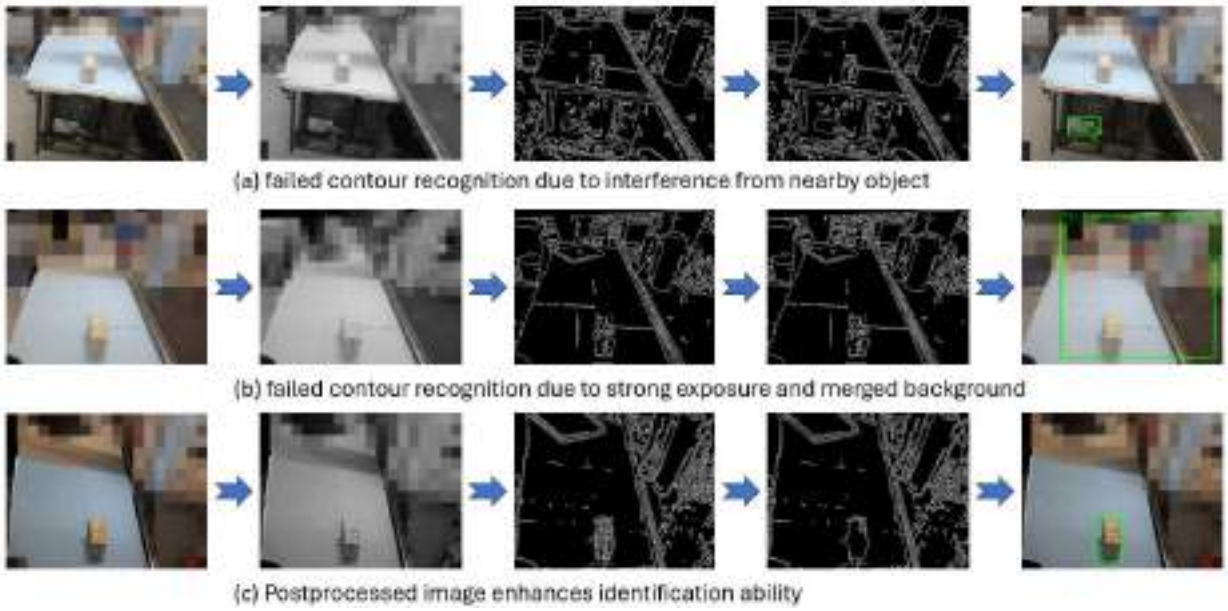


Figure 5. Object identification under different scenarios

height, and material properties like yield stress, stiffness, and early-age strength gain. However, geometry alone may not be sufficient to pinpoint parameters requiring adjustment. By combining object temperature - which reflects on green material properties and concrete hydration behavior - the combined data set could provide insights into which material and/or process parameters may need modification. For instance, localized temperature anomalies, such as excessive gradient or out-of-range values, combined with a slight change in layer height could indicate the potential for future excessive layer deformation or inadequate interlayer bonding, thus enabling predictive interventions on process parameters or material dosing. This comprehensive approach supports dynamic refinements to printing parameters ensuring precision, stability, and structural integrity throughout the 3D printing process.

One of the main challenges of this system lies in the automatic recognition of geometric features in images, which is sensitive to background settings and lighting conditions. When contrast is insufficient, such as under strong lighting, high exposure, or when the sample color closely matches the background, contour detection may fail. For instance, in Figure 5(b), strong exposure caused the background and sample to merge visually, resulting in failed contour recognition. Similarly, Figure 5(a) illustrates interference from other high-contrast contours in the background, which disrupted sample detection. To overcome these challenges, it is necessary to eliminate environmental interferences and enhance recognition accuracy. Techniques such as reducing brightness, increasing contrast, or applying filters can significantly improve detection performance. For example, as shown in Figure 5(c), post-processing the image to enhance contrast led to a substantial improvement in contour recognition, demonstrating the potential of post-processing in ensuring reliable image-based monitoring. These enhancements are essential for maintaining the robustness of the system in diverse operational environments.

The geometry monitoring system developed in this study provides valuable support for real-time quality control in 3D concrete printing. In terms of measurement accuracy, the system shows better performance in width measurements compared to height measurements. As shown in Table 1, the width measurement error is approximately 3.5 mm, while the height measurement error is around 5.4 mm. While these results may not surpass those of some existing systems, the approach offers distinct advantages. By using depth information to generate point clouds, the system captures more detailed geometric features than standard 2D RGB cameras. For example, at a distance of 0.7 m, it can detect deformations as small as 5 mm, which is helpful for identifying potential concrete layer deformations. Figure 6 demonstrates the system's ability to detect complex height deformations, enabling operators to monitor and adjust the printing process as needed.

Table 1 Measurement of sample width and height

Type	Captured(mm)	Actual(mm)
Width	62.4	58.9
Height	89.1	94.5

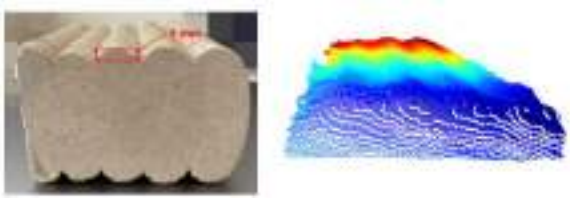


Figure 6. Sample with point clouds image

The use of a structured light-based depth sensor for geometric measurements enables high-resolution data capture but comes with several limitations in various environmental conditions. One major challenge is its sensitivity to lighting conditions, particularly in outdoor applications. In regions with strong solar radiation, the projected infrared pattern used by structured light sensors can be significantly disrupted, leading to reduced measurement accuracy or data loss. This constraint limits their effectiveness in open construction sites or large-scale outdoor 3D printing operations where environmental light fluctuations are unavoidable.

Additionally, the performance of the selected depth sensors is constrained by their frame rate, which is not particularly high. In large-scale applications, such as continuous 3D concrete printing over expansive areas, this limitation may hinder its ability to track fast-moving printing nozzles or dynamically adjust extrusion parameters in real time.

Another practical challenge arises from the presence of extraneous objects in real-world construction environments. Debris, scaffolding, and other surrounding elements can introduce noise in the captured data, making object detection more difficult. These distractions can degrade the reliability of geometric measurements and increase the risk of false detections, which is critical for autonomous or semi-automated construction processes.

From an industrial application perspective, future development should focus on selecting more advanced sensors with enhanced robustness against environmental disturbances. LIDAR-based sensors, for example, are less affected by ambient lighting and could serve as an alternative for outdoor applications, though at a higher cost [27, 28]. Additionally, improving the robustness of object detection algorithms through deep learning-based techniques could mitigate the impact of background clutter and ensure reliable detection of the printed structure. These advancements will be crucial for enabling fully autonomous and scalable 3D concrete printing systems in real-world industrial settings.

5 Conclusion

This study presents the development of a multi-modal sensing system designed to enhance real-time quality control for additive construction by extrusion by integrating thermal, depth and RGB data. As work in progress, we described the system and methodology we use to capture these data. The system aims to provide detailed feedback on geometric features and sample surface temperature. It provides new insights and information for quality monitoring, achieving a better balance between detailed geometric detection and real-time monitoring. Future work will focus on improving the system's resiliency to adapt to various outdoor scenarios and industrial environments. Additionally, future efforts will also involve developing an AI-based framework to process the captured data, enabling real-time adjustment of printing parameters and the ability to predict potential risks, ensuring improved construction quality.

References

[1] M. A. Hossain, A. Zhumabekova, S. C. Paul, and J. R. Kim, "A review of 3D printing in construction and its impact on the labor market," *Sustainability*, vol. 12, no. 20, p. 8492, 2020.

[2] S. C. Paul, G. P. Van Zijl, M. J. Tan, and I. Gibson, "A review of 3D concrete printing systems and materials properties: Current status and future research prospects," *Rapid Prototyping Journal*, vol. 24, no. 4, pp. 784-798, 2018.

[3] G. H. Ahmed, "A review of “3D concrete printing”: Materials and process characterization, economic considerations and environmental sustainability," *Journal of Building Engineering*, vol. 66, p. 105863, 2023.

[4] S. Zaribaf, "S.: Analysis of interlayer bond in 3D printed concrete," Master thesis, Polytechnic University of Catalonia, Barcelona, Spain, 2022.

[5] A. Cicione, J. Kruger, R. S. Walls, and G. Van Zijl, "An experimental study of the behavior of 3D printed concrete at elevated temperatures," *Fire Safety Journal*, vol. 120, p. 103075, 2021.

[6] B. Sun, P. Li, D. Wang, J. Ye, G. Liu, and W. Zhao, "Evaluation of mechanical properties and anisotropy of 3D printed concrete at different temperatures," in *Structures*, 2023, vol. 51: Elsevier, pp. 391-401.

[7] A. Kazemian, X. Yuan, O. Davtalab, and B. Khoshnevis, "Computer vision for real-time extrusion quality monitoring and control in robotic construction," *Automation in Construction*, vol. 101, pp. 92-98, 2019.

[8] E. S. Barjuei, E. Courteille, D. Rangeard, F. Marie, and A. Perrot, "Real-time vision-based

control of industrial manipulators for layer-width setting in concrete 3D printing applications," *Advances in Industrial and Manufacturing Engineering*, vol. 5, p. 100094, 2022.

[9] J. Jhun, D.-H. Lee, A. U. Rehman, S. Kang, and J.-H. Kim, "Development of a Real-time Geometric Quality Monitoring System for Extruded Filaments of 3D Concrete Printing Construction," *IEEE Access*, 2024.

[10] M. Martin, K. Banijamali, and A. Kazemian, "Reality Capture Technologies for Automated Quality Control during Construction 3D Printing," in *Computing in Civil Engineering 2023*, 2024, pp. 361-370.

[11] O. Davtalab, A. Kazemian, X. Yuan, and B. Khoshnevis, "Automated inspection in robotic additive manufacturing using deep learning for layer deformation detection," *Journal of Intelligent Manufacturing*, vol. 33, no. 3, pp. 771-784, 2022.

[12] S. Senthilnathan and B. Raphael, "Using computer vision for monitoring the quality of 3d-printed concrete structures," *Sustainability*, vol. 14, no. 23, p. 15682, 2022.

[13] J. Xu *et al.*, "Inspecting manufacturing precision of 3D printed concrete parts based on geometric dimensioning and tolerancing," *Automation in Construction*, vol. 117, p. 103233, 2020.

[14] A. Kazemian and B. Khoshnevis, "Real-time extrusion quality monitoring techniques for construction 3D printing," *Construction and Building Materials*, vol. 303, p. 124520, 2021.

[15] A. Rehman, I. Kim, and J. Kim, "Towards full automation in 3D concrete printing construction: development of an automated and inline test method for in-situ assessment of structural build-up and quality of concrete. Dev," *Built Environ*, 2023.

[16] K. Banijamali, P. Vosoughi, G. Arce, H. Noorvand, M. Hassan, and A. Kazemian, "Early-Age Strength Monitoring of Sensor-Embedded 3D Printed Structures," in *Construction Research Congress 2024*, 2024, pp. 137-147.

[17] M. K. Mohan, A. Rahul, G. De Schutter, and K. Van Tittelboom, "Extrusion-based concrete 3D printing from a material perspective: A state-of-the-art review," *Cement and Concrete Composites*, vol. 115, p. 103855, 2021.

[18] P. Shakor, S. Nejadi, G. Paul, J. Sanjayan, and A. Nazari, "Mechanical properties of cement-based materials and effect of elevated temperature on three-dimensional (3-D) printed mortar specimens in inkjet 3-D printing," *ACI Materials Journal*, vol. 116, no. 2, pp. 55-67, 2019.

[19] L. Dong, Y. Yang, Z. Liu, Y. Zhang, and C. Wu, "Interface bonding characteristics of 3D printed ultra-high performance concrete after elevated temperatures," *Journal of Building Engineering*, p. 109801, 2024.

[20] P. Shakor, S. Nejadi, G. Paul, and N. Gowripalan, "Effects of different orientation angle, size, surface roughness, and heat curing on mechanical behavior of 3D printed cement mortar with/without glass fiber in powder-based 3DP," *3D Printing and Additive Manufacturing*, vol. 10, no. 2, pp. 330-355, 2023.

[21] Y. Chen, S. He, Y. Gan, O. Çopuroğlu, F. Veer, and E. Schlangen, "A review of printing strategies, sustainable cementitious materials and characterization methods in the context of extrusion-based 3D concrete printing," *Journal of Building Engineering*, vol. 45, p. 103599, 2022.

[22] Canny Edge Detection [Online] Available: https://docs.opencv.org/4.x/da/d22/tutorial_py_canny.html, Accessed:12/10/2024

[23] K. G. Derpanis, "Overview of the RANSAC Algorithm," *Image Rochester NY*, vol. 4, no. 1, pp. 2-3, 2010.

[24] W. Zhou, Y. Zhang, L. Ma, and V. C. Li, "Influence of printing parameters on 3D printing engineered cementitious composites (3DP-ECC)," *Cement and Concrete Composites*, vol. 130, p. 104562, 2022.

[25] V. Mechtcherine *et al.*, "A roadmap for quality control of hardening and hardened printed concrete," *Cement and Concrete Research*, vol. 157, p. 106800, 2022.

[26] R. A. Buswell, W. L. De Silva, S. Z. Jones, and J. Dirrenberger, "3D printing using concrete extrusion: A roadmap for research," *Cement and concrete research*, vol. 112, pp. 37-49, 2018.

[27] E. Lachat, T. Landes, and P. Grussenmeyer, "Investigation of a combined surveying and scanning device: the trimble SX10 scanning total station," *Sensors*, vol. 17, no. 4, p. 730, 2017.

[28] K. M. Lundeen, V. R. Kamat, C. C. Menassa, and W. McGee, "Autonomous motion planning and task execution in geometrically adaptive robotized construction work," *Automation in Construction*, vol. 100, pp. 24-45, 2019.

Integrated Framework for Scaffold Monitoring: From Point Cloud Acquisition to Anomaly Detection

Juhyeon Kim[a], Duho Chung[a], Jeehoon Kim[b], and Hyoungkwan Kim[a]

[a]Department of Civil and Environmental Engineering, Yonsei University, Korea
[b]Fariborz Maseeh Department of Civil, Architectural, and Environmental Engineering, The University of Texas at Austin, USA
E-mail: kah5125@yonsei.ac.kr, jungduho1@yonsei.ac.kr, jeehoon@utexas.edu, and hyoungkwan@yonsei.ac.kr

Abstract –

Scaffolding safety is a critical challenge in the construction industry, often compromised by accidents caused by improper assembly and missing components. This study introduces an automated pipeline for scaffold safety monitoring, integrating robot-assisted point cloud data acquisition, deep learning-based segmentation, scaffold component detection, and safety verification. A quadruped robot equipped with terrestrial LiDAR is utilized to acquire point cloud data, which is then segmented using a deep learning model. The scaffold components are detected through a patch-based algorithm, followed by 3D transformation, clustering, and Delaunay triangulation. The process culminates in the generation of a 3D CAD model, enabling safety evaluations to identify issues such as missing platforms and improperly installed horizontal members. The experimental results demonstrated high performance in CAD model generation and successfully identified structural anomalies in scaffolds, with the proposed method achieving an F1 score of 97.44% in CAD modeling. This approach shows substantial potential for improving scaffolding safety through effective monitoring and early hazard detection.

Keywords –

Scaffold Monitoring; Robotic Data Acquisition; Point Cloud Data; Deep Learning; 3D CAD Modeling; Safety Evaluation

1 Introduction

Scaffolds are vital temporary structures in construction sites, ensuring worker safety while enhancing operational efficiency. However, accidents occurring during the installation of scaffolds remain a significant safety concern, often attributed to missing components or improper assembly [1]. To prevent such accidents, safety guidelines are established to regulate scaffold installation, but ensuring compliance remains a challenge [2]. In practice, adherence to these guidelines relies heavily on manual inspections by supervisors, making it difficult to consistently enforce strict safety protocols. To address these issues, there is a growing demand for the development of automated technologies that can monitor scaffolding conditions and evaluate their safety.

Recent studies have highlighted the potential of using 3D point cloud data for scaffold safety monitoring. With its ability to provide precise spatial information about structures, 3D point cloud data can serve as a valuable tool for detecting and analyzing scaffold components. Luo et al. [3] applied least-squares fitting and random sample consensus (RANSAC) to scaffold point clouds acquired through mobile laser scanning (MLS) to extract scaffold pipes and monitor their deformation. Wang [4] utilized algorithms such as clustering, RANSAC, and m-estimator sample consensus to extract scaffold components from point clouds and verify compliance with installation regulations. Advancements in deep learning have further improved the accuracy of scaffold recognition. Kim et al. [5] developed a scaffold monitoring pipeline that includes data acquisition using a quadruped robot equipped with MLS, deep learning-based point cloud segmentation, and 3D CAD model generation. Kim et al. [6] proposed an algorithm that leverages deep learning techniques to interpret terrestrial laser scanning (TLS) point cloud data and performed regulation checks on scaffold installations.

However, there has been a lack of comprehensive research for scaffold safety monitoring presenting a complete pipeline that encompasses all stages, from data acquisition to 3D model generation and anomaly detection. Furthermore, existing methods for scaffold component detection have often demonstrated reduced performance when applied to different scaffold sites. In previous studies, scaffold component detection methods projected point clouds into 2D before applying techniques such as RANSAC and peak finding. However, this approach results in the loss of 3D geometric

information of the components. A patch-based detection approach could offer a more effective alternative by preserving and utilizing the 3D structure of the components.

This study addresses these limitations by proposing a robust pipeline specifically designed for scaffold monitoring. As shown in Figure 1, the proposed approach integrates efficient point cloud data acquisition using a quadruped robot, deep learning-based point cloud segmentation, scaffold component detection, 3D CAD model generation, and automated anomaly detection into a unified framework. Notably, the method enhances robustness in scaffold component detection through processes such as 3D transformation, patch-based component detection, and Delaunay triangulation for scaffold unit identification.

During the scaffold installation phase, major factors contributing to scaffold accidents include errors in falsework design, incorrect connections, missing components, and improper assembly [7]. However, this study focuses on automatically detecting missing components and improper assembly. The proposed method detects these anomalies by segmenting scaffold point clouds, identifying structural components, and verifying their geometric correctness within the 3D CAD model. By addressing improper assembly and missing components, the framework ensures proactive safety monitoring, enabling early detection of hazardous scaffold conditions. If further refined, the proposed method is expected to contribute significantly to preventing accidents and improving safety conditions in construction environments.

2 Methodology

2.1 Point Cloud Acquisition and Segmentation

As shown in Figure 1(a), a mobile laser scanning system was constructed by mounting a TLS LiDAR (FARO Focus M70) and a mini PC (Intel NUC 11) onto a quadruped robot (Unitree Aliengo). The robot was remotely controlled for movement and point cloud data acquisition. Figure 1(b) illustrates the analysis of the acquired point cloud data using RandLA-Net [8], a deep learning-based 3D segmentation model. The model classifies the point cloud data into seven categories: vertical members, horizontal members, platforms, stairs, rakers, bracings, and background.

2.2 Scaffold Component Detection

In the scaffold component detection step illustrated in Figure 1(c), vertical members, horizontal members, and platforms are extracted from the segmented point clouds. These three components are essential elements of scaffolds and are therefore subject to safety verification; other components are excluded from the assessment as they do not directly affect structural stability.

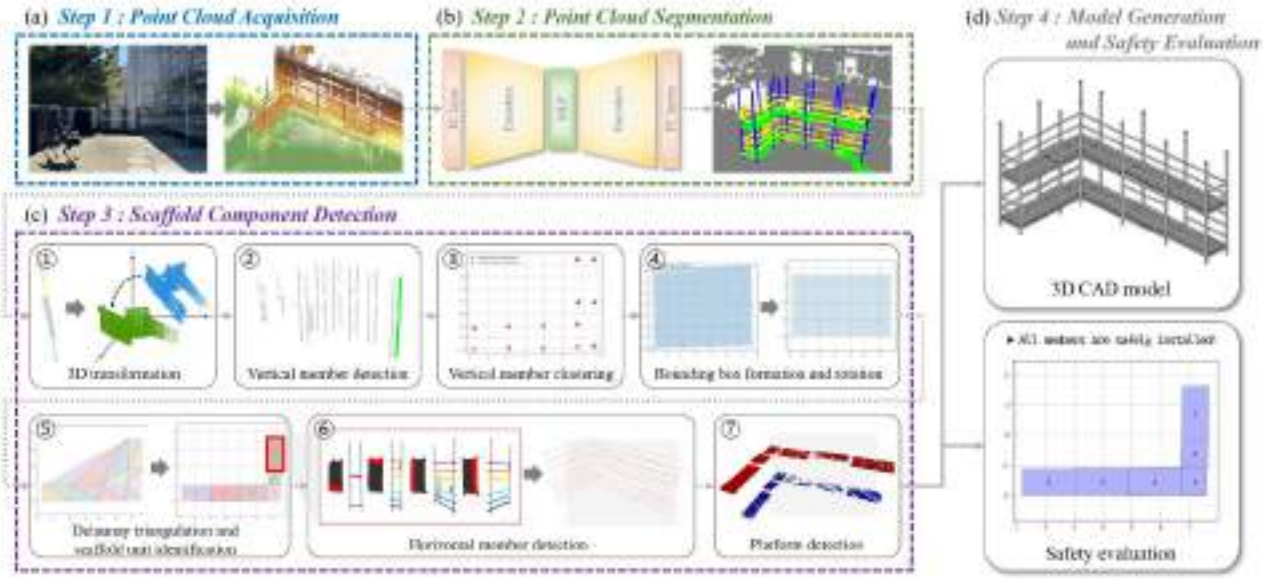


Figure 1. Overview of the scaffold safety monitoring pipeline; (a) robot-assisted point cloud data acquisition, (b) point cloud segmentation using a deep learning-based model, (c) scaffold component detection, and (d) 3D CAD model generation and scaffold safety evaluation.



(a) (b) (c)

Figure 2. 3D patch representation of scaffold components; 3D patches for (a) vertical members, (b) horizontal members, and (c) platforms.

Vertical members, horizontal members, and platforms are identified using a patch-based component matching algorithm proposed by Kim et al. [9]. A 3D patch similar in size and shape to each component is generated and traverses across the target region within the point cloud. If the distance between the points in the patch and the target point cloud is below a predefined threshold, the presence of the component is confirmed. Figures 2(a), 2(b), and 2(c) show green points representing the generated 3D patches for vertical members, horizontal members, and platforms, respectively.

Once the point cloud data are segmented, the largest cluster of vertical member points is identified using density-based spatial clustering of applications with noise (DBSCAN). The entire point cloud is then aligned through a 3D transformation to ensure that the largest vertical member cluster is perpendicular to the XY plane (Figure 1(c), Step 1). This alignment effectively handles tilted scaffold point cloud data. The patch-based component matching algorithm is then applied to the segmented vertical member points to detect vertical components (Figure 1(c), Step 2), using patches measuring 0.1 m × 0.1 m × 4 m. Overlapping vertical members are merged through DBSCAN clustering (Figure 1(c), Step 3). Subsequently, a bounding box is created around the projected vertical member points on the XY plane, and the points are rotated so that the bounding box aligns with the X and Y axes (Figure 1(c), Step 4). Delaunay triangulation is applied using the X and Y coordinates of the vertical members to partition the space. From the partitioned areas, regions forming right-angled rectangles are identified as individual scaffold units (Figure 1(c), Step 5). This scaffold unit identification method is robust and performs effectively regardless of the scaffold's shape or size. Finally, within each scaffold unit, horizontal members and platforms are detected using patches (Figure 1(c), Steps 6 and 7). The dimensions of the patches used to detect these components are determined based on the size of the corresponding scaffold unit, with a patch thickness set at 0.1 m.

2.3 Model Generation and Anomaly Detection

As shown in Figure 1(d), once the scaffold components are detected, their positional information is extracted to generate a 3D CAD model. For example, in Step 6 of Figure 1(c), if a horizontal member is detected using a patch measuring 0.1 m × 0.1 m × 1 m, a pipe with a length of 1 m is created at the corresponding position. The CAD model is automatically generated using the FreeCAD modeler with a Python-based script. The detected vertical and horizontal members are represented as cylindrical CAD objects in FreeCAD, while platforms are modeled as rectangular box-shaped objects. The final CAD model is then exported in FCStd format.

This model can serve as a visual resource for site managers to monitor the installation status of scaffolds and facilitate structural stability analysis. The obtained component information can also be used to directly evaluate the safety of the scaffold. In this study, the presence of platforms and the abnormal lengths of horizontal members in each scaffold unit are identified.

3 Experiments and Results

3.1 Datasets

In this study, the RandLA-Net model pretrained on the dataset from Kim et al. [6] was used for point cloud data segmentation. To validate the proposed method, point cloud data were acquired from an actual scaffold site. Figures 3(a) and 3(b) show the scaffold site where the data were acquired and the quadruped robot used for data acquisition, respectively. A total of 23 point cloud datasets were acquired as the robot moved across the site, and these datasets were registered using FARO SCENE software.



(a) (b)

Figure 3. Data acquisition at the scaffold site; (a) scaffold site where point cloud data were acquired, and (b) quadruped robot used for data acquisition.

To evaluate the performance of the proposed method for scaffold safety verification, components were deliberately installed incorrectly at the scaffold site depicted in Figure 3. As shown in Figure 4(a), the horizontal members of scaffold Unit 8 were installed with an abnormally long length, and their platforms were improperly secured and tilted. Additionally, as shown in Figure 4(b), platforms were missing from scaffold Units 6 and 13.

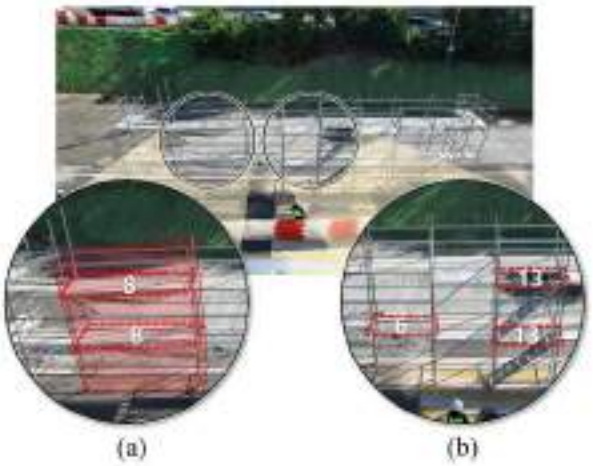
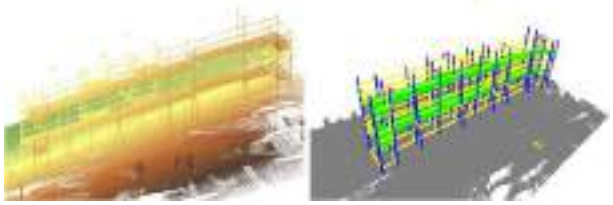


Figure 4. Examples of structural anomalies in scaffolds; (a) improperly installed horizontal members and tilted platforms in scaffold Unit 8, and (b) missing platforms in scaffold Units 6 and 13.

3.2 Experimental Results

3.2.1 3D Semantic Segmentation Results

Figure 5(a) shows the registered point cloud data of scaffolds acquired using the robot. The acquired point cloud data were segmented using a deep learning-based approach, as shown in Figure 5(b). As indicated in Table 1, the segmentation achieved an average F1 score of 91.08%.



(a) (b)

Figure 5. Scaffold point cloud registration and segmentation; (a) registered scaffold point cloud data acquired using the robot, and (b) segmentation results based on the deep learning-based approach.

Table 1. Segmentation performance for scaffold point cloud data.

Scaffold component	Precision	Recall	F1 score
Vertical member	83.14%	99.44%	90.56%
Horizontal member	96.24%	70.46%	81.36%
Platform	88.40%	97.16%	92.57%
Background	99.77%	99.92%	99.84%
Average	91.89%	91.75%	91.08%

3.2.2 Scaffold Component Detection Results

Patch-based detection was performed on the segmented point cloud data to identify points corresponding to vertical members (Figure 6(a)). The detected vertical members were consolidated through clustering (Figure 6(b)) and then rotated to align with the X and Y axes (Figure 6(c)). Using the X and Y coordinates of the vertical members as reference points, Delaunay triangulation was applied to partition the space (Figure 6(d)). From the partitioned areas, scaffold units were estimated (Figure 6(e)). For each scaffold unit, horizontal members and platforms were detected using the patch-based approach (Figures 6(f) and 6(g)).

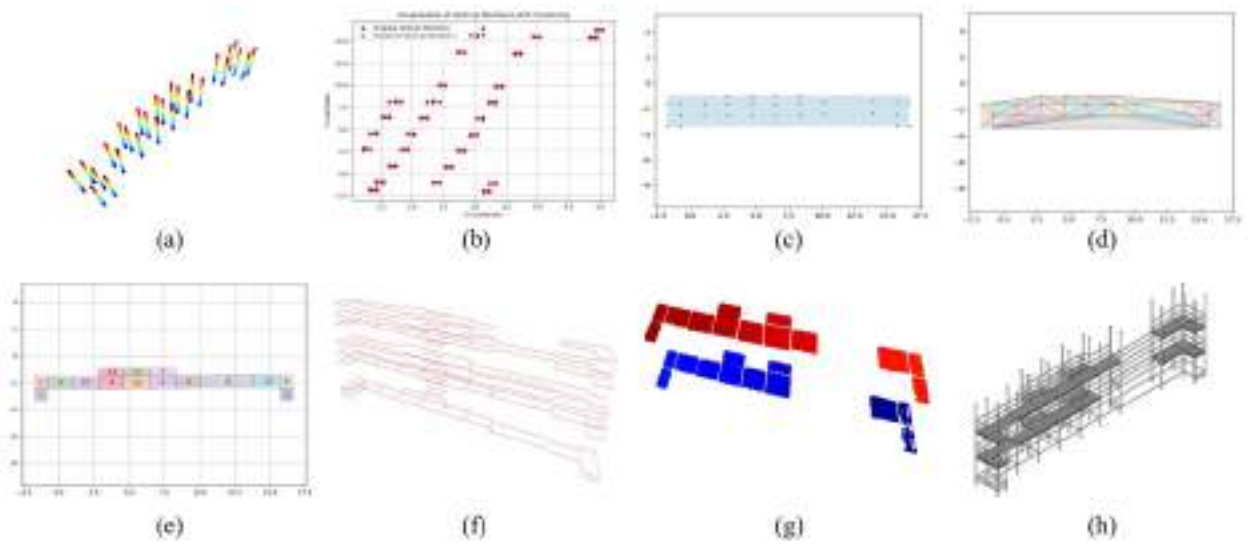


Figure 6. The process of scaffold component detection; (a) patch-based vertical member detection in the segmented point cloud, (b) clustering of detected vertical members, (c) rotation of vertical members to align with the X and Y axes, (d) Delaunay triangulation for space partitioning based on vertical member coordinates, (e) estimation of scaffold units from partitioned areas, (f) horizontal member detection in scaffold units, (g) platform detection in scaffold units, and (h) 3D CAD model generation.

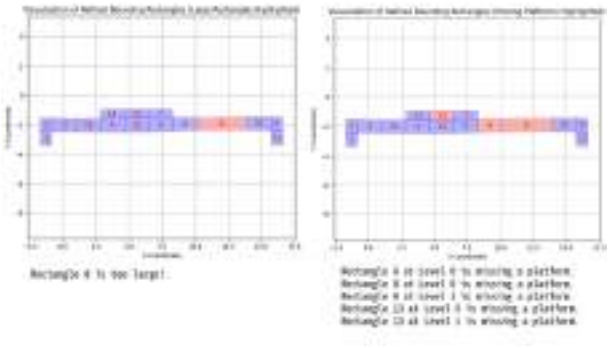
3.2.3 3D CAD Model Generation and Anomaly Detection Results

Figure 6(h) shows the 3D CAD model generated based on the detected scaffold component information. As shown in Table 2, the scaffold components were represented in the 3D CAD model with an F1 score of 97.44%. Although the segmentation of the scaffold point cloud achieved lower performance (Table 1), the modeling accuracy improved by utilizing the raw point cloud data within the scaffold units during the extraction of horizontal members and platforms.

Table 2. Performance of 3D model generation for scaffold components.

Scaffold component	Precision	Recall	F1 score
Vertical member	100%	100%	100%
Horizontal member	95.71%	98.05%	96.87%
Platform	100%	91.30%	95.45%
Average	98.57%	96.45%	97.44%

Figure 7 shows the evaluation and visualization of the scaffold's safety status based on the component information. As shown in Figure 7(a), the algorithm successfully detected the abnormally long horizontal members in scaffold Unit 8. An abnormally long horizontal member must be detected in advance, as it compromises structural stability and increases the risk of collapse. Additionally, Figure 7(b) demonstrates that the algorithm accurately identified the missing platforms in scaffold Units 6 and 13. As shown in Figure 7(b), the platforms in scaffold Unit 8 were abnormally installed at a tilt and were therefore detected as missing.



(a) (b)

Figure 7. Evaluation and visualization of scaffold safety status; (a) detection of abnormally long horizontal members, and (b) identification of missing platforms.

3.3 Discussion

3.3.1 Challenges in CAD Modeling Accuracy

Figure 8(a) highlights the factors that led to the decreased precision in horizontal member representation within the CAD modeling results (Table 2). As shown in the figure, multiple horizontal members at the lowest level were incorrectly detected. This issue arose because noise points acquired near the ground were misclassified as horizontal members. To address this, a preprocessing step to filter out ground noise points could be introduced.

Figure 8(b) illustrates the factors contributing to the decreased recall in CAD modeling of horizontal members. The figure shows Unit 8 of the scaffold, where an abnormally long horizontal members were installed. The installation caused a slight tilt, leading to the failure of the patch-based component matching algorithm. Since the patches move parallel to the X or Y axes, they are less effective in detecting tilted components.

The causes of the reduced recall in platform modeling are demonstrated in the example shown in Figure 8(c). Although platforms were present, they were installed at an angle, similar to the situation in Figure 8(b), and therefore were not detected. However, such improper platform installation poses a significant risk of falls, making the detection of these anomalies in the safety evaluation results meaningful.

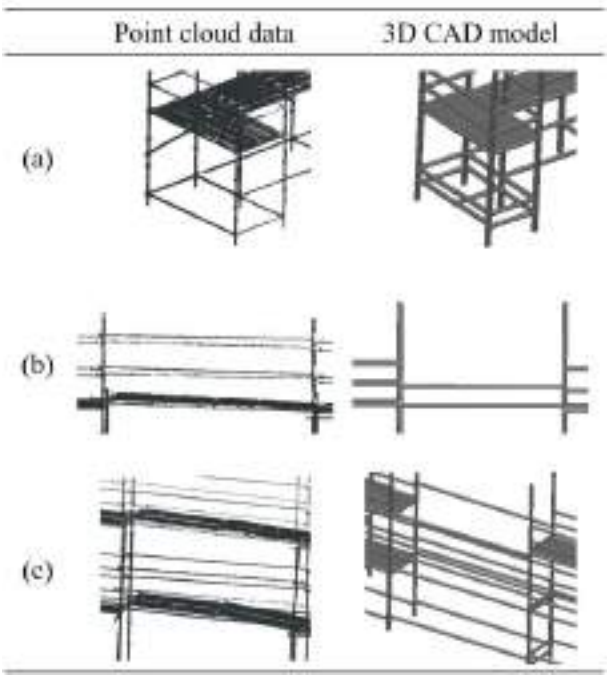


Figure 8. Analysis of false detection issues in CAD modeling for scaffold components; (a) incorrect detection of horizontal members, (b) undetected horizontal members, and (c) undetected platforms.

3.3.2 Future Directions

To achieve comprehensive structural monitoring of scaffolds, it is essential to detect not only improper assembly and missing components but also incorrect connections, such as base plates and joints. However, since these connections are highly localized features, they are difficult to segment using point cloud data alone. Therefore, future studies could explore a sensor fusion approach that integrates image and point cloud data to enable more accurate connection monitoring.

In this study, scaffold safety monitoring was conducted by detecting improper assembly and missing components. The proposed method was validated using abnormally long horizontal members as an example of improper assembly and missing platforms as an example of missing components. However, many additional subcategories exist. For instance, the improper assembly category could include tilted components or elements installed with irregular spacing. Similarly, the missing component category could be expanded to include missing horizontal and vertical members. To enhance the practical applicability of the proposed method, future research will apply and validate the method on scaffolds with a wider range of defects.

Although the proposed method was validated in a testbed environment, future studies should focus on applying and refining the method in real construction sites. This would require acquiring a larger and more diverse scaffold point cloud dataset to further improve 3D segmentation performance. Additionally, to address real-world challenges such as noise and occlusion in point cloud acquisition, techniques such as noise removal and point cloud upsampling should be integrated into the workflow.

4 Conclusion

This study proposed a comprehensive pipeline for scaffold safety monitoring, consisting of robot-assisted point cloud data acquisition, deep learning-based point cloud segmentation, scaffold component detection, CAD model generation, and safety evaluation. Experimental results confirmed that the proposed method effectively detected scaffold components, and generated high-accuracy 3D CAD models. Furthermore, the method successfully detected missing platforms and abnormal horizontal member installations in scaffold units, contributing to the visual assessment of safety conditions. However, the false detection cases of some components suggest the need for further research into noise point cloud preprocessing and the improvement of tilted component detection. For the generalization of the method, further improvement is needed through its application to construction sites with a wider variety of

scaffold shapes and sizes. The proposed method has significant potential for effectively monitoring the installation status and verifying the safety of scaffolds, and is expected to play a key role in improving safety conditions at construction sites.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Ministry of Science and ICT (No.RS-2024-00457308) and the "National R&D Project for Smart Construction Technology (No.23SMIP-A156488-04)" funded by the Korea Agency for Infrastructure Technology Advancement under the Ministry of Land, Infrastructure and Transport, and managed by the Korea Expressway Corporation.

References

[1] Korea Occupational Safety and Health Agency. Practical Guide to Scaffold Safety in Construction Sites. On-line: https://oshri.kosha.or.kr/oshri/publication/commercializationReport.do?mode=view&articleNo=406852&article.offset=0&articleLimit=9, Accessed: 04/12/2024.

[2] Korea Occupational Safety and Health Agency. Safety Operation Guidelines for Scaffolding. On-line: https://www.kosha.or.kr/kosha/data/mediaBankMain.do, Accessed: 11/3/2025.

[3] Luo, R., Zhou, Z., Chu, X., Ma, W., Meng, J. 3D deformation monitoring method for temporary structures based on multi-thread LiDAR point cloud. *Measurement*, 200: 111545, 2022.

[4] Wang, Q. Automatic checks from 3D point cloud data for safety regulation compliance for scaffold work platforms. *Automation in Construction*, 104: 38-51, 2019.

[5] Kim, J., Chung, D., Kim, Y., Kim, H. Deep learning-based 3D reconstruction of scaffolds using a robot dog. *Automation in Construction*, 134: 104092, 2022.

[6] Kim, J., Kim, J., Koo, N., Kim, H. Automating scaffold safety inspections using semantic analysis of 3D point clouds. *Automation in Construction*, 166: 105603, 2024.

[7] Beale, R. G. Scaffold research—A review. *Journal of Constructional Steel Research*, 98: 188-200, 2014.

[8] Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z., ... and Markham, A. RandLA-Net: Efficient semantic segmentation of large-scale point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11108-11117, Seattle, USA, 2020.

[9] Kim, J., Kim, J., Kim, Y., Kim, H. 3D reconstruction of large-scale scaffolds with synthetic data generation and an upsampling adversarial network. *Automation in Construction*, 156: 105108, 2023.

Evaluation of Flight Parameters in UAV-based 3D Reconstruction for Rooftop Infrastructure Assessment

Nick Chodura[1,2], Melissa Greeff[1,3] and Joshua Woods[1,2]

[1]Ingenuity Labs Research Institute, Queen's University, Canada
[2]Department of Civil Engineering, Queen's University, Canada
[3]Department of Electrical and Computer Engineering, Queen's University, Canada

16npc4@queensu.ca, melissa.greeff@queensu.ca, joshua.woods@queensu.ca

Abstract -

Rooftop 3D reconstruction using UAV-based photogrammetry offers a promising solution for infrastructure assessment, but existing methods often require high percentages of image overlap and extended flight times to ensure model accuracy when using autonomous flight paths. This study systematically evaluates key flight parameters—ground sampling distance (GSD) and image overlap—to optimize the 3D reconstruction of complex rooftop infrastructure. Controlled UAV flights were conducted over a multi-segment rooftop at Queen's University using a DJI Phantom 4 Pro V2, with varied GSD and overlap settings. The collected data were processed using Reality Capture software and evaluated against ground truth models generated from UAV-based LiDAR and terrestrial laser scanning (TLS). Experimental results indicate that a GSD range of 0.75–1.26 cm combined with 85% image overlap achieves a high degree of model accuracy, while minimizing images collected and flight time. These findings provide guidance for planning autonomous UAV flight paths for efficient rooftop assessments.

Keywords -

Infrastructure Assessment; Photogrammetry; 3D Reconstruction; Uncrewed Aerial Vehicles; Roof Inspection

1 Introduction

The inspection of rooftops plays a critical role in assessing the structural integrity and maintenance needs of buildings. Traditionally, this task has been carried out by inspectors physically accessing rooftops to identify defects in the building envelope and rooftop equipment. This manual approach is inherently time-consuming, especially for large-scale commercial and industrial structures, and presents significant challenges for visualizing structures.

Advancements in Uncrewed Aerial Vehicle (UAV) technology and photogrammetry have opened new possibilities for building inspections. UAV's equipped with high-resolution cameras can capture images from locations inaccessible to inspectors, enabling comprehensive documentation of structures. These images can be processed using 3D reconstruction techniques to create digital models of the inspected structures. This not only facilitates better visualization of scale and defect locations, but also serves as a historical record for tracking structural changes over time.



Figure 1. LiDAR point cloud of Ellis Hall, Kingston, Ontario with labeled roof sections and height above ground level (AGL).

This study explores the use of autonomous flight paths for generating accurate 3D reconstructions for building inspections. Specifically, it investigates how varying flight parameters, such as image overlap (OL) and ground sampling distance (GSD), affect the accuracy of the reconstructed models. The intended outcome of this research is to be able to optimize these parameters to minimize the flight duration, computation time, and number of image captures - key factors for reducing inspection and processing time in commercial applications - while still maintaining accurate models. There are three primary contributions of this paper:

- Recommendations based on experimental testing for optimal UAV flight parameters in autonomous photogrammetry-based rooftop 3D reconstruction.
- Validation of a UAV-based LiDAR platform as a source of ground truth models using a high-resolution terrestrial laser scanner (TLS).
- Application of cloud-to-cloud comparisons and quantitative pointcloud analysis for infrastructure assessment applications.

2 Related Work

Related work was reviewed from two areas of research: (1) Influence factors for 3D reconstruction, and (2) Quantitative assessment of large-scale mesh models.

2.1 Influence factors for 3D reconstruction

Influence factors for UAV photogrammetry-based 3D reconstruction can be broadly grouped into three main categories: flight parameters, camera settings/image acquisition, and reconstruction methods. While camera settings and reconstruction methods can have significant impacts on reconstruction quality, they were out of the scope of this research and were kept constant throughout the testing. The influence of flight parameters on photogrammetry reconstruction quality has been investigated thoroughly. Several studies have confirmed that flight height, image overlap, ground control point (GCP) quantities/distribution, and georeferencing methods have a significant impact on reconstruction accuracy. GCP's are typically used in larger scale mapping applications such as forestry [1, 2] and glaciology [3], however there has also been research on GCP density and pattern on the reconstruction of dam infrastructure [4]. In this paper's application of building facade 3D reconstruction, there are enough unique features from the building characteristics that an adequate number of tie points can be achieved during feature extraction without the requirement of GCP's.

Flight height has been investigated individually [5, 6], and in combination with overlap and other parameters [7, 8, 7, 9, 10, 11] in various other applications, but less research has been done on how these parameters can influence the reconstruction quality of infrastructure assets. Altman et al. investigated how varying reconstruction methods, number of images, and number of control points (CP's) could influence the reconstruction quality of a church facade [12]. However, all data collected was from terrestrial sources.

2.2 Quantitative assessment of mesh models

Several methods have been established for quantitatively measuring model quality and accuracy. These methods can be broadly categorized into three areas: (1) Reconstruction metrics, (2) Discrete model metrics, and (3) Continuous model metrics. Metrics from the reconstruction process such as number of tie points and number of points in the dense cloud have been shown to correlate with lower control point RMSE values [2]. Discrete metrics consist of control points with known locations that can be used to determine vertical and horizontal error at various locations on the model. These methods can be used to indicate drift or misalignment, but they do not provide continuous information about mesh accuracy across the entire model surface. Continuous model metrics can be divided into measures of mesh visual quality and measures of global geometric accuracy. Mesh visual quality metrics such as local roughness, noise, curvature variation, and topology can be used to quantify the appearance of a mesh surface [13]. Continuous metrics for measuring mesh accuracy are less common in large-scale 3D reconstruction due to the requirement of a ground truth model (GT), which can be difficult to capture at high degrees of coverage and accuracy. Typically, a GT is captured using TLS systems that can produce point clouds at a high degree of accuracy and point density [12, 14, 13, 15].

Chen et al. outlined a workflow for the inspection of a masonry bridge through UAV-based data acquisition, 3D reconstruction, data quality evaluation, and damage detection [15]. They used a TLS point cloud to provide quality metrics for five different potential defects in the reconstructed photogrammetry model. They acknowledge the presence of occlusions in their TLS model due to obstacle-blocking and self-shadowing. These coverage-based defects would be exacerbated for roof reconstructions, where rooftop equipment and obstructions would limit the coverage of the TLS system to an even greater extent. In this paper, a two-stage GT method is proposed. A TLS is used on the rooftop to capture specific areas of high geometric complexity. This is used to validate a UAV-based LiDAR payload that collects data across the entire roof structure.

3 Background

Within the field of photogrammetry, the standard path for autonomous flights is a double-grid boustrophedon pattern using an oblique camera angle [7]. This involves the UAV making back and forth passes at a specified distance in perpendicular directions across the coverage area. This method allows non-horizontal surfaces to be captured from four directions and reduces the chance of occlusions. These paths can be defined with two flight parameters: the image overlap (OL), and (b) the GSD. Because of the oblique camera angle and the varying roof elevations, GSD was used as a variable parameter instead of flight height to better represent the distance between the camera and the target surface.

GSD refers to the resolution of the captured images relative to the real-world dimensions of the scene. GSD depends on the distance between the camera and the object being captured (Z), the camera focal length (f) and the size of a pixel on the sensor (p), which can be determined with the dimensions of the sensor and the image, i.e.,

$$\text{GSD} = \frac{Z}{f}\, p. \tag{1}$$

OL considers the area of shared coverage between two consecutive images. This is expressed as a percentage of

the entire image. Side OL can be determined using the distance between consecutive passes in the same direction. Front OL is defined as the following:

$$\mathrm{OL} = 100 \cdot \frac{\mathrm{GSD}\, w - v\, r_c}{\mathrm{GSD}\, w}, \tag{2}$$

where w is the image width, v is the UAV speed, and r_c is the capture rate of the camera.

4 Methodology

In this study, the approach in assessing flight parameters for 3D photogrammetry has three steps: 1. Data Collection: collected image data from 9 flights of the building in Figure 1 using a DJI Phantom 4 Pro V2. The research team collected UAV-based LiDAR data using a DJI Matrice 350 RTK, and TLS data using a FARO Focus. 2. Model Generation: generated 3D reconstructions using photogrammetry software Reality Capture [16]. LiDAR models are processed in DJI Terra [17], and TLS models are aligned using SCENE point cloud software [18]. 3. Model Evaluation: All models were trimmed, aligned, and compared using 3D point cloud processing software CloudCompare [19].

4.1 Data collection

Ellis Hall on the campus of Queen's University, Kingston, Ontario (latitude: 44.226, longitude: -76.496) was chosen as the test site for this experiment. The building features five distinct roof segments at AGL elevations ranging from 1.71 m to 14.16 m. The roof sections contain a variety of rooftop features, including an observatory dome, various HVAC equipment, fencing and barrier walls, and experimental testing equipment. This variance in elevation and subject matter provides unique levels of obstruction, lighting conditions, GSD, and feature complexity for the experiment.

Images for the photogrammetry reconstructions were collected using a DJI Phantom 4 Pro V2, a consumer multirotor UAV with an on-board 20M camera featuring an 84° field of view and an 8.8 mm focal length, capable of capturing 4k images. The goal of this research was to assess the photogrammetry reconstruction quality of fine details and unique rooftop features in real-world conditions. As such, a process was needed that would provide a continuous, global metric of model accuracy across the entire mesh surface. To do this, A 'ground truth' model (GT) was needed to assess the accuracy of the generated reconstructions. A LiDAR scan of the building was performed using a DJI Matrice 350 RTK, equipped with a Zenmuse L2 LiDAR sensor. The L2 features a frame-based LiDAR, internal IMU system, and an RGB mapping camera [20]. A D-RTK 2 Base Station and on-board RTK Module provide Real-time Kinematic Processing (RTK) for accurate localization during the flight, and Post Processing Kinematics (PPK) was done using DJI Terra [17]. The L2 specifications report a ranging accuracy of 2 cm at 150 m scanning distance, which is reported to improve at closer ranges. This paper proposes the use of a point cloud collected from the Zenmuse L2 as a GT model to provide quantitative evaluations of accuracy for the photogrammetry-based 3D reconstructions.

The UAV-based LiDAR point cloud was validated using the Faro Focus TLS platform. The TLS was placed in 14 locations across Roof Section B and Roof Section D. The Faro Focus produces a point cloud with 2 mm accuracy [21]. Ground control spheres were used to register the scans in Faro's SCENE 3D point cloud software. Three test segments of the roof were chosen to compare the TLS and LiDAR point cloud's. Areas of geometric complexity were chosen to highlight potential deficiencies in the UAV-based LiDAR's effectiveness as a GT model. A cloud-to-cloud (C2C) comparison was conducted between the two point clouds at each test segment. One of these comparisons can be seen in Figure 2. The test segments averaged an 85% F-score at a threshold value of 2 cm.



Figure 2. Recall scalar field for a cloud-to-cloud comparison of the UAV-based LiDAR and TLS point clouds.

4.2 Model generation

Photogrammetry models were generated using Reality Capture photogrammetry software [16]. Reality Capture uses a photogrammetry pipeline consisting of feature extract, feature matching, and a Structure-from-Motion (SfM) algorithm to establish a set of tie points. A meshing stage generates a geometric surface for the scene using a triangulation algorithm. Finally, a Coloring/Texturing stage estimates a color for each vertex based on the input images.

Reality Capture's settings were kept to their default values with the exception of the following alignment parameters: the maximum features per megapixel were increased from 10,000 to 20,000 and the maximum features per image was increased from 40,000 to 100,000. The prior hardness of position and the prior hardness of orientation were decreased from 1.0 to 0.1. Preselector features were

Table 1. Variable Flight Parameters, Flight Data, and Model Data

	Variable Parameters		Flight Data		Model Data	
Flight	GSD (cm)	OL (%)	Duration (min)	Captures	Compute Time (min)	Tie Points
Flight 1	0.51 - 1.01	60	4	42	55 sec	109 270
Flight 2	0.51 - 1.01	70	6.5	76	1 min 26 sec	637 177
Flight 3	0.51 - 1.01	80	9.5	198	2 min 55 sec	2 242 477
Flight 4	0.51 - 1.01	85	15.75	249	5 min 7 sec	1 148 312
Flight 5	0.75 - 1.26	80	7.5	96	1 min 20 sec	406 087
Flight 6	0.75 - 1.26	85	13.5	233	3 min 31 sec	4 571 844
Flight 7	0.75 - 1.26	90	32.5	482	11 min 31 sec	11 434 884
Flight 8	0.98 - 1.49	85	8.25	171	3 min 11 sec	3 073 278
Flight 9	0.98 - 1.49	90	23.5	360	6 min 15 sec	8 108 259

increased from 10,000 to 30,000 and the detector sensitivity was increased from medium to high.

4.3 Model evaluation

The models were evaluated using C2C distances between the photogrammetry-based models and the LiDAR point cloud. All point cloud analysis was done using CloudCompare [19]. The mesh models were sampled at 40,000 points per square meter. This was done to produce a dense point cloud that could accurately represent the mesh surface. All point clouds were then spatially subsampled with a minimum distance of 5 mm to ensure consistent uniform-density between models. The value of 5 mm was selected based on previous work on large-scale photogrammetry models [14].

Models were then trimmed of surrounding surfaces and divided into five rooftop sections. The models were roughly aligned using equivalent point pairs, before a fine alignment was done using an iterative closest point (ICP) algorithm. C2C distances were calculated using a quadratic height function as a local surface model, with a nearest neighbour radius of 10 cm. Three metrics were produced, following procedures in [14].

Precision is the measure of how accurately the points of the model align with the GT, represented as a percentage of points under a threshold distance d, and is defined as:

$$P(d) = \frac{100}{|R|} \sum_{r \in R} [\mathrm{e_{rg}} < d], \tag{3}$$

where R is the set of all points r in the reconstruction. e_{rg} is the distance between a given point r and its nearest point g in the GT point cloud.

Recall is a measure of the completeness of a reconstruction relative to a GT model. Recall is also represented as a percentage and is defined as:

$$R(d) = \frac{100}{|G|} \sum_{g \in G} [\mathrm{e_{gr}} < d], \tag{4}$$

where G is the set of all points g in the GT point cloud.

The third metric combines these two measures to provide a robust qualifier for model quality that considers accuracy and completeness. The F-Score, or harmonic mean, is defined as:

$$F(d) = \frac{2P(d)R(d)}{P(d) + R(d)}. \tag{5}$$

5 Results

Based on qualitative observations, it was observed that the flights at 60% and 70% OL produced models with large distortions, as seen in Figure 4a and F-Scores in Table 2. At these overlap percentages, even some flat rooftop surfaces were not accurately represented by the model. Because of this observation, these OL percentages where not used in future testing. Above 80% OL, models began to better represent the structure, but there were still variations in the reconstruction accuracy of fine details and complex geometry, as seen in Figure 3.

Some clear trends were observed from the C2C quantitative analysis conducted. The flights at 0.75 - 1.26 cm GSD generally outperformed the other flights, even when compared with flights that had significantly more images captured. This was observed for most roof sections: Flight 6 or Flight 7 had the highest F-score for four of the five roof sections. Another interesting observation is that Flight 5, having captured only 96 images, shared the third highest average rank with Flight 4 and Flight 9, which had 249 and 360 captures, respectively. It also had the lowest overlap of the three flights. The analysis also highlights the reduction in model quality for the lowest rooftop sections, D and E, at a higher GSD. At the same OL, Flight 9 had an F-Score that was 12.3% lower for Roof E and 15.9% lower for Roof D, when compared with Flight 7.

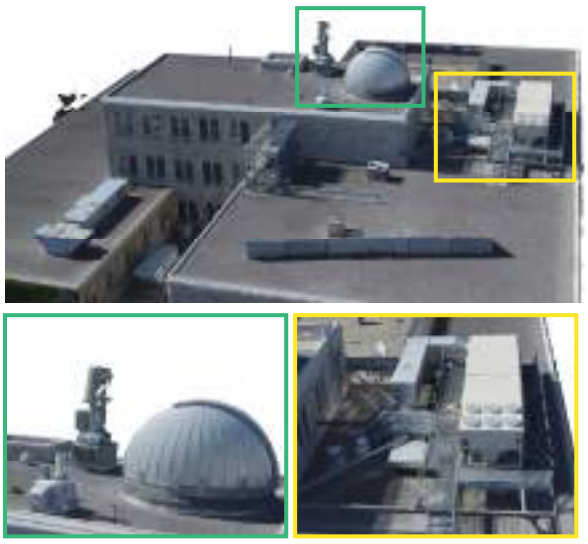
The precision and recall were also plotted across different threshold values for each of the roof sections. In Figure 4 and Figure 5, an example of this is shown for roof sections B, C, and D. It is clear the flights at 0.75 - 1.26 cm GSD (gold, turquoise, and pink on the diagram) consistently perform well across a range of threshold values. These plots also illustrate the under-performance of the 0.98 - 1.49 cm GSD flights on Roof D, shown in lime green and purple on the graphs in Figure 5 and Figure 6.

Table 2. F-Scores at 4cm Threshold

	Flight 1	Flight 2	Flight 3	Flight 4	Flight 5	Flight 6	Flight 7	Flight 8	Flight 9
Roof A	5.64	31.33	92.06	92.79	92.62	**95.82**	**95.64**	94.89	95.51
Roof B	43.55	77.18	90.80	90.56	91.04	**92.39**	**92.19**	90.98	91.56
Roof C	86.78	95.70	98.16	98.12	97.27	**98.17**	**98.80**	97.43	98.01
Roof D	12.63	85.19	84.56	86.36	**89.10**	**90.16**	88.53	70.46	84.36
Roof E	78.29	80.08	78.43	**88.92**	**86.22**	86.03	82.24	76.67	84.01
Mean	29.16	69.13	88.54	91.26	91.18	**92.41**	**91.30**	85.41	90.51
Rank	8.8	7.0	6.0	4.2	4.2	**1.6**	**2.6**	6.4	4.2



(a) Flight 4: 0.51 - 1.01 cm GSD, 85% OL



(b) Flight 6: 0.75 - 1.26 cm GSD, 85% OL

Figure 3. At the same OL, Flight 6 produces a more accurate model than Flight 4. Thin-walled surfaces feature less occlusions, seen in the exhaust ventilator on the left, and complex geometry was represented better, seen in the walkway on the right.

These figures also illustrate the separation in quality for more complex roof sections. Roof C, a section with minimal geometric complexity, features significantly less variation in F-Score across different flights. With the introduction of complex geometry (Roof B) and higher degrees of obstruction (Roof D), the flights begin to display a greater range in quality.

6 Discussion

The models produced with 90% OL provided diminishing returns, or even decreasing quality, over the models at 85% OL. It was also evident that the 0.75 – 1.26 cm GSD flights produced more accurate reconstructions relative to the number of images collected, particularly for the roof sections at higher elevations. For general roof surfaces that have a moderate amount of geometric complexity and limited obstructions, these flight parameters will provide high-quality mesh models that accurately represent the real-world structure.

This paper provides three additional findings with regard to autonomous flight paths for rooftop 3D reconstruction:

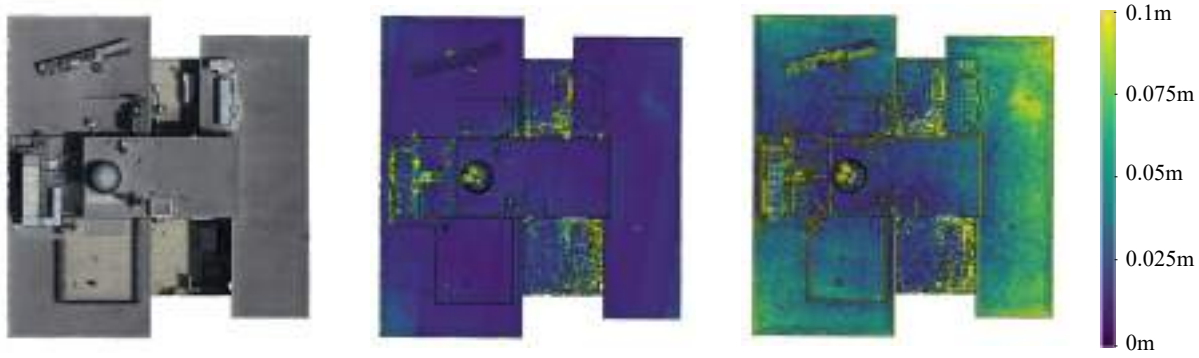
- For roof sections that don't exhibit complex geometry, lower OL can still yield effective results while reducing flight time, image captures, and computation time. This can be seen in Roof C, where 70% OL produced a model with an F-score greater than 95% at a 4cm threshold.

- Thin-walled surfaces and fine details (e.g. chain-link fences, metal bar grating) were not captured accurately even at optimal OL and GSD values. Supplementary images would need to be captured manually around these features for accurate reconstruction.

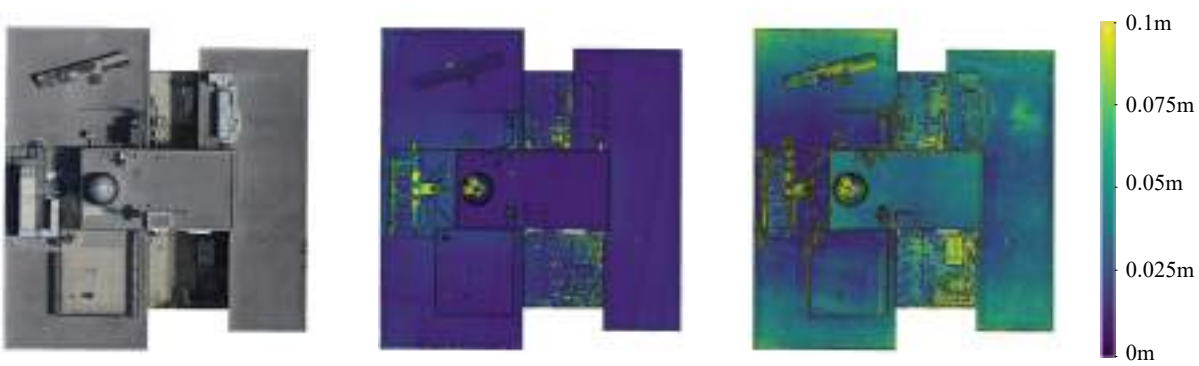
- Lower roof sections exhibited a significant decrease in F-score relative to the other roof sections, regardless of GSD and OL parameters. This could be due to obstruction from other building sections and poorer lighting conditions.



(a) Flight 2: 0.51 - 1.01 cm GSD, 70% OL



(b) Flight 4: 0.51 - 1.01 cm GSD, 85% OL



(c) Flight 7: 0.75 - 1.26 cm GSD, 90% OL

Figure 4. RGB (left), precision (middle), and recall (right) scalar fields for three of the test flights. The model from Flight 2 data produced large regions of high error (>10 cm), and was unable to recreate simple building geometry. With higher overlap percentages, the error becomes primarily concentrated around regions of high detail, and the lower roof sections with poorer lighting and greater obstruction. Note that the rectangular area of high error on the observatory dome was due to rotation of the telescope prior to collection of the GT, however this is consistent across all test flights.

7 Future Work

The number of flights was kept relatively low due to the time-intensive data collection and C2C analysis processes. In future testing, a larger sample size of flights should be compared. Other flight parameters have also been shown to have an effect on reconstruction quality, including ground control points, camera angle, image parameters, camera settings, and georeferencing methods [7]. Future testing could investigate how these parameters may influence model quality in the context of building envelope 3D reconstruction. With added parameters and test flights, the main effects and interaction effects of these parameters can be observed using a statistic model.

The optimal parameters were also tested using other

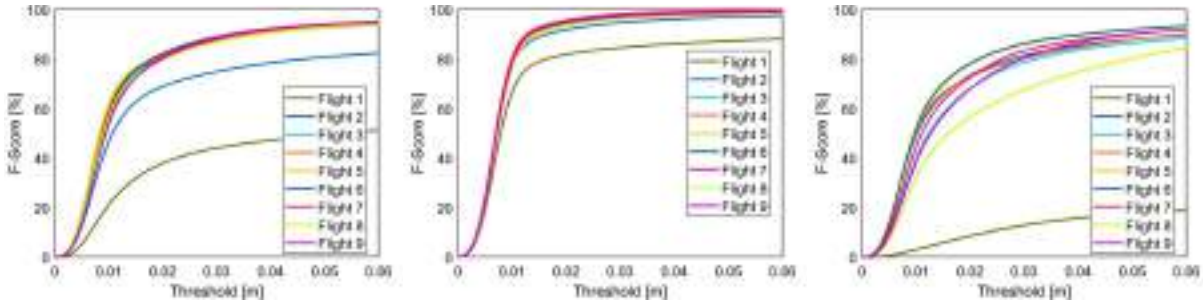


Figure 5. Precision at threshold values up to 6 cm for roof sections B, C, and D (left to right). For rooftops with less features, such as Roof C, the models remain fairly similar across different threshold values. As rooftop complexity is introduced, models begin to spread out.

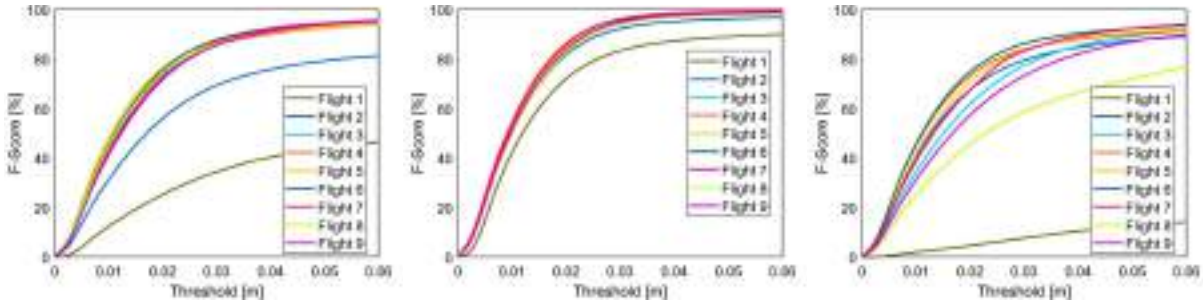


Figure 6. Recall at threshold values up to 6 cm for roof sections B, C, and D (left to right).

UAV platforms including the DJI Mavic 3 Enterprise and the DJI Mini 4 Pro. Flights were also conducted on other buildings, including residential and commercial structures. Future testing could involve quantitative evaluation for different UAV platforms and types of buildings.

8 Conclusion

This paper investigated how varying flight parameters can influence the reconstruction accuracy of rooftop 3D models. Through a series of field tests, the study demonstrated that GSD values of 0.75 - 1.26 cm and an overlap of 85% provided the most efficient flight path while maintaining a high degree of reconstruction accuracy. Using a rooftop TLS configuration, UAV-based LiDAR has been validated as an effective method for providing accurate rooftop point clouds that can be used as suitable ground-truth models for photogrammetry.

9 Acknowledgments

This research was supported by Maintenance Drone Company and Mitacs through the Mitacs Accelerate Program. Additional support was provided by the Ingenuity Labs Research Institute at Queen's University.

References

[1] Jonathan P. Dandois, Marc Olano, and Erle C. Ellis. Optimal altitude, overlap, and weather conditions for computer vision uav estimates of forest structure. *Remote Sensing*, 7(10):13895–13920, 2015. ISSN 2072-4292. doi:10.3390/rs71013895. URL https://www.mdpi.com/2072-4292/7/10/13895.

[2] Erich Seifert, Stefan Seifert, Holger Vogt, David Drew, Jan van Aardt, Anton Kunneke, and Thomas Seifert. Influence of drone altitude, image overlap, and optical sensor resolution on multi-view reconstruction of forest images. *Remote Sensing*, 11(10), 2019. ISSN 2072-4292. doi:10.3390/rs11101252. URL https://www.mdpi.com/2072-4292/11/10/1252.

[3] Saskia Gindraux, Ruedi Boesch, and Daniel Farinotti. Accuracy assessment of digital surface models from unmanned aerial vehicles' imagery on glaciers. *Remote Sensing*, 9(2), 2017. ISSN 2072-4292. doi:10.3390/rs9020186. URL https://www.mdpi.com/2072-4292/9/2/186.

[4] Sizeng Zhao, Fei Kang, Junjie Li, and Chuanbo Ma. Structural health monitoring and inspection of dams based on uav photogrammetry with

image 3d reconstruction. *Automation in Construction*, 130:103832, 2021. ISSN 0926-5805. doi:https://doi.org/10.1016/j.autcon.2021.103832. URL https://www.sciencedirect.com/science/article/pii/S0926580521002831.

[5] N. Anders, M. Smith, J. Suomalainen, et al. Impact of flight altitude and cover orientation on digital surface model (dsm) accuracy for flood damage assessment in murcia (spain) using a fixed-wing uav. *Earth Science Informatics*, 13(3):391–404, June 2020. doi:10.1007/s12145-019-00427-7. URL https://doi.org/10.1007/s12145-019-00427-7.

[6] Benjamin T. Fraser and Russell G. Congalton. Issues in unmanned aerial systems (uas) data collection of complex forest environments. *Remote Sensing*, 10(6), 2018. ISSN 2072-4292. doi:10.3390/rs10060908. URL https://www.mdpi.com/2072-4292/10/6/908.

[7] Yajie Liu, Kevin Han, and William Rasdorf. Assessment and prediction of impact of flight configuration factors on uas-based photogrammetric survey accuracy. *Remote Sensing*, 14(16), 2022. ISSN 2072-4292. doi:10.3390/rs14164119. URL https://www.mdpi.com/2072-4292/14/16/4119.

[8] Darío Domingo, Hans Ole Ørka, Erik Næsset, Daud Kachamba, and Terje Gobakken. Effects of uav image resolution, camera type, and image overlap on accuracy of biomass predictions in a tropical woodland. *Remote Sensing*, 11(8), 2019. ISSN 2072-4292. doi:10.3390/rs11080948. URL https://www.mdpi.com/2072-4292/11/8/948.

[9] Sangku Lee, Jeongha Park, Eunsoo Choi, and Dongkyun Kim. Factors influencing the accuracy of shallow snow depth measured using uav-based photogrammetry. *Remote Sensing*, 13(4), 2021. ISSN 2072-4292. doi:10.3390/rs13040828. URL https://www.mdpi.com/2072-4292/13/4/828.

[10] Taylor Zimmerman, Karine Jansen, and Jon Miller. Analysis of uas flight altitude and ground control point parameters on dem accuracy along a complex, developed coastline. *Remote Sensing*, 12(14), 2020. ISSN 2072-4292. doi:10.3390/rs12142305. URL https://www.mdpi.com/2072-4292/12/14/2305.

[11] Xi Wang, Julia C. Chen, and Gabriel B. Dadi. *Factors Influencing Measurement Accuracy of Unmanned Aerial Systems (UAS) and Photogrammetry in Construction Earthwork*, pages 408–414. doi:10.1061/9780784482438.052. URL https://ascelibrary.org/doi/abs/10.1061/9780784482438.052.

[12] S. Altman, W. Xiao, and B. Grayson. Evaluation of low-cost terrestrial photogrammetry for 3d reconstruction of complex buildings. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2/W4:199–206, 2017. doi:10.5194/isprs-annals-IV-2-W4-199-2017. URL https://isprs-annals.copernicus.org/articles/IV-2-W4/199/2017/.

[13] Erica Nocerino, Elisavet Konstantina Stathopoulou, Simone Rigon, and Fabio Remondino. Surface reconstruction assessment in photogrammetric applications. *Sensors*, 20(20), 2020. ISSN 1424-8220. doi:10.3390/s20205863. URL https://www.mdpi.com/1424-8220/20/20/5863.

[14] Arno Knapitsch, Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics*, 36(4), 2017.

[15] Siyuan Chen, Debra F. Laefer, Eleni Mangina, S. M. Iman Zolanvari, and Jonathan Byrne. Uav bridge inspection through evaluated 3d reconstructions. *Journal of Bridge Engineering*, 24(4):05019001, 2019. doi:10.1061/(ASCE)BE.1943-5592.0001343. URL https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29BE.1943-5592.0001343.

[16] Capturing Reality. Realitycapture, 2025. URL https://www.capturingreality.com. Version 2025.1.

[17] DJI. Dji terra - mapping and 3d modeling software, 2025. URL https://www.dji.com/terra.

[18] Inc. FARO Technologies. Scene - 3d point cloud capture and processing software, 2025. URL https://www.faro.com/en/Products/Software/SCENE-Software.

[19] CloudCompare Development Team. Cloudcompare - 3d point cloud and mesh processing software, 2025. URL https://www.cloudcompare.org. Version 2.12.

[20] DJI. Zenmuse l2 specifications, 2025. URL https://enterprise.dji.com/zenmuse-l2/specs.

[21] Inc. FARO Technologies. Faro knowledge base, 2025. URL https://knowledge.faro.com/Hardware/Focus/Focus.

Algorithm development for automatic detection of progressive damage in tunnel cross-sectional geometry

Christopher Nuñez[1] and Marck Regalado[2]

[a]Faculty of Civil Engineering, National University of Engineering, Perú
E-mail : christopher.nunez.v@uni.pe, marck.regalado.e@uni.pe

Abstract

Accurate assessment of tunnel conditions is vital for ensuring long-term safety and structural integrity, particularly in weak rock masses with complex geological conditions. This study introduces a methodology to analyze tunnel deformation and potential damage using point cloud data acquired in 2019, 2022, and 2024 through terrestrial laser scanning (TLS). The datasets were processed to remove outliers, irrelevant elements, and internal features, resulting in a clean representation of the tunnel's geometry. A novel algorithm integrates point cloud segmentation, centerline alignment, and projection techniques to derive tunnel profiles and monitor structural changes over time. The results demonstrate TLS's effectiveness in capturing high-resolution data for ovalization monitoring, revealing a consistent reduction in cross-sectional area throughout the tunnel and a rightward shift in the 2024 profiles. This approach underscores the potential of TLS for precise structural monitoring and deformation analysis in challenging environments.

Keywords –

Tunnel; Point cloud; Open3D

1 Introduction

Tunnel stability continues to be a major challenge in civil engineering. Various technologies have been applied to optimize monitoring processes, guaranteeing stability and minimizing the probability of tunnel failure throughout its useful life [1], [2]. Failures threatening the safety of tunnels constructed in soft or weak rock masses often occur throughout their long-term service periods [1], [3], [4]. The deformation and failure of rock masses are primarily influenced by discontinuities, which play a crucial role in determining their mechanical properties and overall stability [5]. Additionally, the mechanical behavior of tunnel structures is shaped by a combination of internal and external factors. While historical structural performance is an internal factor, external loads, such as temperature variations and water pressure, significantly affect the long-term properties and stability of the structure [3].

Therefore, ensuring the safe operation of tunnels is essential, as tunnel structures are highly susceptible to disasters due to the complex geological conditions in which they are situated [3]. However, traditional long-term monitoring methods are expensive, time-consuming, and highly dependent on operator subjectivity. These methods usually rely on one-dimensional random sampling, focusing only on limited areas, which leads to the omission of critical details [6], [7]. Additionally, traditional measurement approaches require surveyors to directly access hazardous geological areas. Despite these efforts, there are regions that cannot be manually measured, and the inherent risks involved in such measurements remain a major concern [8].

To address these challenges, many companies have invested in remote sensing technologies, including light detection and ranging (LiDAR), terrestrial laser scanning (TLS), digital photogrammetry, and unmanned aerial vehicles (UAVs) with 3D reconstruction capabilities [5]. For tunnel applications, laser scanning and photogrammetry are commonly employed, as both offer distinct technical advantages and potential value [6]. Firstly, these technologies enable rapid scanning of rock masses, such as temporarily exposed tunnel faces, significantly reducing survey time. Secondly, they allow geologists to perform measurements without physically approaching the tunnel face, thereby enhancing the safety of survey operations [5]. Additionally, sensor-based continuous monitoring techniques and analysis models are increasingly utilized to ensure tunnel functionality and maintain structural integrity over time [2].

Terrestrial Laser Scanning (TLS) offers high resolution and accuracy, making it a reliable method for data acquisition, often surpassing photogrammetry in precision while maintaining flexible measurement distances adaptable to various scenarios. Unlike photogrammetry, TLS does not require a light source, making it particularly suitable for dark underground environments such as tunnels. Its applications extend beyond tunnel monitoring to include stratigraphic

modeling, archaeological excavation, and ancient site exploration [5], [8]. In tunnel engineering, TLS is frequently employed for ovalization monitoring of shield tunnels, with results aligning closely with structural calculations. Additionally, advancements in laser scanning technology have enabled the development of 3D LiDAR point clouds, which have proven effective in monitoring tunnel displacements [9].

After obtaining the point cloud, it is essential to process the data to identify relevant features such as discontinuities. Automated algorithms typically rely on clustering techniques or region-growing approaches, operating on spherical local support regions and utilizing normal vectors to detect features. Clustering methods, such as k-means, fuzzy k-means, ISODATA, DBSCAN, and CFSFDP, often use parameters like dip angle and dip direction derived from normal vectors for discontinuity identification. Additionally, model fitting techniques, including the Hough Transform (HT), Random Sample Consensus (RANSAC), and Principal Component Analysis (PCA), are employed to construct geometric models and estimate optimal plane parameters, enabling precise feature extraction and analysis [5], [10].

In this research, several challenges were identified within the dataset, including outliers, the presence of elements inside the tunnel, and irregularities in the tunnel's shape. To address these issues, we propose an algorithm designed to analyze tunnel damage using point cloud data captured by a terrestrial laser scanner (TLS) over three different years.

The structure of this paper is as follows: Section 2 discusses the materials and methods used to acquire and process the point cloud datasets. Section 3 presents the results, including the extraction of tunnel profiles for each dataset and an analysis of their changes over time. Finally, Section 4 highlights the scientific contributions and conclusions of this study.

2 Materials and Methods

This study focused on a tunnel with vertical walls and an arched roof, located in the Peruvian jungle (see Fig. 1). The tunnel serves as a critical passageway for gas transmission pipelines, facilitating their traversal across mountainous terrain. The structure has an approximate height of 2.10m and a width of 2.50m. The internal profile of the tunnel is reinforced with a layer of shotcrete, providing structural stability and durability.



Figure 1. External view of the tunnel located in Peru.

2.1 Data Acquisition

The tunnel data were collected during three different years: 2019, 2022, and 2024. For the datasets from 2019 and 2022, a Leica C10 laser scanner was employed, while the most recent data from 2024 was acquired using a Trimble X7 3D laser scanner. Table 1 summarizes the technical specifications of the Trimble X7 scanner.

Table 1 Features of the Trimble X7 3D Laser Scanner

Features	Specification
Scan Speed	26,600 points/sec
Scanning Range	350m - 600m
Accuracy	2.5mm at 100m
Noise tolerance	1.5mm - 2.5mm

For data acquisition planning, given the tunnel's 300-meter length, data capture stations were placed every 5 meters. The laser scanner was positioned at a height of 1.8 meters to ensure the capture of all possible points, including those behind pipes. Although the tunnel had sufficient natural lighting, artificial lighting was used to enhance data quality (see Fig. 2).



Figure 2. Data acquisition process inside the tunnel.

After completing the data acquisition process, which required a single day of work with a team of three individuals, the collected data was processed using Trimble software to export the point cloud in .ptl format.

2.2 Data Processing

After data acquisition, the collected data was processed in Python due to its extensive libraries for point cloud data handling and analysis. The primary library utilized was Open3D, an open-source tool for 3D data processing [11]. This library provides robust functionality for working with various point cloud formats, making it a versatile choice for this study. The parameters for data processing were based on the tunnel's measurements, including its height and width, to accurately filter outlier points.

2.2.1 Importing data

The data from 2019 and 2022 were provided as .txt files. Using NumPy, these files were transformed into point clouds containing only positional data. For the 2024 dataset, it was necessary to identify the specific format of the point cloud before processing, in this was .pts.

Open3D provides two primary methods for creating a point cloud: tensors or geometry-based methods. In this study, the geometry-based approach was chosen as it offers a wider range of features for controlling and processing the point cloud effectively. Figure 3 illustrates the 2024 point cloud, visualized using Open3D.

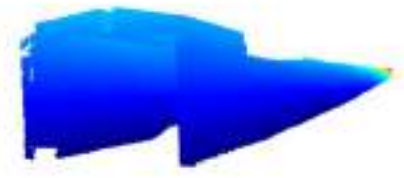


Figure 3. Visualization of the 2024-point cloud data processed in Open3D.

Finally, each dataset was uniformly downsampled using a specified voxel size. The points were grouped into voxels of 0.1m size, and each occupied voxel produced a single point by averaging all the points within it. Table 2 presents the results of the data import process.

Table 2 Summary of the Point Cloud Data

Year	Total points	Filtered Points	Retention Percentage
2019	70124628	348248	0.5%
2022	2259802	231868	10.26%
2024	4722696	208109	4.41%

Note: "Retention Percentage" represents the ratio of filtered points to total points, expressed as a percentage.

2.2.2 Cleaning data

As shown in Figure 3, the scanner captured points outside the tunnel, which are irrelevant to the objectives of this study. Similarly, within the tunnel, the presence of pipes required their removal, as the analysis focused solely on the contour profile. Additionally, several outlier points were identified during the scanning process. Figure 4 illustrates the interior of the tunnel, where the pipes and floor were captured.



Figure 4. The view of the tunnel's interior is represented in the point cloud.

To address these issues, a centerline representing the tunnel's overall geometry was first extracted. Subsequently, all points within a 1-meter distance from the centerline were removed. This process effectively eliminated the floor and the contour of the pipes from the point cloud (see Fig. 5). Next, the KDTree algorithm from the PyntCloud library was applied to identify neighboring points for each data point. Using this neighborhood information, the average distance from each point to its neighbors was computed. Points with an average distance greater than 0.20 meters were classified as outliers, and the corresponding neighboring points were eliminated. This iterative filtering process resulted in a clean point cloud that accurately represents the tunnel's contour shape.

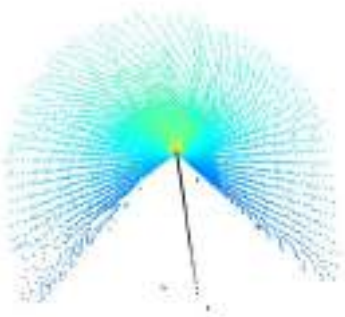


Figure 5. Point cloud after removing the floor and pipe shapes. The black line represents the centerline of the tunnel.

Then, Table 3 shows the number of points in each point cloud before and after the cleaning process, along with the corresponding percentage, which represents the ratio of points remaining after cleaning to the total points before cleaning.

Table 3. Summary of the point cloud data

Year	Downsampled points	Points after cleaning	Percentage
2019	348248	215646	61.9%
2022	231868	192832	83.2%
2024	208109	172915	83.1%

Since the structural movements are measured in millimeters, the focus is on removing outlier points generated near the tunnel floor. However, eliminating noise from the point cloud may impact the analysis of each section, as certain data points could be mistakenly discarded.

3 Results

3.1 Profiles

In this process, the centerline created during the data processing phase was divided into 2600 points. Using the direction of the line, a plane perpendicular to the line was created for each pair of adjacent points. This plane defined a region surrounding a specific number of points from each point cloud (see Fig. 6). With a voxel size of 0.10m, the minimum distance between each plane was also set to 0.10m.



Fig. 6. Points in the surrounding region are defined by the perpendicular planes.

After extracting the points from each region, the projection of each point onto the plane perpendicular to the centerline was calculated. Using the Matplotlib library, a plane was generated with all these points connected by a line, forming the tunnel’s contour (see Fig 7.).

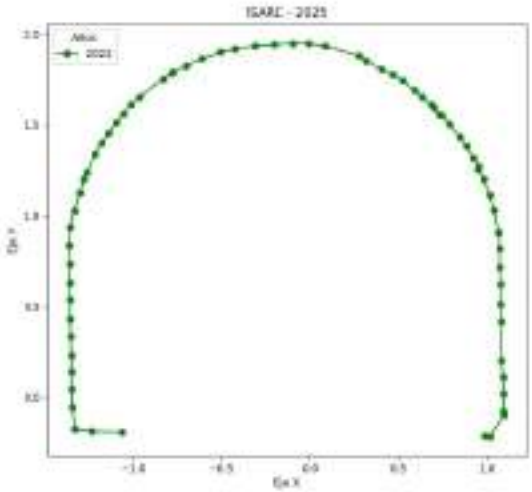


Fig. 7. Tunnel contour formed by points projected onto a plane perpendicular to the centerline.

Subsequently, all profiles were combined into a single 2D plane to analyze the variations in the tunnel's contour over time, highlighting changes observed across different years (see Fig. 8).

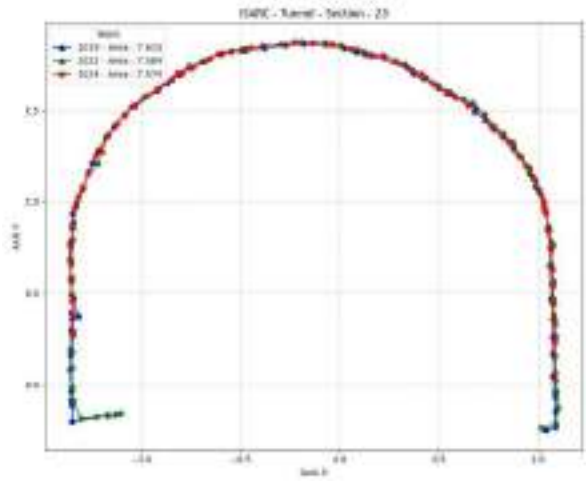


Fig. 7. Tunnel profiles from different years.

3.2 Damage identification

Using the three-point cloud datasets (2019, 2022, and 2024), each section was projected onto a 2D plane to analyze changes in the tunnel's cross-sectional area. Table 3 summarizes the reduction in area for a specific section, showing a consistent decrease over time. Additionally, the results indicate a reduction in tunnel width between 2019 and 2022 (see Fig. 8) and a consistent shift to the right across all sections over the years.

Table 4. Reduction in cross-sectional area

Year	Area (m2)	Percentage
2019	7.615	100%
2022	7.584	99.5%
2024	7.574	99.4%

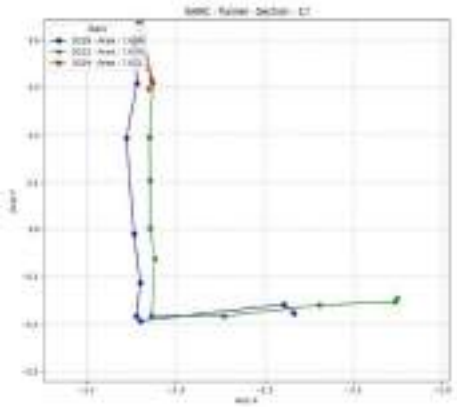


Fig. 8. Reduction observed in the lower-left section of the tunnel's cross-section.

An analysis was conducted along the entire tunnel, comparing cross-sectional areas from each year. The blue line represents the area from 2019, the green line corresponds to 2022, and the red line reflects 2024. The results reveal a noticeable reduction in the cross-sectional area of the tunnel over time, which may be attributed to ground movement or structural degradation of the tunnel. In Fig. 9, the comparison between 2019 and 2022 reveals noticeable changes in the tunnel's cross-sectional area, indicating structural modifications or deformation during this period. Similarly, Fig. 10 illustrates how the tunnel's sections in 2024 have shifted relative to 2022, showcasing continued changes over time

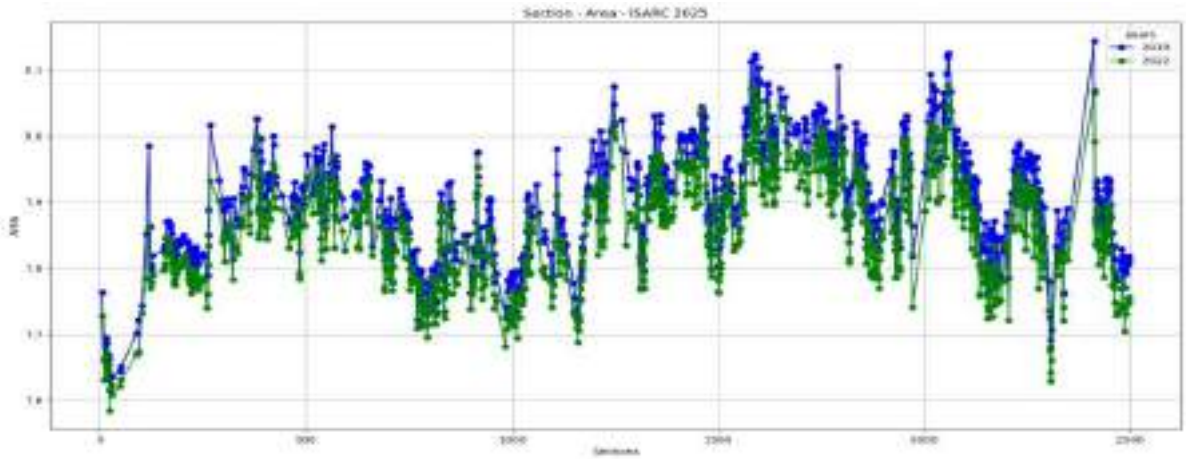


Fig. 9. Comparison of Tunnel Cross-Sections Between 2019 and 2022

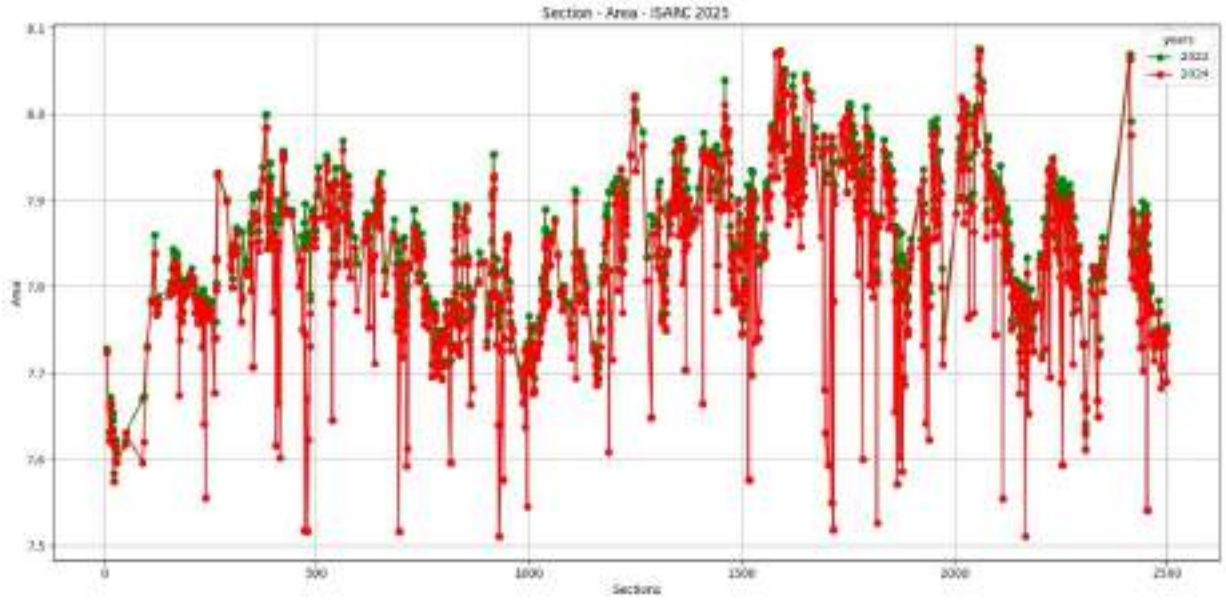


Fig. 10. Comparison of Tunnel Cross-Sections Between 2022 and 2024

4 Conclusions and Discussion

This study presents an innovative methodology for assessing tunnel deformation and structural changes over time using point cloud datasets acquired through terrestrial laser scanning (TLS) in 2019, 2022, and 2024. The following conclusions were derived:

- TLS has proven to be a reliable tool for capturing high-resolution data, facilitating detailed monitoring of the tunnel's structural changes. Its capacity to perform accurate measurements in challenging environments, such as dark underground tunnels, highlights its suitability for infrastructure assessment.
- Analysis of the tunnel profiles revealed a consistent reduction in cross-sectional area across all sections over the years. The results indicate a decline of 0.6% from 2019 to 2024, with notable deformation in specific regions, particularly the lower-left corner. This reduction may be attributed to geological movements or material degradation.
- A systematic rightward shift in the tunnel's cross-sections was observed over time. This movement underscores the need for continuous monitoring to identify and address potential instability risks associated with lateral displacement.
- The proposed algorithm effectively segmented the point cloud data, aligned it with the tunnel centerline, and projected profiles for comparative analysis. This approach enabled the detection of subtle changes in geometry, demonstrating its applicability for tunnel condition assessment.
- The findings highlight the importance of integrating advanced technologies, such as TLS and automated algorithms, into routine tunnel inspections. Such methods provide precise and actionable insights, reducing dependency on manual inspections and enhancing safety in hazardous conditions.

As observed, the findings have significant implications for structural monitoring and tunnel maintenance. Traditional tunnel monitoring techniques rely heavily on visual inspections, geodetic surveys and instrumentation-based systems. Although these methods have been widely used, they have limitations in terms of subjectivity, data resolution and safety. Visual inspections, for example, depend on the experience of engineers and can overlook subtle deformations. Geodetic surveys, which use total stations or leveling instruments, provide accurate measurements at discrete points, but require a lot of time and effort.

The progressive reduction of the cross-section indicates cumulative deformation that could compromise long-term stability. This problem could be due to the geomechanics of the area, such as ground movements, traffic loads in the area, or degradation of the tunnel materials. Early detection of deformations using TLS would allow engineers to implement preventive mitigation measures before critical failures occur. For instance, the use of metal trusses or increasing the thickness of shotcrete, the

characteristics of which must be carefully calculated. On the other hand, being able to detect these changes in the cross-section is a better alternative compared to visual identification or manual measurements, which are subjective and depend on the experience of the inspector. Automating TLS reports makes it possible to generate reports, facilitating informed decision-making for tunnel maintenance.

References

[1] A. Sainoki, S. Tabata, H. S. Mitri, D. Fukuda, and J. ichi Kodama, "Time-dependent tunnel deformations in homogeneous and heterogeneous weak rock formations," *Comput Geotech*, vol. 92, pp. 186–200, Dec. 2017, doi: 10.1016/J.COMPGEO.2017.08.008.

[2] J. P. Yang, W. Z. Chen, M. Li, X. J. Tan, and J. xin Yu, "Structural health monitoring and analysis of an underwater TBM tunnel," *Tunnelling and Underground Space Technology*, vol. 82, pp. 235–247, Dec. 2018, doi: 10.1016/J.TUST.2018.08.053.

[3] X. Tan, W. Chen, T. Zou, J. Yang, and B. Du, "Real-time prediction of mechanical behaviors of underwater shield tunnel structure using machine learning method based on structural health monitoring data," *Journal of Rock Mechanics and Geotechnical Engineering*, vol. 15, no. 4, pp. 886–895, Apr. 2023, doi: 10.1016/J.JRMGE.2022.06.015.

[4] M. S. Kovačević, M. Bačić, K. Gavin, and I. Stipanović, "Assessment of long-term deformation of a tunnel in soft rock by utilizing particle swarm optimized neural network," *Tunnelling and Underground Space Technology*, vol. 110, p. 103838, Apr. 2021, doi: 10.1016/J.TUST.2021.103838.

[5] X. Peng, M. Wang, B. Huang, and P. Lin, "Efficient automated method for characterizing discontinuities in tunnel face rock mass point clouds," *Tunnelling and Underground Space Technology*, vol. 154, p. 106117, Dec. 2024, doi: 10.1016/J.TUST.2024.106117.

[6] H. Yang and X. Xu, "Structure monitoring and deformation analysis of tunnel structure," *Compos Struct*, vol. 276, p. 114565, Nov. 2021, doi: 10.1016/J.COMPSTRUCT.2021.114565.

[7] W. Mu, L. Li, D. Chen, S. Wang, and F. Xiao, "Long-term deformation and control structure of rheological tunnels based on numerical simulation and on-site monitoring," *Eng Fail Anal*, vol. 118, p. 104928, Dec. 2020, doi: 10.1016/J.ENGFAILANAL.2020.104928.

[8] J. wen Zhou, J. lin Chen, and H. bo Li, "An optimized fuzzy K-means clustering method for automated rock discontinuities extraction from point clouds," *International Journal of Rock Mechanics and Mining Sciences*, vol. 173, p. 105627, Jan. 2024, doi: 10.1016/J.IJRMMS.2023.105627.

[9] J. Li, F. Zhou, P. Zhou, J. Lin, Y. Jiang, and Z. Wang, "A health monitoring system for inverted arch of salt rock tunnel based on laser level deformation monitor and wFBG," *Measurement*, vol. 184, p. 109909, Nov. 2021, doi: 10.1016/J.MEASUREMENT.2021.109909.

[10] S. K. Singh, B. P. Banerjee, M. J. Lato, C. Sammut, and S. Raval, "Automated rock mass discontinuity set characterisation using amplitude and phase decomposition of point cloud data," *International Journal of Rock Mechanics and Mining Sciences*, vol. 152, p. 105072, Apr. 2022, doi: 10.1016/J.IJRMMS.2022.105072.

[11] Q.-Y. Zhou, J. Park, and V. Koltun, "Open3D: A Modern Library for 3D Data Processing," *arXiv:1801.09847*.

3D Measurement System for Soil Loading by an Autonomous Backhoe using OPERA

Takayoshi Hachijo[1], Yutaro Fukase[1], Takashi Yokoshima[1], Yuki Miyashita[1], Shunsuke Kimura[1], Masanori Suzuki[2], Yuichiro Kasahara[3], Tomoya Kouno[3], Koshi Shibata[3], Ryo Kurazume[3], Daisuke Endo[4], Genki Yamauchi[4] and Takeshi Hashimoto[4]

[1]Institute of Technology, Shimizu Corporation, Japan
[2]Civil Engineering Technology Division, Shimizu Corporation, Japan
[3]Graduate School of Information Science and Electrical Engineering, Kyushu University, Japan
[4]Advanced Technology Research Team, Public Works Research Institute, Japan

hachijo_t@shimz.co.jp

Abstract -

We developed a 3D measurement system for autonomous construction and conducted demonstration experiments of autonomous construction using OPERA (Open Platform for Earthwork with Robotics and Autonomy), which is a research and development platform developed by the Public Works Research Institute. As a use case, we carried out soil loading into a dump truck by a backhoe in a test field. In this paper, we give overviews of our system and the demonstration experiments.

Keywords -

3D measurement; OPERA; Autonomous construction; Point cloud; Soil loading; ROS 2

1 Introduction

The Japanese construction industry is facing a labor shortage because of a declining birthrate, an aging population, and youth job disengagement. To address this issue, the Ministry of Land, Infrastructure, Transport and Tourism is promoting i-Construction 2.0 [1] in order to establish efficient construction processes based on digital technologies. The aim of i-Construction 2.0 is to revolutionize construction sites via three types of automation, namely, automation of construction work, automation of data linkage, and automation of construction management, the expectations being reduced labor and improved productivity.

Extensive research has been conducted on the development of autonomous construction machinery and the utilization of 3D data. R. Heikkilä et al. [2] investigated case studies of automated construction in Europe, highlighting the increasing use of methods that employ 3D data, which plays a critical role in automation. I. Niskanen et al. [3] implemented the 3D LiDAR on the boom of the backhoe to recognize the shape of soil loaded on the vessel of the dump truck. A. Assadzadeh et al. [4] have employed deep learning to estimate the posture of the backhoe using RGB image data captured by cameras installed at the site.

In support of the aforementioned efforts, our work involves automating construction work by using autonomous construction machinery. As an implementation of autonomous civil construction work, we target soil loading associated with shield tunneling as shown in Figure 1. Soil excavated from the shield tunnel is transported to an external soil pit via pipelines and conveyor belts and then loaded into a dump truck by a backhoe and transported to a treatment plant. The backhoe operates depends on when the dump truck arrives, which in turn depends on traffic conditions, and this often causes waiting time for the backhoe operator. Therefore, automating these soil-loading tasks would increase efficiency and productivity. Furthermore, in this case the locations of the backhoe, truck parking, and soil pit hardly ever change. We can install sensors in the environment to detect the shape of the earth mound in the soil pit and the position of the dump truck. This requires fewer sensors than in more-dynamic situations, and this system is applicable even if the dump truck has no sensors of its own. For these reasons, our aims were to automate the soil loading involved in shield tunneling and conduct demonstration experiments in a test field. In this paper, we give an overview of our system the experiments performed in the test field.

2 System overview

We have developed three components for realizing autonomous construction OPERA (Open Platform for Earthwork with Robotics and Autonomy), which is a platform for autonomous construction; ROS2-TMS for Construction, which is a management system for controlling construction machinery; and a 3D measurement system. Figure 2 shows an overview of the system, in which two LiDARs detect the shape of the earth mound and three LiDARs detect the dump truck. The three components

are on the same network and communicate using ROS 2 [5] to enable autonomous construction. In ROS 2, the communication method for sending and receiving commands and messages is called topic communication, and these commands and messages are called topics; sending topics is called publishing and receiving topics is called subscribing. In this paper, we use the same terminology.

2.1 OPERA

We use OPERA [6, 7] to control autonomous construction machines. Developed by the Public Works Research Institute (PWRI), OPERA is a research and development platform for automating earth works; it contains a set of common control signals, middleware, simulators, a demonstration environment, and autonomous construction machinery, and it is available on GitHub for open development [8]. Moreover, OPERA was developed in ROS 2, which allows construction machinery and sensors to be controlled by common commands independent of the manufacturer.

2.2 ROS2-TMS for Construction

We use ROS2-TMS for Construction [9] to manage the tasks of autonomous machinery developed by Kyushu University. This task management system enables the accumulation and management of environmental information from sensors in construction sites and machinery; it realizes autonomous construction by giving commands to robots and autonomous construction machinery based on the flowchart and data . TMS_IF for OPERA [10] is an interface developed to publish commands via OPERA based on the indicated tasks; it was also developed in ROS 2 and uses the open package Moveit2 [11] for motion planning and the open package Nav2 [12] for autonomous driving. We control the construction machinery by providing information from Mongo DB[13] as the database using ROS 2.

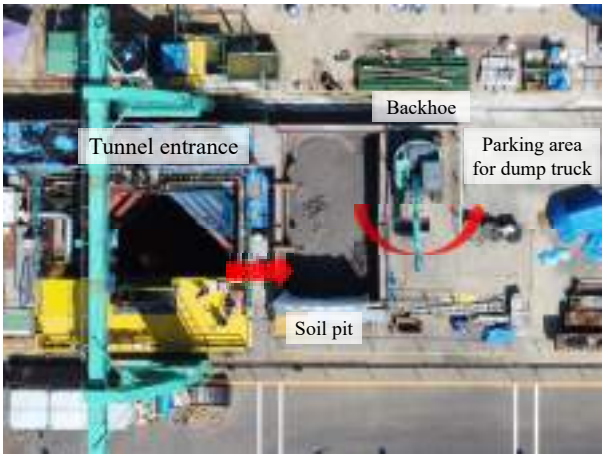


Figure 1. Use case.

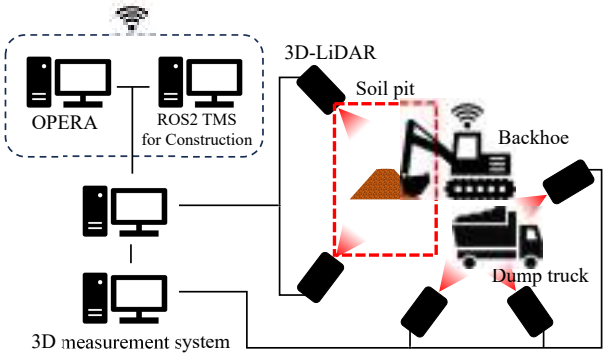


Figure 2. Overview of system.

2.3 3D measurement system

We developed the measurement system by using multiple 3D LiDARs to detect the shape of the earth mound and the location of the dump truck. The resulting point-cloud data enable accurate recognition of the shapes of the targets, and this allows calculation of the best position for the backhoe to excavate more soil. We explain the registration method for the point-cloud data from the multiple 3D LiDARs in Section 2.3.1, the excavation recognition methodology in Section 2.3.2, and the loading recognition methodology in Section 2.3.3.

2.3.1 Registration of point-cloud data

Because the LiDAR coordinate system uses the sensor center as the origin, the point-cloud data obtained by each LiDAR are in different coordinate systems. In this study, we define the center of the backhoe as the origin of the world coordinate system W, and we calculate the homogeneous transformation matrix $H_{L\to W}$ from each LiDAR coordinate system L to W.

As shown in Figure 3, we placed three spheres in the scan range of all the LiDARs , and we estimate the center coordinates of each sphere from the obtained point-cloud data as shown in Figure 4; Figure 4(b) shows the cropped point-cloud data around one of the spheres from Figure 4(a). Using four points selected randomly from the cropped point-cloud data of the sphere, we obtain its center coordinates $s(a, b, c)$ by solving the following equation:

$$\begin{pmatrix} x_1 - x_0 \\ x_2 - x_0 \\ x_3 - x_0 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} x_1^2 - x_0^2 + y_1^2 - y_0^2 + z_1^2 - z_0^2 \\ x_2^2 - x_0^2 + y_2^2 - y_0^2 + z_2^2 - z_0^2 \\ x_3^2 - x_0^2 + y_3^2 - y_0^2 + z_3^2 - z_0^2 \end{pmatrix} /2. \tag{1}$$

Next, we calculate the L_2 norm between s and each cloud point and count the points for which the L_2 norm is within threshold. This process is iterated 1000 times, after

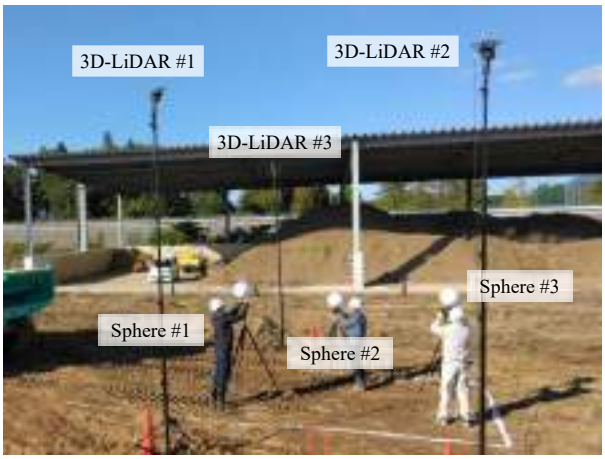


Figure 3. Installation of spheres.

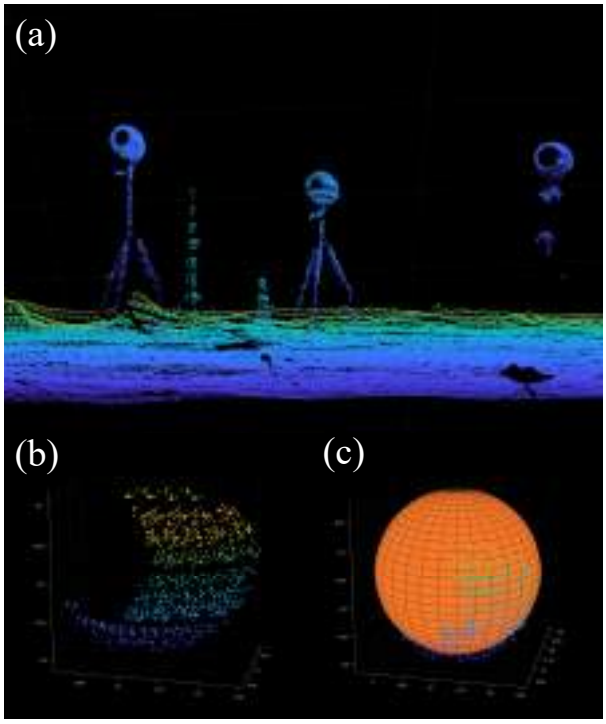


Figure 4. Point-cloud data of spheres: (a) LiDAR-obtained point-cloud data; (b) cropped point cloud of a sphere; (c) estimated sphere.

which the s with the most points is adopted. We estimate $S_L = \{s_1, s_2, s_3\}$ by calculating for the three spheres.

We then placed prisms at a mechanically known point on the backhoe and the center positions of the spheres. The sphere coordinates S_W in the world coordinate system were obtained by using a total station.

We obtain the homogeneous transformation matrix $H_{L \to W}$ from S_L and S_W and transform each LiDAR coordinate system into the world coordinate system. Figure 5 shows the point-cloud data transformed by the homogeneous transformation matrix.

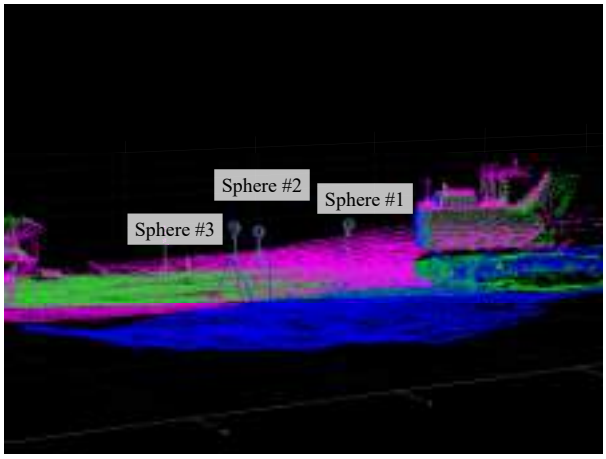


Figure 5. Integrated point-cloud data.

2.3.2 Recognition for excavation

We estimate the excavation points from the shape of the earth mound by using the point-cloud data transformed to the world coordinate system. We extract the point-cloud data in the soil pit and divide them into a planar grid. At each grid location, the lowest z coordinate is the height of the grid, and these grid data form the height map from which the highest grid point is estimated as being the excavation point. Figure 6 shows the height map of the earth mound. The excavation points are published constantly in the database of ROS2-TMS for Construction as the ROS 2 topics.

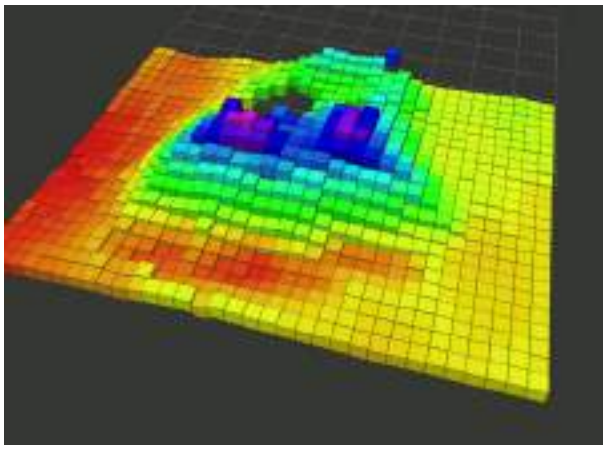


Figure 6. Height map of earth mound.

2.3.3 Recognition for loading

First, we estimate the current position of the dump truck. We obtain the point-cloud data of an empty dump truck P_d in advance and find the coordinates at the four corners of the vessel O_d from P_d. RANSAC is used to derive the homogeneous transformation matrix that converts P_d into the current point-cloud data P_r. By using the matrix to transform O_d, the current position of the dump truck is estimated as the current vessel coordinates O_c.

Next, we estimate the loading point by using O_c to extract the inlier point-cloud data in P_r, and the extracted point-cloud data are divided into a planar grid. The points with the highest z coordinate give the height of the grid, and these gird data form the height map. In this study, the grid map was padded and smoothed via convolution. From the height map, the lowest grid point is estimated as being the loading point. In addition, the loading volume inside the vessel is estimated by comparing P_d and P_r.

3 Overview of experiment

We conducted the task of excavating and loading soil with one autonomous backhoe (ZX200, Hitachi Construction Machinery Co., Ltd.) and one crawler carrier (IC200, Kato Works Co., Ltd.) at the PWRI DX field in Tsukuba City, Ibaraki Prefecture, Japan. Figure 7 shows the experiment. The earth mound was surrounded by pipes that simulated the soil pit. Five LiDARs (Livox) were installed to detect the shape of the earth mound and the location of the crawler carrier and estimate the excavation and loading points. Those points were then published to the database of ROS2-TMS for Construction as the ROS 2 topics. The backhoe generated a trajectory based on the target coordinates in the database and performed excavation and loading operations. When the inside of the vessel reached a certain loading capacity, a loading completion flag was published to the database, and the operation was completed.

3.1 Excavation test

Figure 8 shows the excavation by the backhoe, which avoided the pipes by us preparing walls as obstacles in the simulation environment in advance. Figure 8(a) shows the backhoe waiting for the start flag of ROS2-TMS for Construction, and Figure 8(b) shows the excavation task starting and the backhoe having subscribed the excavation point from the database. Figure 8(c)–(f) show the backhoe executing the excavation tasks twice and excavating the top of the earth mound. In the study of Niskanen et al.[3], the 3D LiDAR was installed on the boom of the backhoe to measure the vessel. In contrast, we install the 3D LiDAR at the site. By installing the 3D LiDAR at the site, it is possible to recognize the entire soil mound, including the backside, which cannot be observed from the backhoe. Additionally, since point cloud data of the soil mound can be continuously acquired regardless of the backhoe's posture, the backhoe can operate without waiting for measurement and estimation times.

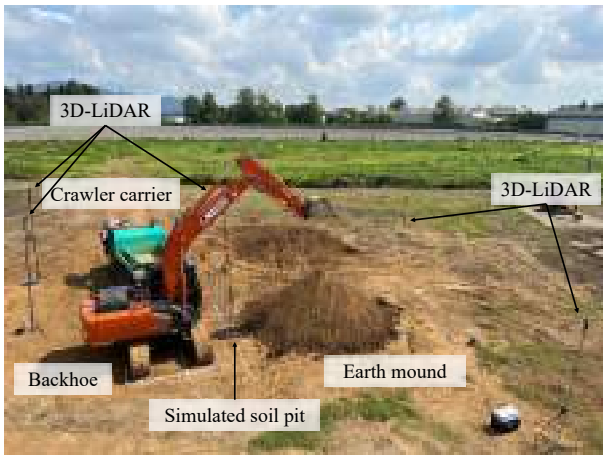


Figure 7. Overview of experiment.

In this study, the backhoe is assumed to remain stationary. Since the coordinate system of the backhoe is obtained through prism measurements, it is necessary to perform measurements each time the backhoe moves. This is our future work, and we will explore methods to automatically recognize the position and orientation of the backhoe in the future.

3.2 Loading test

Figure 9 shows the point-cloud data at different times; in each panel, the purple rectangle is the estimated outline of the vessel and the red sphere is the loading point at the relevant time. The vessel is recognized when the evaluation value exceeds a certain level from the registration results, and the lowest point inside the vessel is estimated as the loading point in the height map in this rectangle. By performing coordinate transformation from the position and orientation of the vessel, it is possible to continuously estimate the position and orientation of the dump truck. Acquiring point cloud data from multiple angles using three LiDARs improves the accuracy of registration, thereby stabilizing the estimation of position and orientation. The vessel is not recognized in Figure 9(a) but is recognized in Figure 9(b). In Figure 9(c), the loading point is estimated in the vessel and published to the database. In Figure 9(d)–(h), the position of the loading point changes as the soil is loaded. Figure 10 shows the inside of the vessel before and after loading, and the loading volume was estimated from the point-cloud data after the loading operation.

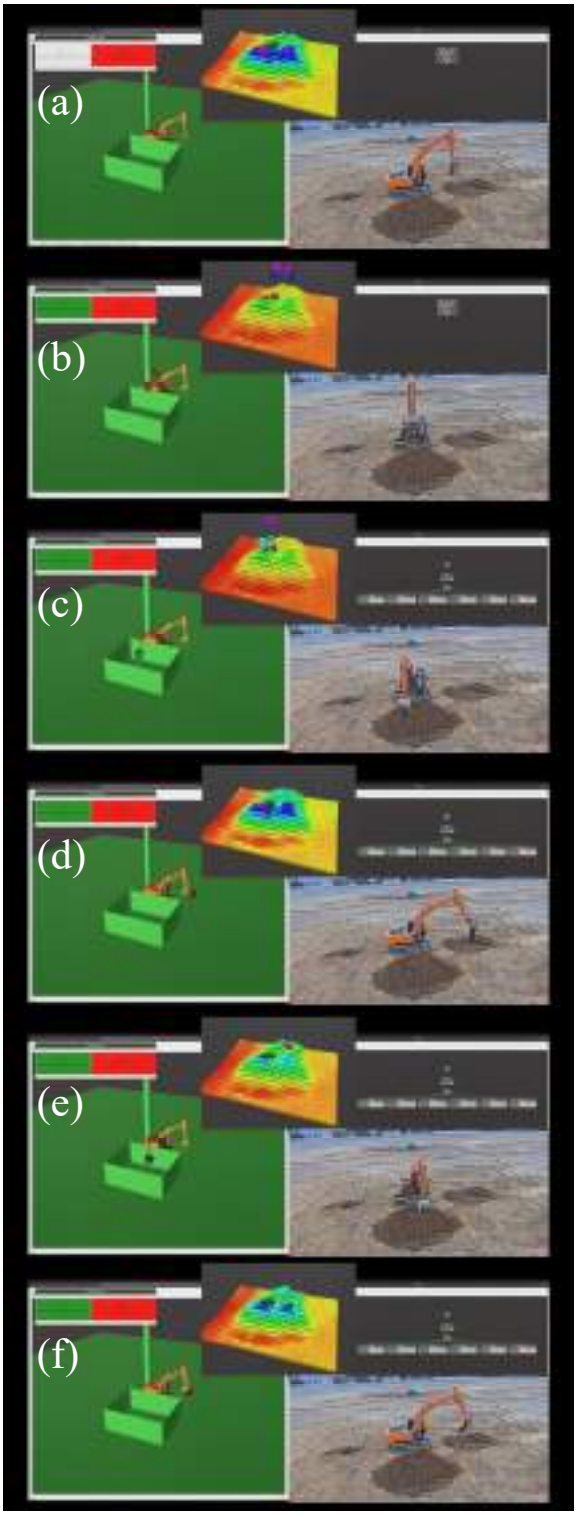


Figure 8. Each figure shows the simulation (left), the actual backhoe (right), and the height map (center).

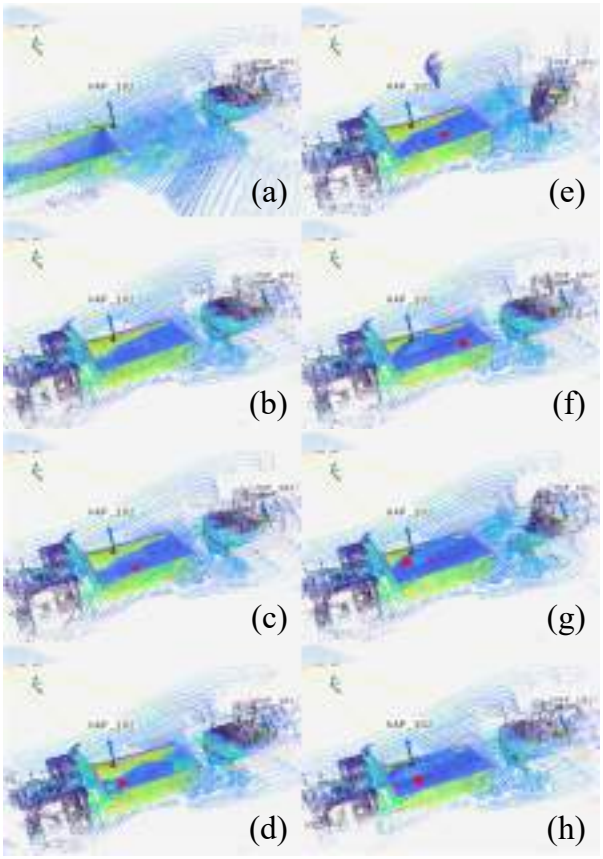


Figure 9. Estimating the loading point at each time. The purple rectangle is the estimated outline of the vessel and the red sphere is the loading point at each time.

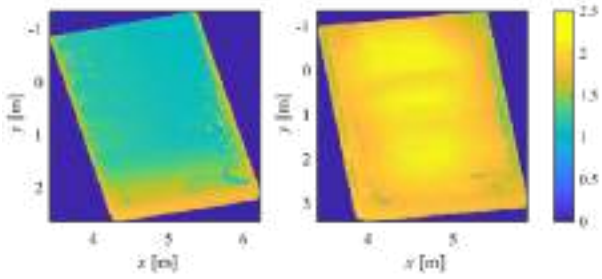


Figure 10. Inside of vessel.

4 Conclusion

We have developed a 3D measurement system for autonomous construction. We configured the task of loading soil into a dump truck by an autonomous backhoe as the use case, and we conducted demonstration experiments of this task in the test field.

Installing multiple LiDARs around the targets enabled the latter to be measured without any blind spots. The co-

ordinate transformation method using spheres facilitated registration of the point-cloud data obtained by the LiDARs.

The obtained point-cloud data allowed detection of the shape of the earth mound. The excavation points were published to ROS2-TMS for Construction as the task management system, and the backhoe excavated autonomously based on the excavation points.

The obtained point-cloud data allowed detection of the vessel position and the shape of the loaded soil. The loading points were also published to ROS2-TMS for Construction.

The integration of multiple point-cloud data which was achieved by our developed system, ensured reliable data acquisition. We will focus on exploring optimal excavation methods using the acquired data.

In this measurement system, the LiDAR-obtained point-cloud data are large and can overload the network. Therefore, in future work we will construct a robust network and continue to develop the system for application to actual construction sites.

References

[1] MLIT. i-construction 2.0. On-line: `https://www.mlit.go.jp/tec/constplan/content/001738240.pdf`, Accessed: 01/11/2024.

[2] R. Heikkilä and M. Jaakkola. Automation of road construction - the state of the art in europe. In *Proceedings of the 23rd International Symposium on Automation and Robotics in Construction(ISARC2006)*, pages 7–10, Tokyo, Japan, 2006.

[3] I. Niskanen, G. Duan, H. Suzuki, D. Endou, M. Immonen, Hiltunen M., G. Yamauchi, P. Tyni, T. Hashimoto, and R. Heikkilä. Determining payload on platform of lorry in real time using integrated 3-d lidar from excavator boom. In *IEEE Transactions on Instrumentation and Measurement*, volume 73, pages 1–7, 2024.

[4] A. Assadzadeh, M. Arashpour, H. Li, R. Hosseini, F. Elghaish, and S. Baduge. Excavator 3d pose estimation using deep learning and hybrid datasets. *Advanced Engineering Informatics*, 55:101875, 01 2023. doi:10.1016/j.aei.2023.101875.

[5] Open Robotics. Ros 2 documentation: Humble. On-line: `https://docs.ros.org/en/humble/index.html`, Accessed: 01/11/2024.

[6] G. Ymauchi, D. Endo, H. Suzuki, and T. Hashimoto. Proposal of an open platform for autonomous construction machinery development. In *Proceedings of the 40th International Symposium on Automation and Robotics in Construction(ISARC2023).*, pages 186–191, 2023.

[7] D. Endo, Y. Matsusaka, G. Ymauchi, and T. Hashimoto. Research on an open source physical simulaotr for the autonomous construction machinery development. In *Proceedings of the 41th International Symposium on Automation and Robotics in Construction(ISARC2023).*, pages 1303–1306, 2024.

[8] github. pwri-opera. On-line: `https://github.com/pwri-opera`, Accessed: 01/11/2024.

[9] github. irvs/ros2_tms_for_construction. On-line: `https://github.com/irvs/ros2_tms_for_construction`, Accessed: 01/11/2024.

[10] Y. Kasahara, T. Itsuka, K. Shibata, T. Kouno, R. Maeda, K. Matsumoto, S. Kimura, Y. Fukase, T. Yokoshima, G. Ymauchi, D. Endo, T. Hashimoto, and R. Kurazume. Task management system for construction machinery using the open platform opera. In *Proceedings of the IEEE International Conference on Robot Human Interactive Communication (RO-MAN)*, August 2024.

[11] PickNik Robotics. Moveit 2 documentation. On-line: `http://abcd.efg.hij`, Accessed: 01/11/2024.

[12] Open Navigation. Nav2. On-line: `https://docs.nav2.org/`, Accessed: 01/11/2024.

[13] Mongo DB. Documentation. On-line: `https://www.mongodb.com/docs/`, Accessed: 01/11/2024.

One-Shot Indoor Positioning Using 360-Degree Photos

Daeyoung Gil[a] and Ghang Lee[a, b*]

[a]Department of Architecture and Architectural Engineering, Yonsei University, Republic of Korea
[b]Technical University of Munich, Institute for Advanced Studies, Munich, Germany
[*]Corresponding author
E-mail: jacob89@yonsei.ac.kr, glee@yonsei.ac.kr

Abstract

Vision-based indoor positioning approaches are increasingly gaining attention due to their cost-effectiveness and scalability. The core principle of vision-based indoor positioning is to identify the most visually similar space to the reference space images. A major challenge in previous vision-based indoor positioning methods arises from the need for large annotated datasets to ensure robust model performance, necessitating coverage of space images from various angles and points of view. To address this limitation, we propose a method that applies feature matching to 360-degree images, enabling indoor positioning with just a single reference image per space. This eliminates the need for multiple reference images by preprocessing the input image through feature-descriptor-based alignment and cube-map projection, allowing the image to be adjusted to better match the position and shape of the reference image. Experiments conducted on six different floor plans achieved an accuracy of 72.57% using only one reference image, confirming the feasibility and efficiency of this lightweight approach to indoor positioning.

Keywords –

360-degree photo; Image feature matching; Indoor positioning; Computer vision; Few-shot learning; One-shot learning

1 Introduction

Unlike outdoor positioning often relies on global positioning system (GPS) [1], indoor environments require different approaches because GPS signals cannot be received. The most common indoor positioning methods are signal-based techniques such as Wi-Fi [2], ultra-wideband (UWB) [3], and radio frequency identification (RFID) [4], which typically rely on triangulation methods and thereby offer robust performance. However, these techniques can be heavily influenced by indoor objects, such as furniture and walls, that interfere with signal propagation. Moreover, they require the installation of dedicated equipment, which can limit their applicability. In these contexts, vision-based indoor positioning methods present a more intuitive and robust alternative than signal-based methods [5].

Indoor positioning research based on visual information often involves extracting features from images and then determining proximity to predefined reference points [6]. Recent advances in machine learning and deep learning have enabled systems to automatically identify salient features from video frames, which can subsequently be used to estimate a user's location relative to these reference points. By harnessing visual cues, such systems can potentially overcome the limitations of wireless signal-based approaches, offering finer spatial resolution and rich contextual information about the surrounding environment.

Despite the advantages offered by vision-based indoor positioning methods, several challenges remain. A major issue is the heavy dependence on large, labeled image datasets; as the number of positioning reference points increases, so does the need for more extensive data collection. Another concern is the potential confusion arising in visually similar spaces [7]. Even in the same location, images can vary significantly depending on the camera's extrinsic parameters, making it difficult to distinguish between different areas that share analogous visual features. Therefore, to maintain the benefits of image-based indoor positioning while overcoming these challenges, methods are needed that can operate effectively with fewer images and reduce reliance on specific camera parameters.

In this paper, we propose a one-shot indoor positioning method for residential buildings that leverages 360-degree images. Because many residential buildings repeat the same floor plan on every floor, they offer an environment well-suited for vision-based indoor positioning. To address the conventional reliance on large-scale datasets, we advocate the use of 360-degree images instead of standard photographs. Compared to conventional images, 360-degree images exhibit lower dependence on camera angles and positions, and they can be manipulated through projection to highlight specific visual information [8]. Building on these advantages, we

prepare a single reference image for each room as training data and use a corresponding query image to determine the user's current location by comparing their visual similarity. As part of this process, we employ feature extraction to adjust the orientation of the 360-degree images. We then perform cube-map projections on the aligned images, segmenting them so that each direction can be compared directly. Through this approach, robust indoor positioning is achievable with only one representative 360-degree image per room.

2 Related Studies

2.1 Image-based Indoor Positioning System

Image-based positioning offers the advantage of leveraging dense visual information with minimal equipment while providing intuitive insights into spatial relationships. Traditionally, such methods have relied on feature-based techniques like the Scale-Invariant Feature Transform (SIFT) or Oriented FAST and Rotated BRIEF (ORB) [9] for keypoint detection and matching. By correlating extracted features with reference images or pre-built maps, researchers have demonstrated accurate localization even in the most intricate, multi-room settings. Concurrently, recent advances in neural network architectures have led to the widespread development of similarity-based methods [7]. Although these approaches share the overarching goal of precise localization, they primarily differ in how they extract and represent features from given images.

Despite their promise, these methods often require large-scale image datasets and carefully tuned parameters to ensure robustness under varying conditions. Keypoint-descriptor-based methods (e.g., SIFT, ORB) circumvent extensive training procedures, but tend to be highly sensitive to environmental changes in lighting or perspective [10]. Consequently, additional processes, such as rigorous parameter tuning or ensemble modelling, are often recommended to improve reliability. Moreover, in tall buildings where multiple floors replicate near-identical room layouts, distinguishing among different spaces becomes increasingly problematic. This issue is exacerbated by the limited field of view (FoV) in conventional images, which can capture only a fraction of the environment and thus yield inconsistent results when camera angles differ.

To overcome these challenges, 360-degree imaging has gained momentum as an alternative. By capturing the entire environment from a single vantage point, such images minimize the need for multiple photographs taken at various angles or distances [11]. Not only does this broader coverage enhance feature extraction, but it also simplifies the alignment process during positioning. Furthermore, positioning methods that employ scene understanding of 360-degree photos through segmentation-based techniques have also been introduced [12,13]. However, deep learning-based methods, particularly those leveraging convolutional neural networks and other advanced architectures, demand vast amounts of training data to achieve robust performance. Generating labeled image datasets for every possible indoor environment can be both time-consuming and cost-intensive, which is where few-shot learning becomes highly relevant.

2.2 Few-shot Learning in Construction

Few-shot learning techniques enable models to generalize from a limited number of examples [14], thereby reducing the reliance on large, labeled datasets. They encompass various approaches, including meta-learning [14] and similarity-based methods [15], and provide a promising solution by requiring only minimal data per class. In the construction domain, where collecting extensive labeled datasets can be costly and labor-intensive [16], few-shot learning helps alleviate the burdens typically associated with vision-based monitoring techniques [17].

In the context of indoor positioning, few-shot learning substantially decreases the need for exhaustive image capture and labeling in each room or space. By focusing on representative images or leveraging meta-level patterns across different tasks, these approaches maintain high accuracy even in visually similar environments—a feature particularly advantageous in high-rise residential buildings with repeated floor plans. Through the integration of 360-degree imaging and few-shot learning, it becomes possible to capture comprehensive spatial information while avoiding the data-intensive bottlenecks often encountered in deep learning methods.

2.3 Research Gap

In summary, two major challenges arise when implementing indoor positioning systems that rely primarily on visual information and image similarity [7]. One involves calibrating the camera's extrinsic parameters [11], and the other concerns assembling the large dataset required to construct the model [8]. As a potential solution, the use of 360-degree images has been proposed in prior work. Within this approach, some studies focus on calibration, while others involve subdividing images to increase the available data. However, the former still requires a feasibility assessment for practical use, and the latter necessitates additional preprocessing, which restricts direct application under few-shot conditions.

In this study, we combine the strengths of these prior methods to propose an indoor positioning technique that

functions with only a small number of images. Through feature matching, we correct the camera's external parameters, and by employing CMP-based 360-degree images, we make full use of omnidirectional visual information. As a result, accurate positioning becomes possible even with minimal data, providing a novel approach to vision-based indoor localization.

3 Research Proposal

Our proposed method is divided into two primary steps (Figure 1). The first step is a preprocessing stage performed when the query image is provided. During this stage, we align the 360-degree image and, if necessary, apply inpainting to remove the photographer or other unintended objects. This process corrects for variations in camera angle and eliminates noise that could otherwise skew comparisons, thereby ensuring an equal basis for evaluating similarity.

In the second step, we compare the corrected query image with each reference image. By applying a projection technique, we segment both the query image and the reference image into multiple directions, splitting them into six sections via cube-map projection (CMP), so that images corresponding to the same direction can be compared individually [8]. This approach addresses the distortions inherent in the equirectangular projection (ERP) format of 360-degree images, which can vary based on the camera's position inside the space. We repeat these comparisons for all reference images in the indoor environment, and the final positioning result is determined by selecting the space with the highest average similarity score. Further details on each step of the process are provided in the following subsections.

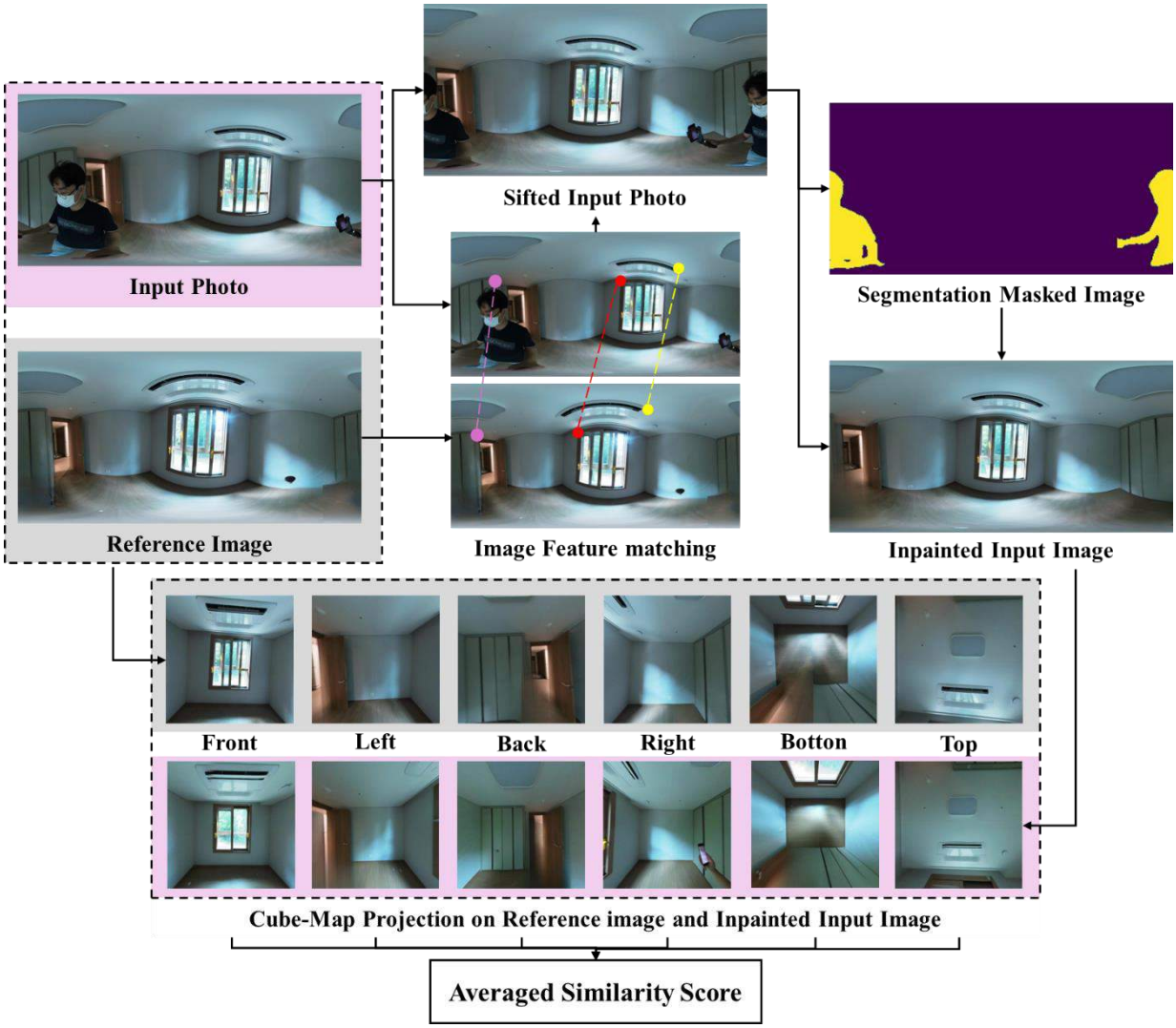


Figure 1. Overall flow of preprocessing on input image and similarity comparison with reference image

3.1 Preprocessing on Input Image

First, once the input photo is provided, serving as the query image, both the query image and the reference image undergo feature extraction using keypoint-descriptor methods. In this study, we employ an ensemble model comprising ORB [9], BRISK [18], and AKAZE [19]. [20][10]We then measure the horizontal distance between the keypoint-pairs that yield the highest match scores. This distance is used to shift the input image so that it is aligned to face the same direction as the reference image.

Next, if the image contains a human subject, it introduces noise that must be removed. We accomplish this by performing inpainting technique, which requires a segmentation mask to identify the person's location. In this process, we employ semantic segmentation to generate the mask. Specifically, we use the BEiT model [21], pre-trained on the ADE20K dataset [22], for segmentation, and the LaMa model for inpainting [23].

3.2 Similarity Comparison

After aligning the orientation of the input image and removing noisy objects, we compare the resulting image with a reference image to evaluate their similarity. In this study, we use a CMP approach for this comparison, aiming to compensate for the substantial distortions in ERP images that can occur even within the same space. In typical residential buildings, each room can be approximated as having a roughly cubic shape; therefore, once the reference image is properly adjusted, each CMP section captures a distinct wall in the room. This setup allows both the CMP-based reference image and the pre-processed input image to be compared on a wall-by-wall basis.

To quantify similarity between images, we employ a pretrained Siamese network that outputs a similarity score [24]. The final similarity score between one input photo and a reference image is the average of the scores obtained from the six CMP sections. Repeating this process for every reference image in each room, we identify the final location as the room whose reference image yields the highest average similarity.

4 Experiment & Results

4.1 Dataset Preparation and Setting for Experiment

Performance experiments on the proposed method were conducted in residential buildings. Specifically, we selected six housing units, and categorized their spaces into eight areas: ["Bathroom", "Bedroom", "Living Room", "Dress Room", "Balcony", "Utility Room", "Unit Room", "Kitchen"]. In each space, we captured 360-degree images. When a single unit contained multiple spaces of the same type (e.g., multiple bathrooms or bedrooms), we labeled them numerically to distinguish them as separate spaces. As a result, we compiled 288 images across 11 spatial categories for our experimental dataset (Table 1).

In addition, we prepared separate reference images for each room. To correct for potential distortions that could arise during the similarity-check process, these reference images were taken from the center of each room. Once converted via CMP, the resulting six images were oriented to provide frontal views of each wall in the space.

Table 1. Photo distribution of each room in the test Dataset

Balcony	**Dressroom**	**Kitchen**	**Livingroom**
18	20	30	30
Bedroom	**Bathroom**	**Unit**	**Utility**
34 / 31 / 35	24 / 21	24	21

4.2 Performance Check

In our experiments, the primary parameters requiring careful configuration were the number of features used during the feature matching step and the FoV value applied in the CMP process. In this study, we limited the number of matched features to one and fixed the FoV at 90, based on empirical findings that the performance gains from using a larger number of features were outweighed by the degradation caused by increased noise.

Furthermore, for feature matching, we employed three representative keypoint descriptors (ORB, BRISK, and AKAZE) and ensemble model. Propose ensemble model selects the best descriptor on a per-match basis by choosing the one that yields the highest match score, thereby enhancing the overall reliability of the matching process.

Based on the ensemble model, the highest performance among the tested descriptors was 72.57%, largely attributable to BRISK, which was the primary descriptor used in the ensemble (Table 2). Notably, previous studies employing similar approaches relied on AKAZE [11], suggesting that no single descriptor offers superior performance in all scenarios [10]. Instead, a composite approach like the ensemble appears to be more effective overall.

A closer examination of the data reveals varying classification performance across different rooms (Figure 2). Bedrooms 1, 2, and 3 were frequently misclassified due to their high degree of visual similarity, while spaces

with more distinctive features, such as balconies or dressing rooms, showed better performance. In contrast, although bathrooms displayed relatively favorable recall values, their precision scores dropped significantly. This was attributable to low feature-matching scores across all descriptors, which led to reduced overall similarity scores. One likely cause is the large mirror occupying one wall, which may have interfered with the feature-matching process. Future research should address this issue to further enhance system performance.

Table 2. Average accuracy according to each descriptor

Descriptor	ORB	BRISK	AKAZE	Ensemble
Accuracy	69.44%	71.53%	68.40%	72.57%

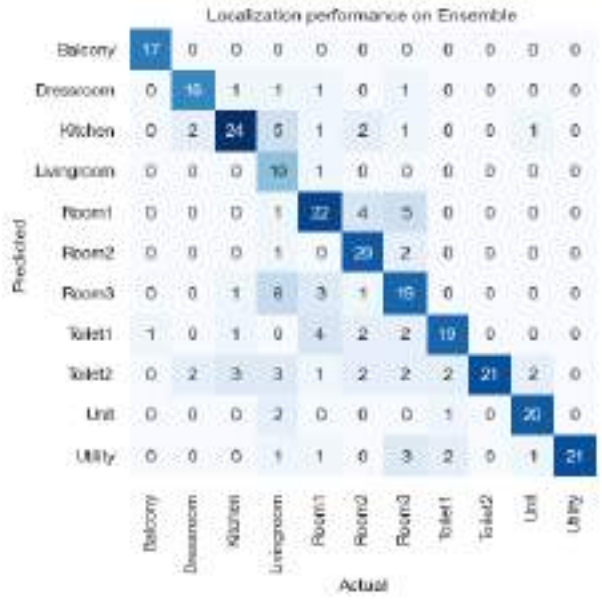


Figure 2. Confusion matrix of the ensemble model for each room

4.3 Ablation Studies

To evaluate how our proposed method, which utilizes feature matching for image alignment and CMP for image comparison, contributes to overall performance, we conducted an ablation study by selectively removing each of these components. First, (a) we measured accuracy using only the reference image and the input photo, without either feature matching or CMP, resulting in 9.13% accuracy. This is the result of the previous method based on a 360-degree image approach that relies solely on simple feature descriptors [11], excluding all the suggestions from this study. Second, (b) we performed feature matching for image alignment on the 360-degree images but used the ERP images for the comparison step, yielding 8.14% accuracy. Third, (c) we omitted the feature matching step and used only CMP, reaching 8.56% accuracy (Table 3).

Table 3. Performance changes according to application of suggestions

	(a) Without proposal	(b) Only feature matching	(c) Only CMP comparison	(d) Proposed method
Accuracy	9.13%	8.14%	8.56%	72.57%

These results are significantly lower than the 72.57% accuracy achieved by (d) our complete method. Moreover, given that the dataset contains 11 classes, the near 9–10% performance suggests results approaching random chance, thereby underscoring the limited utility of each individual component on its own. Consequently, our findings indicate that the benefits of the proposed method manifest only when feature matching and CMP are applied in tandem. Furthermore, considering that existing studies typically rely on feature matching alone [11], the considerable distortion inherent in 360-degree images can introduce major complications, supporting the necessity of our combined approach for robust indoor positioning.

5 Conclusion

This study presents a 360-degree image-based indoor positioning method that can be applied in small, repetitive indoor spaces using a single reference image. Unlike conventional images, 360-degree images provide substantially more visual information, which can alleviate the demands for a large dataset. However, to address the potential distortion and noise inherent in these images, we propose a two-step approach comprising of a preprocessing phase and a comparison phase. The method achieved an accuracy of 72.56% using only a single image per room. This represents a significant improvement compared to traditional feature-matching-based methods that are difficult to apply in practice [11], as confirmed by the ablation study. Furthermore, the performance is comparable to previous studies that required large amounts of training data [7], demonstrating that this study successfully achieved its goal of developing a practical model with minimal image input.

The primary contribution of this work lies in proposing a practical indoor positioning system that requires only one 360-degree image per room, alleviating the burden of preparing a large dataset in previous vision-based methods or the high cost of installing signal-based systems. Another key contribution of this study lies in the automated calibration approach by combining the use of

feature matching and an image similarity comparison technique. When image shifting via feature matching is integrated with CMP-based image similarity, we can achieve a level of performance suitable for practical use, even with only a single image per room. These benefits can also be extended to other studies requiring approximate location tracking, thus broadening the potential applications of our method.

Nevertheless, future research should aim to further enhance positioning accuracy, particularly in spaces like bathrooms that exhibit lower performance. Additionally, although the proposed method demands substantial computational resources and processing time, we anticipate that subsequent optimization and module simplification in future work will help mitigate these limitations.

6 Acknowledgement

This work was supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant, funded by the Ministry of Land, Infrastructure and Transport (No. RS-2024-00407028).

References

[1] A.S. Rao, M. Radanovic, Y. Liu, S. Hu, Y. Fang, K. Khoshelham, M. Palaniswami, T. Ngo, Real-time monitoring of construction sites: Sensors, methods, and applications, Automation in Construction 136 (2022) 104099. https://doi.org/10.1016/j.autcon.2021.104099.

[2] M. Abbas, M. Elhamshary, H. Rizk, M. Torki, M. Youssef, WiDeep: WiFi-based Accurate and Robust Indoor Localization System using Deep Learning, in: 2019 IEEE International Conference on Pervasive Computing and Communications (PerCom, 2019: pp. 1–10. https://doi.org/10.1109/PERCOM.2019.8767421.

[3] P. Dabove, V. Di Pietra, M. Piras, A.A. Jabbar, S.A. Kazim, Indoor positioning using Ultra-wide band (UWB) technologies: Positioning accuracies and sensors' performances, in: 2018 IEEE/ION Position, Location and Navigation Symposium (PLANS), 2018: pp. 175–184. https://doi.org/10.1109/PLANS.2018.8373379.

[4] T. Sanpechuda, L. Kovavisaruch, A review of RFID localization: Applications and techniques, in: 2008 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, 2008: pp. 769–772. https://doi.org/10.1109/ECTICON.2008.4600544.

[5] Y. Wei, B. Akinci, A vision and learning-based indoor localization and semantic mapping framework for facility operations and management, Automation in Construction 107 (2019) 102915. https://doi.org/10.1016/j.autcon.2019.102915.

[6] P. Pascacio, S. Casteleyn, J. Torres-Sospedra, E.S. Lohan, J. Nurmi, Collaborative Indoor Positioning Systems: A Systematic Review, Sensors 21 (2021) 1002. https://doi.org/10.3390/s21031002.

[7] I. Ha, H. Kim, S. Park, H. Kim, Image retrieval using BIM and features from pretrained VGG network for indoor localization, Building and Environment 140 (2018) 23–31. https://doi.org/10.1016/j.buildenv.2018.05.026.

[8] Y. Wei, B. Akinci, Panorama-to-model registration through integration of image retrieval and semantic reprojection, Automation in Construction 140 (2022) 104356. https://doi.org/10.1016/j.autcon.2022.104356.

[9] E. Rublee, V. Rabaud, K. Konolige, G. Bradski, ORB: An efficient alternative to SIFT or SURF, in: 2011 International Conference on Computer Vision, 2011: pp. 2564–2571. https://doi.org/10.1109/ICCV.2011.6126544.

[10] S.A.K. Tareen, Z. Saleem, A comparative analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK, in: 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), 2018: pp. 1–10. https://doi.org/10.1109/ICOMET.2018.8346440.

[11] T. Yashiro, H. Hirayama, K. Sakamura, An Indoor Localization Service using 360 Degree Spherical Camera, in: 2020 IEEE 2nd Global Conference on Life Sciences and Technologies (LifeTech), 2020: pp. 17–18. https://doi.org/10.1109/LifeTech48969.2020.1570617174.

[12] H. Xu, Q. Zhao, Y. Ma, S. Wang, C. Yan, F. Dai, Free-Viewpoint Navigation of Indoor Scene with 360° Field of View, Electronics 12 (2023) 1954. https://doi.org/10.3390/electronics12081954.

[13] J. An, D.H. Lee, H.H. Cho, O.H. Jeong, Indoor Positioning System Using Smartphone and 360° Camera, in: 2021 IEEE International Conference on Smart Internet of Things (SmartIoT), 2021: pp. 342–343. https://doi.org/10.1109/SmartIoT52359.2021.00062.

[14] S. Thrun, L. Pratt, Learning to Learn, Springer Science & Business Media, 2012.

[15] Z. Cui, Q. Wang, J. Guo, N. Lu, Few-shot classification of façade defects based on extensible classifier and contrastive learning, Automation in Construction 141 (2022) 104381. https://doi.org/10.1016/j.autcon.2022.104381.

[16] D. Gil, G. Lee, Zero-shot monitoring of construction workers' personal protective equipment based on

image captioning, Automation in Construction 164 (2024) 105470. https://doi.org/10.1016/j.autcon.2024.105470.
[17] J. Kim, S. Chi, A few-shot learning approach for database-free vision-based monitoring on construction sites, Automation in Construction 124 (2021) 103566. https://doi.org/10.1016/j.autcon.2021.103566.
[18] S. Leutenegger, M. Chli, R.Y. Siegwart, BRISK: Binary Robust invariant scalable keypoints, in: 2011 International Conference on Computer Vision, 2011: pp. 2548–2555. https://doi.org/10.1109/ICCV.2011.6126542.
[19] P. Alcantarilla, J. Nuevo, A. Bartoli, Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces, in: Procedings of the British Machine Vision Conference 2013, British Machine Vision Association, Bristol, 2013: p. 13.1-13.11. https://doi.org/10.5244/C.27.13.
[20] D.G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, International Journal of Computer Vision 60 (2004) 91–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94.
[21] H. Bao, L. Dong, S. Piao, F. Wei, BEiT: BERT Pre-Training of Image Transformers, (2022). https://doi.org/10.48550/arXiv.2106.08254.
[22] B. Zhou, H. Zhao, X. Puig, T. Xiao, S. Fidler, A. Barriuso, A. Torralba, Semantic Understanding of Scenes through the ADE20K Dataset, (2018). https://doi.org/10.48550/arXiv.1608.05442.
[23] R. Suvorov, E. Logacheva, A. Mashikhin, A. Remizova, A. Ashukha, A. Silvestrov, N. Kong, H. Goka, K. Park, V. Lempitsky, Resolution-robust Large Mask Inpainting with Fourier Convolutions, (2021). https://doi.org/10.48550/arXiv.2109.07161.
[24] G. Koch, R. Zemel, R. Salakhutdinov, Siamese neural networks for one-shot image recognition, in: ICML Deep Learning Workshop, 2015. https://www.cs.cmu.edu/~rsalakhu/papers/oneshot1.pdf.

Human Pose Estimation using Automated Multi-Camera Calibration

Israt Sharmin Dola[1], Inbae Jeong[2], Youjin Jang[3], and Moein Younesi Heravi[4]

[1,2] Dept. of Mechanical Engineering, North Dakota State Univ., Fargo, USA

[3,4] Dept. of Civil, Construction and Environmental Engineering, North Dakota State Univ., Fargo, USA

israt.dola@ndsu.edu, inbae.jeong@ndsu.edu, y.jang@ndsu.edu, moein.younesiheravi@ndsu.edu

Abstract

3D human joint estimation is essential for enabling effective human-robot interaction in construction automation, facilitating precise monitoring of worker movements to enhance safety, ergonomics, and operational efficiency. Vision-based systems utilizing multi-camera setups offer diverse perspectives to address challenges such as occlusions, projection ambiguities, and sensor noise. However, these systems depend heavily on accurate camera calibration to align views and synchronize measurements. Traditional manual calibration methods are time-consuming, labor-intensive, and prone to human error, making them unsuitable for dynamic construction sites where frequent camera repositioning is required due to shifting conditions and tasks. This study proposes a novel framework that leverages external marker detection for real-time, automated camera calibration, eliminating the need for manual intervention. This approach significantly reduces setup time, minimizes errors, and ensures reliable performance in rapidly changing environments. Furthermore, the framework integrates an Extended Kalman Filter (EKF) to fuse 2D joint locations from multiple cameras, effectively handling sensor noise and the nonlinear nature of human motion. By combining marker-based calibration with EKF-based fusion, the proposed framework delivers a robust and automated solution for 3D human joint estimation, enhancing safety, efficiency, and adaptability in construction automation applications.

Keywords – 3D Human Joint Estimation; Multi Camera System; Automated Camera Calibration; Extended Kalman Filter

1 Introduction

Construction automation has emerged as a transformative approach to improving efficiency, safety, and ergonomics on job sites. Central to this advancement is the ability to accurately recognize human activities and estimate 3D poses, enabling seamless interaction between workers and automated systems. Modern construction sites increasingly feature human-robot collaboration, where workers and autonomous machines must operate in close proximity to accomplish complex tasks. However, these environments present significant hazards, including heavy machinery, work at heights, uneven terrain, and dynamic interactions between workers and automated systems [1], [2], [3]. In such settings, precise monitoring of worker movements is essential not only for ensuring safe interactions but also for optimizing task coordination and preventing errors. By reconstructing human movements through 3D joint estimation, construction managers can identify unsafe behaviors, assess ergonomic risks, and refine workflow strategies. This capability enhances safety protocols, task efficiency, and human-robot collaboration, reducing the likelihood of injuries from improper posture, repetitive strain, or equipment misuse [4].

Various 3D joint estimation approaches leverage advanced sensing, filtering, and modeling techniques. Among these, vision-based methods are widely adopted, utilizing camera data to track joint positions in three dimensions. RGB-D and 3D cameras help overcome the dimensional limitations of 2D cameras by providing depth information [5], [6], [7]. However, such systems often struggle with resolution constraints and sensitivity to lighting variations [8], [9], [10]. Recent monocular 3D pose estimation techniques have advanced significantly by leveraging Convolutional Neural Networks (CNNs) to regress 3D joint locations from single RGB images [11]. These approaches have expanded the application of vision-based motion capture to unconstrained

environments, but real-world deployment remains challenging due to occlusion, motion blur, and variable lighting conditions, necessitating further improvements.

While monocular methods offer promising results, they often struggle with depth ambiguities. Multi-camera setups address this limitation by capturing multiple viewpoints of a subject, enabling robust triangulation of joint positions. However, such setups require accurate camera calibration to ensure proper alignment across views. Once calibrated, their performance can be enhanced using data fusion techniques such as the Kalman Filter (KF) and Extended Kalman Filter (EKF) to mitigate sensor noise and missing data, while particle filters are particularly effective in handling nonlinearities and occlusions [12], [13], [14]. Additionally, deep learning-based approaches, including Recurrent Neural Networks (RNNs), Generative Adversarial Networks (GANs), and sequence-to-sequence models, further improve motion prediction and temporal consistency in pose estimation [15], [16], [17].

Despite these advancements, significant challenges remain, particularly in dynamic and complex environments like construction. A major limitation is the reliance on accurate camera calibration, which is essential for precise depth perception and spatial alignment. Misaligned views due to improper calibration compromise depth calculations, reducing 3D estimations accuracy. Traditional calibration methods, such as the Direct Linear Transformation and two-step approaches, provide high precision in estimating intrinsic and extrinsic camera parameters. However, they are manual, time-consuming, and best suited for controlled environments, making them impractical for dynamic and unpredictable construction settings [18], [19].

To address these limitations, researchers have explored alternative approaches. Active vision-based methods leverage controlled camera movements to simplify mathematical modeling, but their dependency on precise motion control makes them challenging to implement in real-world applications. Similarly, advancements in computational techniques, such as neural networks combined with global optimization algorithms, have shown promise for automated calibration. These methods effectively model nonlinear relationships in calibration, but their reliance on large training datasets, intensive computation, and careful initialization can limit their applicability in time-sensitive or resource-constrained settings [20], [21].

While these techniques perform well in controlled environments, they encounter significant challenges in dynamic settings where frequent recalibration is required. On construction sites, cameras must be regularly repositioned to accommodate shifting tasks, occlusions, changing lighting conditions, and complex human motion patterns. Manual recalibration in such scenarios is time-consuming and inefficient, emphasizing the need for a real-time, automated calibration system that can maintain accuracy without disrupting ongoing operations.

In this study, we propose an automated calibration framework that integrates a real-time, self-updating multi-camera calibration process for 3D joint estimation. Unlike traditional methods, where camera calibration is performed separately before pose estimation, our framework simultaneously calibrates camera positions, orientations, and fields of view (FoV) while estimating 3D joint positions, enhancing both efficiency and adaptability. To achieve this, we employ an EKF-based multi-view fusion approach, where 2D joint detections from multiple cameras are integrated to generate robust 3D pose estimates, effectively reducing sensor noise and motion uncertainties. Additionally, we introduce an external marker-based calibration technique that automatically updates camera parameters throughout the estimation process, eliminating the need for manual recalibration, which is often time-consuming and prone to errors. This approach minimizes setup errors, enhances real-time adaptability, and ensures consistent calibration even in dynamic and complex applications such as construction automation.

2 Methodology

This study aims to simultaneously estimate the 3D joint positions of a human and automatically calibrate a multi-camera system. To achieve this, a state vector, X_t is defined at a given time t representing all the unknown variables to be estimated: human 3D joint positions and camera parameters (position, orientation, and FoV).

The 3D joint positions are denoted as $H_{t,j}$, where $j \in \{1, \dots n\}$, and n represents the total number of joints in the human body. Each $H_{t,j}$ specifies the global 3D coordinates of the j-th joint in a Cartesian coordinate system. The camera parameters are expressed as $C_{t,i}$, where $i \in \{1, \dots m\}$, and m denotes the total number of cameras. Each $C_{t,i}$ includes both the extrinsic parameters (3D position and orientation) and intrinsic parameters (horizontal and vertical fields of view) of the i-th camera.

The state vector, X_t is defined as:

$$X_t = \left[H_{t,1}, H_{t,2} \cdots H_{t,n}, C_{t,1}, C_{t,2} \cdots C_{t,m}\right]^T \quad (1)$$

Here, each joint position $H_{t,j}$ is a 3D vector representing the global coordinates $(x_{h,j}, y_{h,j}, z_{h,j})$ of the j-th joint:

$$H_{t,j} = [x_{h,j} \quad y_{h,j} \quad z_{h,j}]^T \quad (2)$$

Similarly, the camera parameter $C_{t,i}$ represents the 3D position $(x_{c,i}, y_{c,i}, z_{c,i})$, orientation defined by roll $\varphi_{c,i}$, pitch $\varphi_{c,i}$, yaw $\psi_{c,i}$, as well as the horizontal ($F_{h_{c,i}}$) and vertical ($F_{v_{c,i}}$) FoV:

$$C_{t,i} = \left[x_{c,i}, y_{c,i}, z_{c,i}, \varphi_{c,i}, \varphi_{c,i}, \psi_{c,i}, F_{h_{c,i}}, F_{v_{c,i}}\right]^T \quad (3)$$

This formulation serves as the foundation for simultaneously refining 3D joint positions and calibrating the multi-camera system. The proposed methodology, illustrated in Figure 1, comprises three main steps: 2D joint estimation, marker detection and alignment, and EKF-based integration. 2D joint estimation extracts pixel coordinates of human joints, while marker detection establishes a global reference frame through ArUco marker corners. These measurements feed into the EKF, which refines the state vector, X_t to estimate 3D joint positions and camera parameters, ensuring precise alignment. The following sections provide a detailed explanation of each step.

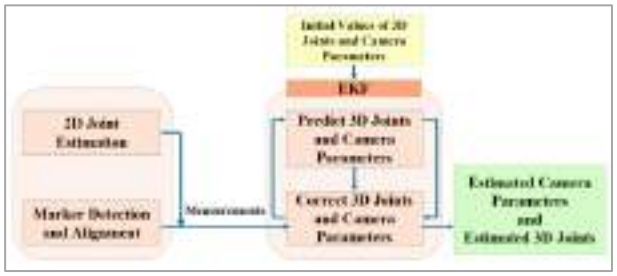


Figure 1. Overall framework

2.1 2D Joint Estimation

The performance of the 3D joint estimation largely depends on accurate 2D joint detection. This study employs MediaPipe's Pose module; a deep learning-based framework developed by Google for high-accuracy, real-time application [22]. It utilizes the lightweight BlazePose model, which features a detector-tracker architecture that reduces latency by tracking previously detected keypoints rather than re-detecting them in each frame. BlazePose outputs 33 keypoints [23]; but this research selects 15 primary keypoints, including major body joints (shoulders, elbows, wrists, hips, knees, and ankles) and key facial landmarks, while excluding keypoints that are not essential for pose estimation [2]. This selection provides a comprehensive 2D representation of the human body, as shown in Figure 2.

Before 2D joint estimation, images from all cameras are synchronized to ensure temporal alignment across views for consistent detection. Each frame is processed individually, with the model outputting pixel coordinates and a confidence score (0–1) for each keypoint. To enhance accuracy, only keypoints with confidence scores above 0.8 are retained, reducing false detections from occlusions or challenging poses. The extracted 2D joint data is structured into arrays for each camera view, enabling synchronized multi-camera processing. This serves as input for the fusion algorithm, facilitating precise 3D joint reconstruction. Leveraging MediaPipe's efficient architecture, this approach achieves low latency, making it ideal for real-time multi-view analysis.

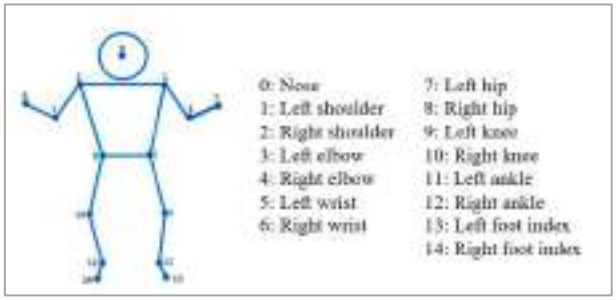


Figure 2. Defined 2D human joints

2.2 Marker Detection and Alignment

In this study, ArUco markers are chosen for their unique binary patterns, which enable easy identification and robust tracking. Unlike checkerboards or circular grids, ArUco markers are resilient to moderate lighting variations and can be reliably detected from diverse angles and distances, making them ideal for multi-camera setups [24].

For this phase, an ArUco marker with a predefined size is used as the reference for automatic camera calibration. The bottom-left corner of the marker is defined as the origin of the global coordinate system, $(0,0,0)$. Based on the marker's size, s, the 3D coordinates of the other corners are defined as $(s,0,0)$, $(0,s,0)$ and $(s,s,0)$. These fixed coordinates serve as reference points for aligning all camera views within a unified coordinate system.

The process begins with detecting the precise corner coordinates of the marker in the camera image planes. These detected coordinates serve as observed measurements for the fusion algorithm, which subsequently estimates both extrinsic and intrinsic camera parameters. The extrinsic parameters, represented by the 3×4 transformation matrix $[R|t]$, define each camera's orientation, R relative to the marker and its position, t in 3D space. Meanwhile, the intrinsic parameters such as FoV, influence how 3D points are projected onto the image plane. This simultaneous estimation ensures that all camera views are accurately aligned within a unified global coordinate system, reducing projection errors and improving multi-view consistency.

2.3 Estimation of Human Pose and Camera Parameter

EKF serves as the core algorithm for simultaneously estimating the 3D human joint positions and the intrinsic

and extrinsic parameters of the multi-camera system by integrating synchronized 2D joint detections and marker-based measurements to iteratively refine the state vector X_t (Equation (1)), which encapsulates both the human joint positions and the camera parameters.

The EKF follows a prediction-correction cycle. In the prediction phase, the system propagates the state and uncertainty forward using a process model, where each joint position evolves independently over time, modeled as:

$$x_{t+1,j} = x_{t,j} + \eta_t \tag{4}$$

where $\eta_{t,j} \sim \mathcal{N}(0, Q_t)$ represents Gaussian process noise, with Q_t as a covariance matrix capturing uncertainties for each joint and camera. The prediction step maintains a Gaussian assumption centered around the current estimate while updating the state distribution over time. Unlike methods relying on pre-known 3D joint positions or calibrated camera parameters, our approach uses the EKF framework to estimate both the joint positions and camera parameters dynamically. The EKF estimates and refines these parameters through its fusion of multi-camera observations and marker-based measurements, ensuring adaptability to dynamic environments.

In the correction phase, 2D joint detections and marker corner coordinates are incorporated to refine the state vector. The observation model predicts the expected pixel coordinates (u_e, v_e) for both joints and marker corners based on the current state estimate. Figure 3 illustrates how a point (joint or marker corner) is transformed from the world coordinate system to the camera coordinate system and finally to the pixel coordinate system.

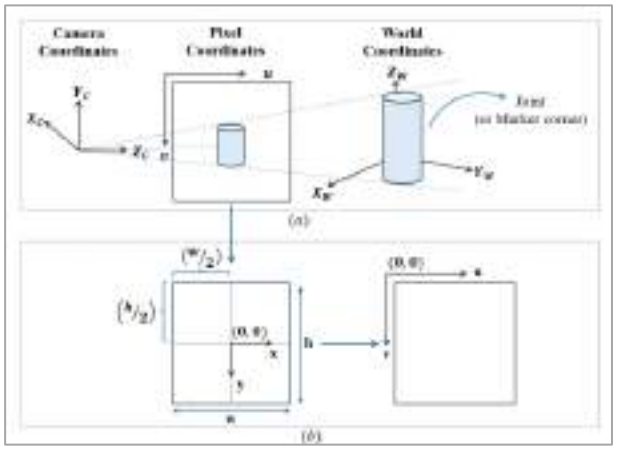


Figure 3. Imaging geometry for joint or marker corner (a) projection joint or marker onto image plane (b) joint or marker in pixel coordinates

To map global joint positions to each camera i's local frame, the EKF employs a transformation matrix T_i and a rotation matrix R_i which is defined as:

$$v_{h,j|C_i} = R_i^{-1} \cdot T_i^{-1} \cdot \mathrm{v}_{h,\mathrm{j}|\mathrm{G}} \tag{5}$$

where $\mathrm{v}_{h,\mathrm{j}|\mathrm{G}}$ and $\mathrm{v}_{h,\mathrm{j}|\mathrm{C_i}}$ represent the j-th joint location in the global frame and in the camera i's frame respectively. They are defined as:

$$\mathrm{v}_{h,\mathrm{j}|\mathrm{G}} = \left[x_{h,j}, y_{h,j}, z_{h,j}, 1\right]^T \tag{6}$$

$$v_{h,j|C_i} = \left[x_{h,cam.j}, y_{h,cam,j}, z_{h,cam,j}, 1\right]^T \tag{7}$$

Similarly, each corner of the marker is mapped to each camera i's local frame:

$$v_{m,k|C_i} = R_i^{-1} \cdot T_i^{-1} \cdot \mathrm{v}_{m,\mathrm{k}|\mathrm{G}} \tag{8}$$

where $\mathrm{v}_{m,\mathrm{k}|\mathrm{G}}$ and $\mathrm{v}_{m,\mathrm{k}|\mathrm{C_i}}$ represent the marker's k-th corner location in the global frame and in the camera i's frame respectively. They are defined as:

$$\mathrm{v}_{m,\mathrm{k}|\mathrm{G}} = \left[x_{m,k}, y_{m,k}, z_{m,k}, 1\right]^T \tag{9}$$

$$v_{m,k|C_i} = \left[x_{m,cam.k}, y_{m,cam,k}, z_{m,cam,k}, 1\right]^T \tag{10}$$

The transformation matrix T_i, where $(x_{c,i}, y_{c,i}, z_{c,i})$ represents the translation of camera i in the global coordinate system. is defined as:

$$T_i = \begin{bmatrix} 1 & 0 & 0 & x_{c,i} \\ 0 & 1 & 0 & y_{c,i} \\ 0 & 0 & 1 & z_{c.i} \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{11}$$

The rotation matrix R_i accounts for the camera's orientation, combining rotations along the $x, y,$ and z axes:

$$R_i = R_z(\theta_{c,i}) R_y(\phi_{c,i}) R_x(\psi_{c,i}) \tag{12}$$

where $\theta_{c,i}$, $\phi_{c,i}$, and $\psi_{c,i}$ are the roll, pitch, and yaw angles of camera i, respectively. These transformations map global 3D positions to the local coordinate system of each camera, resulting in the camera-frame coordinates $(x_{cam}, y_{cam}, z_{cam})$.

The transformed 3D coordinates are then projected onto the image plane to compute the expected pixel coordinates (u_e, v_e) normalized by the image width (w) and height (h) to ensure values range between 0 and 1. Using the horizontal and vertical FoV (FoV_h and FoV_v) the equations for the expected pixel coordinates are:

$$u_e = \frac{1}{2}\left(1 + \frac{x_{cam}}{z_{cam}\, tan\left(\frac{FoV_h}{2}\right)}\right) \tag{13}$$

$$v_e = \frac{1}{2}\left(1 + \frac{y_{cam}}{z_{cam}\, tan\left(\frac{FoV_v}{2}\right)}\right) \tag{14}$$

In this formulation, we use the pinhole camera model,

detailed in figure 4, where FoV inherently defines the camera's intrinsic properties seamlessly integrating with our projection model. This approach eliminates the need for focal length as a separate parameter, making the formulation more compact and efficient while maintaining accuracy in projection calculations. This formulation is applicable to both human joints and marker corners, denoted as $\left(u_{e_{h,j}}, v_{e_{h,j}}\right)$ for the j th human joint and $\left(u_{e_{m,k}}, v_{e_{m,k}}\right)$ for the k-th marker corner.

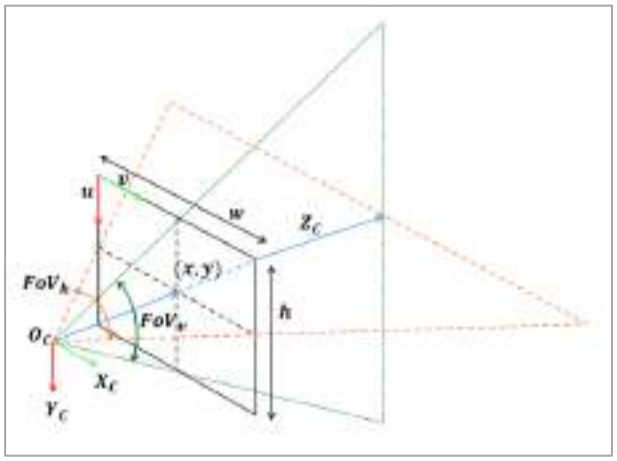


Figure 4. Expected measurements from planar projection

For n cameras, the expected measurement vector $z_{t,e}$ combines the expected 2D pixel coordinates of the detected human joints, $z_{t,e,h}$ and marker corners, $z_{t,e,m}$ from each camera i. It is defined as:

$$z_{t,e} = \left[z_{t,e,h}, z_{t,e,m}\right]^T \quad (15)$$

$$z_{t,e,h} = \left[u_{e,h,i,1}, v_{e,h,i,1}, \ldots, u_{e,h,i,n}, v_{e,h,i,n}\right]^T \quad (16)$$

$$z_{t,e,m} = \left[u_{e,m,i,1}, v_{e,m,i,1}, \ldots u_{e,m,i,k}, v_{e,m,i,k}\right]^T \quad (17)$$

Here $u_{e,h,i,j}$ and $v_{e,h,i,j}$ are the expected pixel coordinates of joint j in camera i's image plane, and $u_{e,m,i,k}$ and $v_{e,m,i,k}$ are the pixel coordinates of marker k's corners.

Once the actual measurements $z_{t,a}$, obtained directly from sensors as noisy 2D pixel coordinates of detected joints and marker corners, are available, the measurement residual,y_t is computed:

$$y_t = z_{t,a} - z_{t,e} \quad (18)$$

where $z_{t,a}$ is derived from the outputs of the 2D joint detection and marker detection algorithms steps discussed earlier in the methodology, while $z_{t,e}$ represents the expected measurements predicted by the observation model.

To incorporate this residual into the state update, the Kalman gain K_t is computed as:

$$K_t = P_t H_t^T (H_t P_t H_t^T + R_t)^{-1} \quad (19)$$

where H_t is the Jacobian of the observation model relative to the state vector, capturing the sensitivity of the measurements to changes in state. R_t is the measurement noise covariance matrix which captures the uncertainty in each camera's observation and P_t is the predicted state covariance matrix, representing uncertainty prior to incorporating the measurements.

In the measurement update step, the EKF integrates the observed measurements $z_{t,a}$. The updated state x_{t+1} is computed as:

$$x_{t+1} = x_{t+1}^- + K_t y_t \quad (20)$$

and the updated state covariance, P_{t+1} is:

$$P_{t+1} = (I - K_t H_t) P_{t+1}^- \quad (21)$$

where x_{t+1}^- and P_{t+1}^- the predicted state and covariance before the measurement update.

This EKF-based filtering algorithm effectively combines predicted and observed data, yielding precise and robust 3D joint estimations and camera parameter calibrations in dynamic, multi-camera environments.

3 Experimental Setup

To evaluate the proposed EKF-based multi-camera calibration and 3D joint estimation system, the implementation was carried out in the Webots simulation environment. Webots was chosen for its ability to replicate real-world scenarios in a controlled setting, providing access to ground truth data, enabling repetitive and reproducible experiments, and allowing systematic testing of algorithms under various configurations.

The simulation was designed to emulate a real-world multi-camera setup with four static cameras and a 5 cm ArUco marker. A pedestrian proto model simulated a human subject, with predefined joint positions replicating realistic human motion. Planar projection was utilized to simulate the 2D image planes of the cameras, enabling accurate marker and joint detections essential for subsequent calibration and estimation steps. The key parameters of the simulation environment are as follows:

- Image Resolution: 640 × 640 pixels.
- Marker Size: 5 cm (edge length).
- Camera Configuration: Planar layout with four static cameras.
- Subject: Webots pedestrian proto model.
- Number of Trials: 30.

The evaluation focused on quantifying the accuracy of the estimated 3D joint positions, camera positions,

camera orientations, and camera FoVs. The Root Mean Square (RMS) error was used as the primary metric for all evaluations:

$$RMS = \sqrt{\frac{1}{N \cdot K} \sum_{i=1}^{N} \sum_{k=1}^{K} \left\| x_{i,k}^{tr} - x_{ik}^{est} \right\|^2} \qquad (22)$$

where, N is the total number of measurements, K is the total number of cameras, $x_{i,k}^{tr}$ and x_{ik}^{est} are the true and estimated values of the quantity being evaluated. x can represent joint positions, camera positions, camera orientations, or FoVs.

By simulating a multi-camera system with realistic projections, human motion, and marker detections, the experimental setup successfully replicated the challenges of real-world environments.

4 Results

This section presents the proposed methodology's findings in the Webots simulation environment. Figure 5 illustrates the successful detection of joints in the pedestrian model and all four corners of the ArUco marker in the image planes from all four cameras. The joint detection provides 2D coordinates for estimating 3D joint positions, while marker corner detection establishes a global reference for camera calibration by defining a consistent coordinate system.

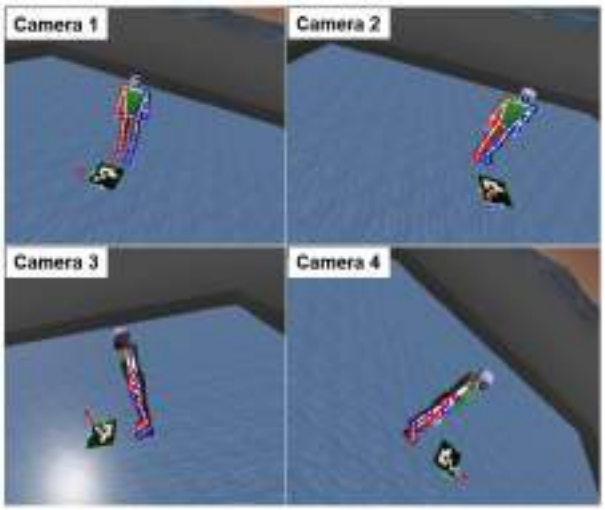


Figure 5. Detection of joints and ArUco marker corners across all camera views

The performance of the proposed methodology was evaluated by analyzing the error trends for joint positions, camera positions, camera orientations, and camera FoV. The results showed an RMS error of 0.18m for joint positions and $0.32m$ for camera positions, while the RMS errors for camera orientations and FoV were approximately 11° and 9° , respectively. Figure. 6 illustrates the error trends for joint positions, camera positions, camera orientations, and camera FoV over time. The graphs show a steady reduction in errors as the EKF iteratively refines the state estimates using multi-camera observations and marker-based measurements.

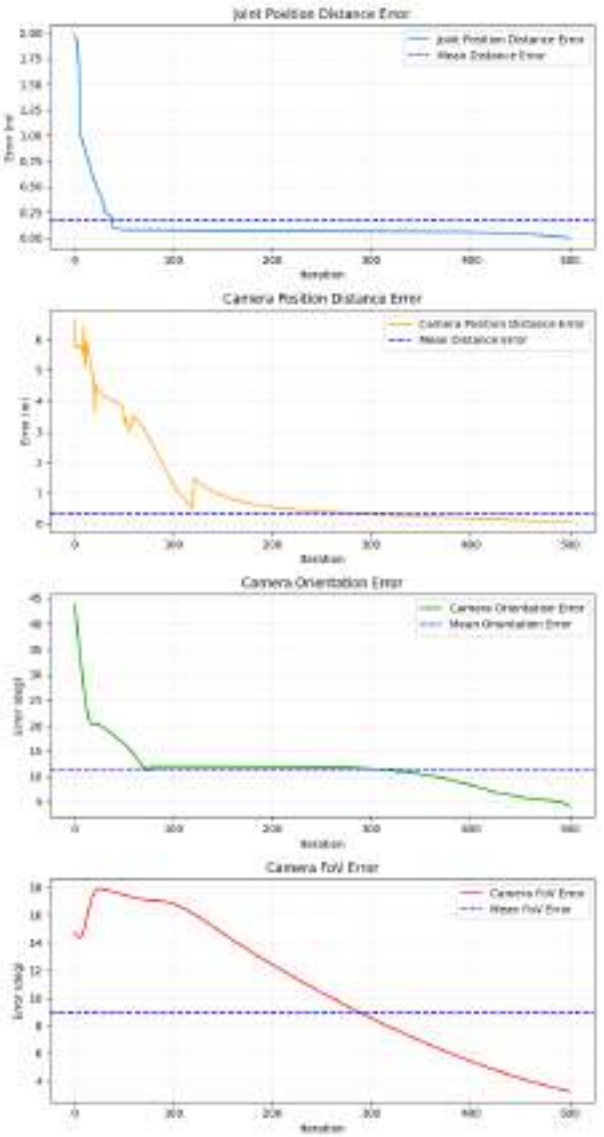


Figure 6. Error trends of joint positions and camera parameters

These RMS errors indicate that the proposed system achieved automatic camera calibration for 3D joint estimation. The consistently small deviations across trials highlight the robustness of the methodology in handling challenges such as sensor noise, nonlinearities, and multi-camera synchronization. Combined with the error trends shown in Figure 6, these findings validate the

reliability and accuracy of the framework in achieving its objectives in a controlled simulation environment.

5 Conclusion

This study proposed a novel framework for automatic camera calibration for 3D joint estimation using a multi-camera system. By integrating an EKF for multi-camera data fusion and an external marker-based calibration method, the framework eliminates the need for manual calibration. Simulations conducted in Webots demonstrated its capability to accurately estimate 3D joint positions and maintain reliable calibration with four cameras. Results showed that the automated calibration method reduced reliance on manual interventions, highlighting the framework's potential for real-world applications where rapid deployment and adaptability are critical.

Despite promising results, certain limitations present opportunities for future research. Simulations, while providing controlled and reproducible results, may not fully capture real-world complexities such as dynamic lighting, occlusions, and irregular motion patterns. Future work should validate the framework in real-world settings to assess its robustness. While our approach models intrinsics via FoV, camera parameter calibration remains a challenge, as real-world systems often require calibration of focal length, principal point, and distortion coefficients. Additionally, the reliance on external markers may pose challenges in cluttered environments where occlusions impact visibility. Optimizing marker placement, increasing marker count, or exploring marker-free calibration could enhance performance. Scalability is another concern, as real-time EKF updates and multi-camera processing demand high computational resources. Leveraging parallel processing or Edge AI acceleration could improve efficiency for real-time deployment. Lastly, while EKF performed well, exploring alternative filtering techniques could further improve accuracy or reduce computational overhead.

With these advancements, the proposed framework has the potential to evolve into a robust and adaptable tool for enhancing worker safety and efficiency in high-risk, dynamic environments. This progress would pave the way for more intelligent and scalable systems in construction automation, furthering the integration of AI and robotics in real-world applications.

Acknowledgement

The authors express their gratitude for the support provided by the North Dakota Economic Diversification Research Fund and the National Science Foundation (NSF) EPSCoR RII Track 2 Program under the grant number OIA-2119691. The findings and opinions expressed in this article are those of the authors only and do not necessarily reflect the views of the sponsors.

References

[1] I. Jeong, Y. Jang, J. Park, and Y. K. Cho, "Motion Planning of Mobile Robots for Autonomous Navigation on Uneven Ground Surfaces," Journal of Computing in Civil Engineering, vol. 35, no. 3, May 2021, ASCE, doi: 10.1061/(asce)cp.1943-5487.0000963.

[2] Y. Jang, I. Jeong, M. Younesi Heravi, S. Sarkar, H. Shin, and Y. Ahn, "Multi-Camera-Based Human Activity Recognition for Human–Robot Collaboration in Construction," Sensors, vol. 23, no. 15, Aug. 2023, MDPI, doi: 10.3390/s23156997.

[3] K. Kim and Y. K. Cho, "Effective inertial sensor quantity and locations on a body for deep learning-based worker's motion recognition," Automous Construction, vol. 113, May 2020, Elsevier, doi: 10.1016/j.autcon.2020.103126.

[4] F. Hoenigsberger et al., "Machine Learning and Knowledge Extraction to Support Work Safety for Smart Forest Operations," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer Science and Business Media Deutschland GmbH, 2022, pp. 362–375. doi: 10.1007/978-3-031-14463-9_23.

[5] J. Wang, L. Liu, W. Xu, K. Sarkar, and C. Theobalt, "Estimating Egocentric 3D Human Pose in Global Space." 2021. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 11500–11509. IEEE. doi: 10.1109/ICCV48922.2021.01130.

[6] D. Tome et al., "SelfPose: 3D Egocentric Pose Estimation From a Headset Mounted Camera," IEEE Trans Pattern Anal Mach Intell, vol. 45, no. 6, pp. 6794–6806, Jun. 2020, doi: 10.1109/TPAMI.2020.3029700.

[7] J. Shotton et al., "Real-Time Human Pose Recognition in Parts from Single Depth Images," 2011. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2011), Colorado Springs, CO, USA, pp. 1297–1304. IEEE. doi.org/10.1109/CVPR.2011.5995316.

[8] Zequn Zhang, Yuanning Liu, Ao Li, and Minghui Wang, A Novel Method for User-Defined Human Posture Recognition Using Kinect. Proceedings of the 7th International Congress on Image and Signal

Processing (CISP 2014), Dalian, China, pp. 736–740. IEEE, 2014. doi.org/10.1109/CISP.2014.7003875.

[9] L. Minh Dang, K. Min, H. Wang, M. Jalil Piran, C. Hee Lee, and H. Moon, "Sensor-based and vision-based human activity recognition: A comprehensive survey," Pattern Recognit, vol. 108, Dec. Elsevier. 2020, doi: 10.1016/j.patcog.2020.107561.

[10] A. Jalal, Y. H. Kim, Y. J. Kim, S. Kamal, and D. Kim, "Robust human activity recognition from depth video using spatiotemporal multi-fused features," Pattern Recognit, vol. 61, pp. 295–308, Jan. 2017, Elsevier. doi: 10.1016/j.patcog.2016.08.003.

[11] D. Mehta et al., "Monocular 3D human pose estimation in the wild using improved CNN supervision," In Proceedings of the 2017 International Conference on 3D Vision (3DV), pp. 506–516. IEEE. doi: 10.1109/3DV.2017.00064.

[12] M. Gholami, A. Rezaei, H. Rhodin, R. Ward, and Z. J. Wang, "TriPose: A Weakly-Supervised 3D Human Pose Estimation via Triangulation from Video," arXiv preprint, arXiv:2105.06599. [Online]. Available: http://arxiv.org/abs/2105.06599.

[13] L. Chen, H. Ai, R. Chen, Z. Zhuang, and S. Liu, "Cross-view tracking for multi-human 3D pose estimation at over 100 FPS." In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, pp. 3276–3285. IEEE. doi:10.1109/CVPR42600.2020.00334.

[14] X. Li, Z. Fan, Y. Liu, Y. Li, and Q. Dai, "3D pose detection of closely interactive humans using multi-view cameras," Sensors (Switzerland), 19(12), 2831. MDPI. doi: 10.3390/s19122831.

[15] L.-Y. Gui, Y.-X. Wang, X. Liang, and J. M. F. Moura, "Adversarial Geometry-Aware Human Motion Prediction," 2018. In Proceedings of the European Conference on Computer Vision (ECCV), pp. 786–803. Springer. .

[16] A. Jain, A. R. Zamir, S. Savarese, and A. Saxena, "Structural-RNN: Deep learning on spatio-temporal graphs," 2016. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5308–5317. IEEE. doi: 10.1109/CVPR.2016.573.

[17] J. Martinez, M. J. Black, and J. Romero, "On human motion prediction using recurrent neural networks," 2017. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017), pp. 4674–4683. IEEE. doi: 10.1109/CVPR.2017.497.

[18] Z. Qi, L. Xiao, S. Fu, T. Li, G. Jiang, and X. Long, "Two-step camera calibration method based on the SPGD algorithm," Applied Optics, 51(26), 6421–6428. Optica Publishing Group. doi: 10.1364/ao.51.006421.

[19] L. Song, W. Wu, J. Guo, and X. Li, "Survey on camera calibration technique," 2013. In Proceedings of the 2013 5th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC 2013), pp. 389–392. IEEE. doi: 10.1109/IHMSC.2013.240.

[20] D. Xianzhi, "Research on Camera Calibration Technology Based on Deep Neural Network in Mine Environment," 2020. In Proceedings of the 2020 International Conference on Computer Vision, Image and Deep Learning (CVIDL 2020), pp. 375–379. IEEE. doi: 10.1109/CVIDL51233.2020.00-68.

[21] Wang Qi, Fu Li, and Liu Zhenzhong, "Review on Camera Calibration," 2010. In Proceedings of the 2010 Chinese Control and Decision Conference (CCDC), Xuzhou, China, pp. 3354–3358. IEEE. doi.org/10.1109/CCDC.2010.5498574.

[22] C. Lugaresi et al., "MediaPipe: A Framework for Building Perception Pipelines," 2019. arXiv preprint, arXiv:1906.08172. [Online]. Available: http://arxiv.org/abs/1906.08172.

[23] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, and M. Grundmann, "BlazePose: On-device Real-time Body Pose tracking," Jun. 2020, arXiv preprint, arXiv:2006.10204. [Online]. Available: http://arxiv.org/abs/2006.10204.

[24] M. Kalaitzakis, B. Cain, S. Carroll, A. Ambrosi, C. Whitehead, and N. Vitzilaios, "Fiducial Markers for Pose Estimation: Overview, Applications and Experimental Comparison of the ARTag, AprilTag, ArUco and STag Markers," Journal of Intelligent and Robotic Systems: Theory and Applications, vol. 101, no. 4, Apr. 2021, Springer. doi: 10.1007/s10846-020-01307-9.

LiLoc – Defect Image Localization in Laser Scans

Patrick Herbers[1], Lisa Freifrau von Rössing[1], and Markus König[1]

[1]Chair of Computing in Engineering, Ruhr University Bochum, Bochum, Germany

patrick.herbers@rub.de, lisa.freiinvonroessing@rub.de, koenig@inf.bi.rub.de

Abstract -

Deterioration and maintenance are a constant part of the operational phase of a structures life cycle. Managing and analyzing a structure's history of damage and repair is essential for present operational decisions. Images are a central way of documenting the location of damages during maintenance, but often lack references to the structure. Mapping defects not only spatially, but also over time, can give inspectors a deep insight into the health of a structure. Therefore, we present *LiLoc*, a module for localizing images of inspection documentation in laser scans. Images are matched to the panoramic images of a laser scan, and matrix transformations for mapping points to laser scans are determined. Extracted matrices can be used to transform damage masks and other overlaying information. We then compare the algorithms on two bridge data sets.

Keywords -

Localization, Maintenance, Laser Scanning, Inspection, Defect Mapping

1 Introduction

The maintenance effort required for infrastructure is at an unprecedented high with many concrete structures in Europe and North America reaching the end of their life span. Among the various activities involved in infrastructure upkeep, inspections stand out as the most critical component for assessing the health and integrity of structures. This is especially true for concrete bridges, for which regular inspections throughout their lifespan are an absolute necessity for public safety.

Currently, the inspections in Germany are in most cases conducted by traversing the bridge, identifying old and new defects, documenting the bridge's current state with images and written notes, and finally summarizing the inspection results in a standardized report, consisting of text and images. The images document defects visually, and are used for referencing the defects in later inspections to enable the evaluation of defect progression. Textual information is used to further describe the defect, such as material and size. Usually, the textual information also describes the location of the defect, such as the building component identifier, and less often, road kilometer or cardinal direction. Due to a lack of detail, this recorded location is often imprecise. Finding the defect again is therefore difficult, even though it is required as a basis for the analysis of the structural condition and the defect progression. Enhancing the largely analog and inefficient inspection process through digital tools enables earlier detection of defects and defect progression, thereby providing a stronger foundation for making informed operational decisions.

One possible solution is creating a defect map, meaning a module for a Digital Twin (DT) of the structure that stores defect information and location over time. The planning and execution of maintenance measures for bridge structures benefits from the concept of DTs. Inspectors can easily locate defects with the defect map in a 3D model, while operators can efficiently analyze large amounts of data on a holistic level. Especially the comprehensive map of all defects on a structure over time improves the informational value of inspections by also giving exact spatial information. This defect map will save time during inspections and allow operators to make more informed decisions during the life cycle of a bridge. While a lot of data is available with many inspections reaching back decades, mapping defects from older or current inspections manually is however inefficient.

The major hurdle to overcome for automatically creating a defect map is transforming the data from existing databases. Since the location of defects is often only documented in inconsistent and imprecise textual descriptions, automatically placing a defect on a 3D model becomes challenging. Inspection images may contain hints to the location of a defect, but often lack context, consisting of close ups and zoomed-in pictures for a detailed documentation of defects. Thus, a way to localize images from historical databases is an essential step for the creation of a defect map.

Furthermore, a 3D model into which defects can be mapped is required. Capturing the as-built state of a structure by creating a digital model using laser scanning has been used in practice for several years. As laser scanning allows for the efficient generation of an as-built model, and images of defects are readily available, combining the two methods enables comprehensive defect mapping in conjunction with detailed visual and textual information.

Therefore, we propose an approach to map defects using one-time laser scans and images from multiple inspections through feature matching.

2 Related Works

Improving the mapping of defects is an important topic in structural health research, with a focus on accurately and automatically localizing defects in a 3D model. An often-used tool for concurrent 3D model creation and defect mapping are Unmanned Aerial Systems (UAS). For example, Lin et al. [1] use an UAS to take images of a bridge. The images are used to create a 3D model with Structure from Motion (SfM), and therefore first clustered based on GPS coordinates for to prevent mismatches due to the inherent self-similarity of bridge structures. Defects are then detected on the same images. Since the location of the latter within the 3D model is known, the detected defects can be mapped into the model easily. However, extensive flight planning is necessary to account for GPS inaccuracy. A similar approach is used by Tan et al. [2] for external building walls. Here, GPS is used to compute the UAS position, which is then used to map defects onto a Building Information Modelling (BIM) model.

Research has shown that localization with GPS can lack the accuracy required for defect mapping, especially in areas with little reception, such as underneath a bridge. Therefore, additional technology has been tested to increase localization and thereby mapping accuracy. Yoon et al. [3] utilize an UAS with LiDAR, GPS, and IMU to achieve an accuracy of 10 cm even underneath a bridge. The system is however only evaluated on flat surfaces, such as one side of a bridge column, and issues due to GPS inaccuracy are reported, such as image curving, distortion, and areas missed by the UAS. In another example, Zhao et al. [4] use an UAS with LiDAR, Ground Control Points (GCP) and Check Points (CP) to take images of a concrete dam. Next, cracks larger than 5 mm are detected and mapped with Scale-Invariant Feature Transform (SIFT) [5] and SfM. However, the authors note that the GCPs are critically necessary for reliably mapping the defects, but the setup of GCP and CP is time-intensive and therefore inefficient.

Alternative, non-GPS methods for mapping defects have been tested, mainly relying on camera data. Simultaneous localization and mapping (SLAM), specifically ORB-SLAM, is used on images taken by an UAS to localize defects underneath a longer bridge structure by Jiang et al. [6]. The images were taken from a distance of 2 m and the method results in a localization error of around 25 cm. Bartczak et al. [7] use images taken by an UAS for photogrammetric reconstruction of a 3D model by employing sequential image pairing, explicitly foregoing GPS due to inaccuracies. Defects are detected in the UAS images and thereby mapped to the resulting 3D model.

The previous literature shows the general suitability of using UAS for defect mapping. If an image is used for 3D model creation and defect detection concurrently, the defects are mapped into the resulting model automatically. However, UAS rely on GPS or other methods to localize themselves and thereby images. The previously presented studies show that these methods are still time- and/or cost-intensive, and prone to localization errors. Additionally, although the research is promising, there are further issues with the usage of UAS for defect mapping. First, UAS require a trained user for utilization, which adds further costs to the inspection process, as they cannot replace inspectors. UAS have to keep a safe distance to walls, which decreases the defect resolution in an image. In Xiao et al. [8] for example, close-up shots were taken, but could not be used for model generation successfully and were therefore discarded, leaving their data unavailable for defect detection. The detection of fine cracks with a width of 0.1 mm is required during inspections. Second, usage of UAS is highly regulated and can be restricted in environments such as the vicinity of critical infrastructure, which highway bridges are part of.

In summary, while the most common method of 3D model creation and defect mapping for bridges with UAS does have accuracy issues with regards to localization and resolution, it's not an unsuitable method, and it's apparent that it will play a major role in future inspection processes. However, due to usage restrictions and efficiency considerations, other methods of defect mapping are necessary. A simple alternative for 3D model creation are laser scans with portable scanners, which require little training and are explicitly encouraged in German inspection regulations [9]. Laser scanners usually record either panoramic or stitched images of their scanning pose to capture color information in the point cloud. While panoramic images can be taken during laser scans and are mapped to the scans point cloud, the images are not suitable for defect detection. Therefore, scans are usually not used for defect mapping.

Both laser scanning and UAS are used for as-built model creation but only consider the current state of the structure. A wealth of historical defect images exists, and new images are taken regularly during in-person inspections. Automatically mapping this defect data into a laser scan fulfills the need for a defect map. It also offers additional benefits, as the historical data is very valuable during inspections, e.g., to assess defect progression.

This historical data is not considered by any of the studies described previously, but the textual location descriptions have been considered the two following studies. Göbels and Beetz [10] localize defects via text from bridge defect inspection reports with an ontology called RELOC.

The localization is accurate to the structural component and the directions left/right/top/bottom, with the level of detail corresponding to the detail of the inspectors notes. Heise et al. [11] show the importance of the spatial location of defects in reports, and propose a linked data approach to road and bridge infrastructure data. Through mapping defect location between cross-domain systems, such as road and bridge databases with differing spatial reference systems, defect information can be referenced into other systems and historical data can be integrated. It is obvious that neither of these two works consider image data, which are not as prone to human error as text-based descriptions, or achieve an exact localization. As mentioned before, none of the other works discussed earlier consider historical data at all. Therefore, we focus on efficiently and accurately localizing current and historical defect images in present laser scans.

Since the historical data consists predominantly of individual images, image registration is a possible solution for mapping this data to the panoramic images in the laser scans. Image registration is the process of aligning two images of the same subject with different camera angles, resulting in a homography. One approach for this is key point based image matching. SIFT and adjacent methods extract comparable key points in both images, which are then used for image alignment. SIFT has previously been used for image retrieval in panoramic images, for example by Zamir and Shah [12].

In recent years, image matching using neural networks has become a dominant area of research. Neural network based key point extractors and matching methods, such as XFeat [13] and LightGlue [14] show a significant improvement in speed and accuracy over previous methods. A common research problem is image matching for panoramic images, such as street scenes or from laser scanners. Orhan and Bastanlar [15] use Convolutional Neural Networks (CNNs) to generate searchable feature sets on panoramic images by using a sliding window approach instead of rectifying multiple images. Kendall et al. [16] use a CNN for 6-DoF camera pose regression on a learned environment.

Image registration can be applied to all kinds of images. It is not limited to matching defect images with panoramic images, but can also map a historical image to a current image taken from a different angle. This is especially relevant considering UAS-image-based 3D models, into which in the future, historical defects can also be mapped.

3 LiLoc

We present *LiLoc*, a Python module for localizing images in LiDAR scans with the explicit purpose of creating a defect map for structures. The module is made up of three stages: *Image Extraction*, *Image Matching*, and *Projection*.

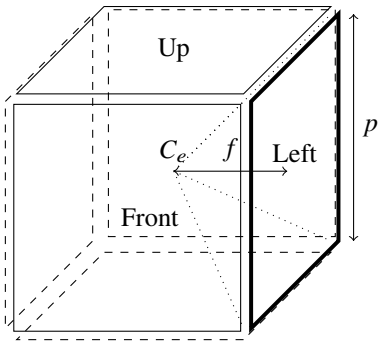


Figure 1. Cube configuration of the laser scan images. Every global camera pose C_e is associated with six images (Up, Down, Left, Right, Front, Back). f is the focal length, p is the image width.

3.1 Image Extraction

First, laser scan images and poses are extracted from a LiDAR scan. As input, we assume point clouds in the open general-purpose *E57* format [17] with multiple scan locations and RGB images. For best results, the scan locations should already be geo-referenced or registered among each other.

The images are extracted as rectified images in a cube configuration (see Figure 1). Every image is saved with an associated pose frame as a transformation matrix with translation and rotation. Pose frames are exported as JSON files with references to the associated JPEG file. We also attempt to extract camera information, such as focal length, from the *E57* file.

3.2 Image Matching

Image matching is done by first identifying visually distinct keypoints in defect and scan images. *LiLoc* supports keypoint detection using Scale-Invariant Feature Transform (SIFT), as well as XFeat [13] for finding keypoints with significant features. XFeat uses a convolutional neural network for finding features, which provides faster and more accurate keypoint detection than older methods. While SIFT produces a variable amount of keypoints, XFeat produces a maximum amount of 4096 keypoints, which are filtered by keypoint quality. All keypoints are saved per image and can be cached to disk for later reuse.

The second stage of image matching is finding keypoint matches. For image matching, we provide two matching methods: Exhaustive matching and cross matching. Exhaustive matching matches every image with every other

image, such as to find matches between all images of an inspection, or a defect instance, if one is specified. Cross matching matches every image of a set A with every image of a set B. This is useful for finding matches between laser scan images and inspection images, or between different capture dates.

Supported matching algorithms are k-Nearest-Neighbour (kNN) and LightGlue [14], a learned feature matching algorithm using deep neural networks. Both matching algorithms return a set of keypoint matches.

If the number of matches exceeds a threshold value, *LiLoc* attempts to find a homography between image pairs using OpenCV. This results in a perspective transformation matrix T. Results are written out in JSON format.

3.3 Projection

In the case of laser scan matching, we can use the match transform T to project points from the defect image onto the laser scan or a 3D model. For this, the intrinsic and extrinsic camera matrix (C_i and C_e respectively) are required for the transformation. The extrinsic camera matrix can be constructed from the position of the scanner and the direction of the scan image. The formula for the intrinsic camera matrix is as follows:

$$C_i = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{1}$$

We assume a pinhole camera model with a cube projection arrangement using square images with resolution P for width and height. Due to the cube arrangement, the focal length is equal to the half the image width/height $f_x = f_y = P/2$, resulting in a 90° field of view. Since the camera center is also the image center, it also follows $c_x = c_y = P/2$. Thus, we can simply write the intrinsic camera matrix as:

$$C_i = \begin{bmatrix} P/2 & 0 & P/2 \\ 0 & P/2 & P/2 \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

The extrinsic matrix $C_e = (R|t)$ encodes rotation R and translation t of the camera. Combining the extrinsic and intrinsic coordinates, we can obtain a directional vector $\vec{r}$ from an image coordinate $\vec{p}$.

$$\vec{r} = C_e C_i \vec{p} \tag{3}$$

The image coordinate is equal to the image center for now, but can be expanded when the defect position in the image is given. From position t towards direction $\vec{r}$, we use a cylinder cast (analogous to a ray cast) with a radius of 10 cm to find the closest point in the laser scan. We use the median of the closest 10 points to reduce the impact of outliers. This results in a 3D position for the defect image.

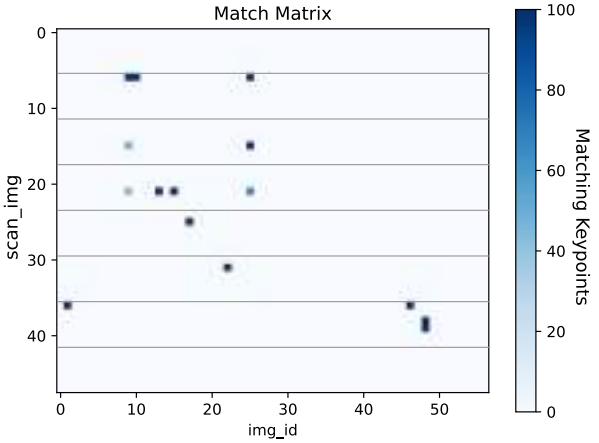


Figure 2. Match matrix of 57 defect images of to 8×6 laser scan pose images of Bridge A (scan-to-defect) using XFeat + LightGlue.

4 Case Study Results

Our approach was tested on two case studies consisting of two different highway bridges. For the lasers scans, a Leica RTC 360 was used, and photos were taken with Android and iOS smartphones.

Bridge A is a concrete bridge with a highway on-ramp. A full laser scan with 8 poses was taken, and images of concrete-based defects were collected on the same day. The individual scans were combined to a point cloud, and the point cloud was then registered to a pre-existing BIM model. For Bridge B is a 600 m pre-stressed concrete bridge. Partial laser scans were taken at five poses and 504 pictures of present defects in different camera angles were collected.

For Bridge A, we compare SIFT + kNN-Matching with XFeat + LightGlue. We only focus on XFeat + LightGlue for Bridge B.

4.1 Scan matching

For localization of the defect images in the laser scans, we attempt a match of every defect image with every scan image. For this, we extracted the scan images as described in Section 3.1. The matching threshold for all tests was set to 30 matches.

For this test we use XFeat as the keypoint extractor and LightGlue as the keypoint matcher. Bridge A with 8 scan locations (resulting in 48 rectified scan images) and 57 defect images results in 15 matches (see Figure 2). 10 images were correctly matched to at least one scan image, allowing for localization in the point cloud. Even with limited context, the matching performed well. Figure 3



Figure 3. Example of a positive match between a panoramic image (left) and a defect documentation image (middle). The defect image is projected onto the scan pose (right).

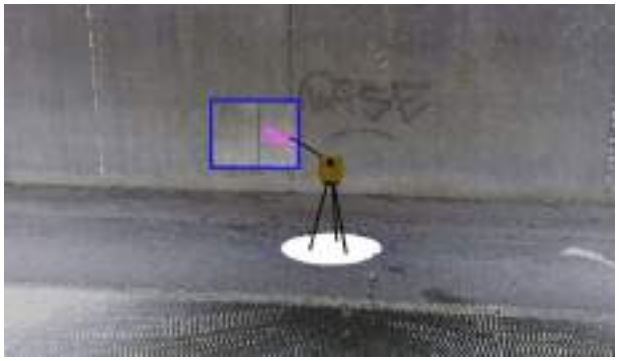


Figure 4. Projection of a defect image of graffiti on the concrete wall below Bridge A.

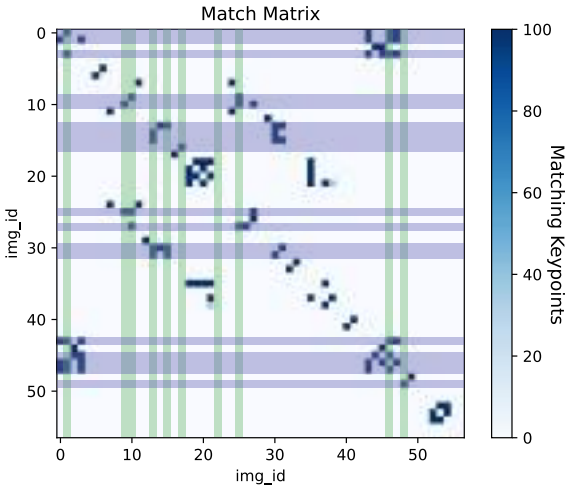


Figure 5. Match matrix of 57 defect images of Bridge A (defect-to-defect). Green lines show images that could be localized to a scan position (10 total). Blue lines include images that match to at least one localized image (18 total).

shows an example of spalling on the ceiling of the highway bridge. The detailed image of the defect was correctly matched to the scan image. Images that could not be matched to any scan image consist of extreme close-ups or of areas that the scanner did not cover. No false positives occurred in this limited data set.

We tested the projection capabilities with example images. Figure 4 shows the point cloud below the bridge. From the laser scan position, the matching defect image was projected onto the nearby wall. This links the defect to the point cloud with a 3D position.

As a baseline comparison, this data set was tested with SIFT features and kNN-matching. This resulted in 67 matches. Only four of the matches were correct (6% precision). Surprisingly, these four matches were matches that XFeat and LightGlue did not recognize. The low precision shows that classic keypoint matching methods are not suitable for finding matches in larger databases.

Bridge B with 9 scan positions (54 scan images) and 504 defect images resulted in 17 matches. We assume that the low number of matches is due to the limited scan coverage of the 600 m bridge. Again, no false positives were found. Notable is that most matches contain text written with chalk on the structure, like demarcations and measurements conducted by workers. Every matched defect image also corresponds to only one scan position, since the laser scan was limited to only significant locations with almost no overlap.

4.2 Defect matching

We used the exhaustive matching method for matching all defect images. This $n \times n$ matching allows for finding images of the same defect in the database.

Bridge A amounts to 58 matches in total (see Figure 5). Again, no false positives were found in this limited data set. Considering the scan matches from Section 4.1, we can subsequently localize overlapping and zoomed images of the same defect. 18 images are first removed to a direct scan match, and 22 images can be traced to at least one scan position in total. Some sets of images form a cluster, where 4–5 images of the same defect could be localized among each other, but could not be matched to a scan position.

For Bridge B we matched 504 images against each other. Overall, 904 matches were detected (see Figure 6). Since the image data set had up to ten pictures of the same defect from different perspectives, the matching clusters are larger than in Bridge A. Even though images were sometimes taken at extreme perspectives, the images were matched correctly (see Figure 7).

171 matches for Bridge B were manually classified as false positives (81% precision). The largest portion of false positives (120) turned out to be images with a crack

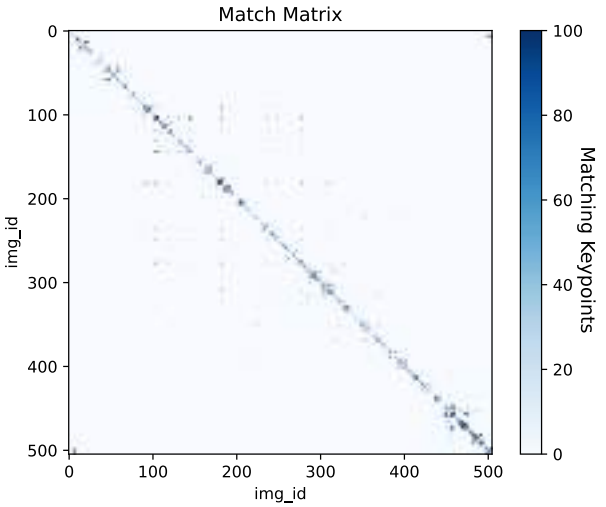


Figure 6. Match matrix of 504 defect images of Bridge B (defect-to-defect).

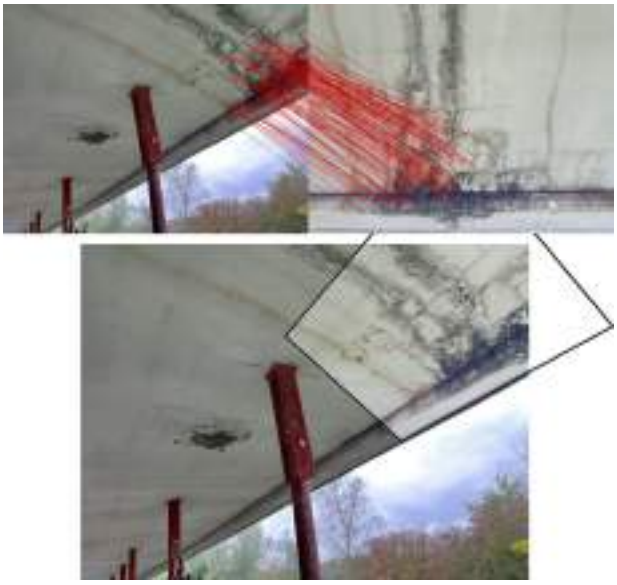


Figure 7. Example of a positive match of two images of the same defect, one with context, one with detail.



Figure 8. False positives for defect matching. Top: A crack width reference card is recognized as a match. Bottom: A similar scaffolding element is recognized as a match.

width reference card, a tool often used by inspectors to measure cracks (see Figure 8). The other false positives were mainly due to scaffolding present at the bridge, which shows a high degree of self-similarity. Increasing the keypoint match threshold to 100 reduces the number of total matches to 667 and the false positives to 47 (93% precision).

Tracing all images of Bridge B to at least one scan position allows localizing more defects. Filtering for false positives, 92 images can be traced to at least one scan position, allowing for localization in the point cloud.

4.3 Discussion

Using XFeat and LightGlue for matching shows significantly better results than using SIFT and kNN matching. SIFT and others have been developed for image stitching or live image tracking, such as SLAM, instead of recalling images taken at separate times with different cameras. While still useful in some of these situations (e. g. loop closure), matching entire databases produces too much false positives to be useful. Meanwhile, AI matching methods prove to be sensitive enough to detect significant points even in uniform concrete surfaces.

While only a small portion of defect images could be matched to scan images in our data set directly, exhaustive matching could fill some gaps. Using exhaustive matching to detect defect image clusters allowed for tracing these clusters to at least one scan image. Despite this, some images still could not be traced to a scan position. This is possibly due to incomplete laser scans or defect images that did not provide enough context for a successful localization.

False positives were mostly due to similar objects in the images, such as scaffolding or instruments. Still, some false positives are due to the self-similar nature of bridges. We have also neglected counting false negatives, as defining a false negative in terms of localization can be difficult.

5 Conclusion

We presented *LiLoc*, a tool for localizing defect images in laser scans, and tested it on two self-collected data sets. The tool supports extracting scan images from point clouds, matching images in exhaustive and cross-matching mode, and mapping defect images to a 3D coordinate. The code is documented on GitHub[1].

This tool paves the way for researchers, engineers and on-site inspectors utilizing the historical data of inspections to make maintenance easier and more comprehensible. With this image matching tool, historical data can be used in conjunction with modern digital tools. This approach can be integrated into existing inspection processes as historical data can be linked to new digital technologies. Databases can be enriched with location specific data, closing the gap between image data, laser scans, and DTs.

The results show that with AI keypoint extraction, reliable image localization is achievable. Learned features perform better on concrete structures than classic methods such as SIFT. Even difficult matches of pure concrete walls can reliably be matched based on the imperfections in the concrete. This has previously been a challenge with classic methods.



Figure 9. Example of a match between two images of the same defect, taken in 2005 and 2014 respectively.

Although the results are promising, some limitations are evident. Some objects or artifacts in the image can confuse the matching algorithm, such as measurement instruments or burnt in text on the image (such as the capture date). One option would be to detect these objects using a semantic segmentation network filter out keypoints with a mask. Lighting conditions are also a significant impact factor. In this paper, we have not compared large-scale data sets with images under different lighting conditions or camera modes (e. g. flash photography). The data set should be expanded upon for realistic applications. In this study, we have also only quantified the method's performance with data that was recorded concurrently. Figure 9 shows an example of a successful match of real historical inspection data with a time difference of nine years. Due to the difficulty of obtaining historical data sets in a large scale, this study focused on data captured without a large time difference. In future research, the data set should be expanded with more data recorded at different times and conditions. Crafting data sets that reflect these challenges is a necessary step in future work.

Future research will focus on improving the algorithms used in this paper, and adding new keypoint and matching algorithms. Both XFeat and LightGlue allow for training the network to adjust to specific conditions. Training these algorithms specifically for the materials used in the construction industry may improve precision significantly. Removing keypoints on equipment, as well as other camera artifacts such as burnt-in dates may be masked out in a pre-processing step. Other local feature extractors such as GlueStick [18] work with lines in addition to points, which may prove suitable for the built environment.

[1] https://github.com/RUB-Informatik-im-Bauwesen/LiLoc

Acknowledgment

This research was funded by the mFUND research programme of the Federal Ministry for Digital and Transport (BMDV) (funding code: 01F2259A). This paper is based on parts of the research project carried out at the request of the BMDV, represented by the Federal Highway Research Institute, under research project No. 69.0017/2023. The authors are solely responsible for the content.

References

[1] Jacob J. Lin, Amir Ibrahim, Shubham Sarwade, and Mani Golparvar-Fard. Bridge Inspection with Aerial Robots: Automating the Entire Pipeline of Visual Data Capture, 3D Mapping, Defect Detection, Analysis, and Reporting. *Journal of Computing in Civil Engineering*, 35(2):04020064, March 2021. ISSN 1943-5487. doi:10.1061/(ASCE)CP.1943-5487.0000954.

[2] Yi Tan, Geng Li, Ruying Cai, Jun Ma, and Mingzhu Wang. Mapping and modelling defect data from UAV captured images to BIM for building external wall inspection. *Automation in Construction*, 139:104284, July 2022. ISSN 0926-5805. doi:10.1016/j.autcon.2022.104284.

[3] Sungsik Yoon, Gi-Hun Gwon, Jin-Hwan Lee, and Hyung-Jo Jung. Three-dimensional im-

age coordinate-based missing region of interest area detection and damage localization for bridge visual inspection using unmanned aerial vehicles. *Structural Health Monitoring*, 20(4):1462–1475, July 2021. ISSN 1475-9217, 1741-3168. doi:10.1177/1475921720918675.

[4] Sizeng Zhao, Fei Kang, and Junjie Li. Concrete dam damage detection and localisation based on YOLOv5s-HSC and photogrammetric 3D reconstruction. *Automation in Construction*, 143: 104555, November 2022. ISSN 09265805. doi:10.1016/j.autcon.2022.104555.

[5] David G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2): 91–110, November 2004. ISSN 1573-1405. doi:10.1023/B:VISI.0000029664.99615.94.

[6] Shang Jiang, Yuyao Cheng, and Jian Zhang. Vision-guided unmanned aerial system for rapid multiple-type damage detection and localization. *Structural Health Monitoring*, 22(1): 319–337, January 2023. ISSN 1475-9217. doi:10.1177/14759217221084878.

[7] E. T. Bartczak, M. Bassier, and M. Vergauwen. RTK-Based Photogrammetric Reconstruction and Spatial Filtering. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-1/W2-2023:1873–1880, December 2023. ISSN 2194-9034. doi:10.5194/isprs-archives-XLVIII-1-W2-2023-1873-2023.

[8] Jing-Lin Xiao, Jian-Sheng Fan, Yu-Fei Liu, Bao-Luo Li, and Jian-Guo Nie. Region of interest (ROI) extraction and crack detection for UAV-based bridge inspection using point cloud segmentation and 3D-to-2D projection. *Automation in Construction*, 158:105226, February 2024. ISSN 09265805. doi:10.1016/j.autcon.2023.105226.

[9] Bundesministerium für Verkehr und digitale Infrastruktur. Richtlinie zur einheitlichen Erfassung, Bewertung, Aufzeichnung und Auswertung von Ergebnissen der Bauwerksprüfungen nach DIN 1076 (RI-EBW-PRÜF), 2017-02-22, 2017. URL `https://www.bast.de/DE/Publikationen/Regelwerke/Ingenieurbau/Erhaltung/RI-EBW-PRUEF-Erhaltung.html`.

[10] Anne Göbels and Jakob Beetz. Relative Location Ontology: An ontological model for representing directional topological relationships between spatial entities in oriented space. In *Proceedings of the 2nd Linked Data in Architecture and Construction Workshop 2024*, Bochum, Germany, June 2024.

[11] Ina Heise, Anne Göbels, Andre Borrmann, and Jakob Beetz. Enabling Comprehensive Querying of Road and Civil Structure Data using Graph-based Methods. In *Proceedings of the 41st International Conference of CIB W78*, Marakesh, Morrocco, October 2024.

[12] Amir Roshan Zamir and Mubarak Shah. Accurate Image Localization Based on Google Maps Street View. In Kostas Daniilidis, Petros Maragos, and Nikos Paragios, editors, *Computer Vision – ECCV 2010*, Lecture Notes in Computer Science, pages 255–268, Berlin, Heidelberg, 2010. Springer. ISBN 978-3-642-15561-1. doi:10.1007/978-3-642-15561-1_19.

[13] Guilherme Potje, Felipe Cadar, Andre Araujo, Renato Martins, and Erickson R. Nascimento. XFeat: Accelerated features for lightweight image matching. In *2024 IEEE / CVF Computer Vision and Pattern Recognition (CVPR)*, 2024.

[14] Philipp Lindenberger, Paul-Edouard Sarlin, and Marc Pollefeys. LightGlue: Local feature matching at light speed. In *International Conference on Computer Vision (ICCV)*, 2023.

[15] Semih Orhan and Yalin Bastanlar. Efficient Search in a Panoramic Image Database for Long-term Visual Localization. In *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 1727–1734, Montreal, BC, Canada, October 2021. IEEE. ISBN 978-1-6654-0191-3. doi:10.1109/ICCVW54120.2021.00198.

[16] Alex Kendall, Matthew Grimes, and Roberto Cipolla. PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 2938–2946, Santiago, Chile, December 2015. IEEE. ISBN 978-1-4673-8391-2. doi:10.1109/ICCV.2015.336.

[17] Daniel Huber. The ASTM E57 file format for 3D imaging data exchange. In *Proceedings of SPIE Electronics Imaging Science and Technology Conference (IS&T), 3D Imaging Metrology*, volume 7864, 2011-01, 2011.

[18] Rémi Pautrat, Iago Suárez, Yifan Yu, Marc Pollefeys, and Viktor Larsson. GlueStick: Robust image matching by sticking points and lines together. In *International Conference on Computer Vision (ICCV)*, 2023.

A Robust Pipeline for Facade Geometry Estimation Using AprilTags and Digital Imaging

Aravind Srinivasaragavan[1], Danya Liu[1], Kepa Iturralde[1,2] and Christoph Holst[1]

[1]Chair of Engineering Geodesy, School of Engineering and Design,
Technical University of Munich, 80333 Munich, Germany
[2]Chair of Digital Transformation in Construction, Institute of Construction Management,
Faculty of Civil and Environmental Engineering, University of Stuttgart, 70569 Stuttgart, Germany

aravind.srinivasaragavan@tum.de, danya.liu@tum.de, kepa.iturralde@ibl.uni-stuttgart.de, christoph.holst@tum.de

Abstract -

The accurate modeling of building facades plays a critical role in architecture, urban planning, and energy efficiency. Traditional methods of facade mapping are often labor-intensive, costly, and prone to error. Recent advancements in computer vision and sensor technology have introduced innovative approaches for capturing precise geometric data, including the use of AprilTags—fiducial markers that enable efficient facade reconstruction through digital imaging techniques. This paper explores a method that involves arranging AprilTags in a grid pattern on building surfaces, capturing multiple images of the tagged facade, and using triangulation techniques to infer the 3D geometry of the structure with high accuracy. The resulting digital models have a wide range of applications, including optimizing solar panel placement, creating detailed replicas for heritage conservation, and enhancing urban planning through improved simulations of light, shadows, and heat.

Keywords -

Facade Modeling; AprilTags; 3D Reconstruction; Stereo Vision; Camera Calibration; Fiducial Markers; Distance Estimation

1 Introduction

The accurate modeling of building facades has become increasingly important in various fields such as architecture, urban planning, and energy efficiency. Traditional methods of mapping and measuring facades can be time-consuming, costly, and prone to error. Recent advancements in computer vision and sensor technology have introduced new approaches for capturing detailed and precise geometric data of structures. One such method involves the use of AprilTags—fiducial [1] markers that can be placed on the facade—and digital imaging techniques [2] to estimate the real-world distances between these markers, effectively reconstructing the facade's geometry.

In this technique, AprilTags are arranged in a grid pattern on the building's surface, either manually or with the assistance of a drone. A camera then captures multiple images of the tagged facade from different angles, and, combined with information such as the camera's focal length and intrinsic parameters, these images are used to infer the relative positions of the tags. By triangulating the distances between the tags, the method can reconstruct the 3D geometry of the facade with high accuracy. This process allows for the creation of a digital model that closely represents the real-world structure, enabling applications in design, planning, and analysis.

The resulting geometric data has diverse applications beyond simple structural visualization. For example, in the field of renewable energy, the technique can be used to precisely model building surfaces for the optimal placement of solar panels. Similarly, it can assist in heritage conservation, where detailed digital replicas of historical facades can be created for restoration or archival purposes. Furthermore, this method could be applied to urban studies, enabling more accurate simulations of shadows, light exposure, and heat absorption in city planning. As building facades continue to evolve in complexity and functionality, the ability to quickly and accurately model these surfaces opens up new possibilities for both practical and theoretical applications in the built environment.

2 Problem Description and Constraints

The accurate modeling of building facades is a critical task in fields such as architecture, urban planning, and energy efficiency. However, traditional methods of mapping and measuring building facades are inherently challenging due to several factors. These methods are often time-consuming, labor-intensive, and prone to human error, especially when dealing with large or complex structures. Furthermore, the equipment required for manual measurements can be costly and may not always yield the level of precision necessary for advanced applications.

Recent advancements in computer vision and sensor technology offer an opportunity to address these challenges [3]. Specifically, the use of AprilTags—fiducial markers designed for robust detection—has emerged as a promising approach for facade modeling. These markers enable precise distance measurements and allow for the reconstruction of a facade's geometry through digital imaging.

First, the placement of AprilTags requires a systematic and controlled process, typically arranged in a grid pattern on the building's surface. This can be done manually or with the assistance of drones, but both approaches present logistical challenges, particularly for large buildings or difficult-to-access facades. Second, the accuracy of the reconstruction depends heavily on the quality and quantity of the images captured. Factors such as camera calibration, lighting conditions, and the angles from which the images are taken can significantly impact the precision of the results. Additionally, the computational process of triangulating the positions of the tags requires careful handling of camera intrinsic parameters, such as focal length, to ensure reliable 3D geometry generation.

These constraints highlight the need for a method that balances practicality and accuracy while minimizing costs and manual effort. Despite these challenges, the use of AprilTags for facade modeling represents a scalable and efficient solution that could potentially overcome the limitations of traditional techniques, enabling diverse applications in design, analysis, and planning. In short, the main problem is to determine the spacial geometry of the placed AprilTags as accurately as possible. The spacial geometry of the Apriltag grid directly gives one information about the geometry of the building facade. In this paper, we propose a pipeline, which only involves a Camera with a fixed focal length and a Python pipeline, which can be used to very accurately determine the distance to an Apriltag. This information can then be used to estimate the spacial geometry of the AprilTag grid.

3 Technical Preleminaries

In the process of reconstructing the geometry of building facades, one common challenge is the presence of noise and inaccuracies in the observed 3D data points, especially when these points are assumed to be co-planar. While techniques like AprilTag-based distance estimation offer high accuracy in reconstructing building geometry, small errors in the measurements or in the camera calibration can introduce deviations that affect the precision of the reconstructed model. In the next section, we will discuss one possible method to mitigate this issue.

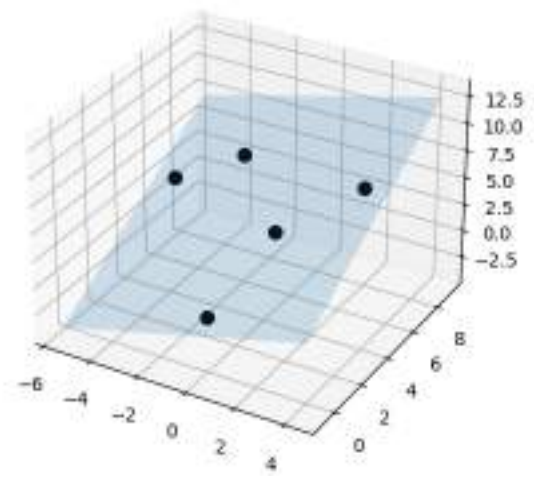


Figure 1. The Best-Fitting plane passing through 5 points

3.1 Best-Fitting Plane: The Key Improvement Strategy

A practical solution to this problem is the concept of the best-fitting plane—a mathematical approach that minimizes the error between observed data points and a plane that best represents their true position.

The best-fitting plane is typically computed using methods like least squares fitting, where the goal is to find the plane that minimizes the sum of squared distances between the observed data points and the plane itself. This method is particularly useful when dealing with 3D data points that are known a-priori to lie approximately on the same plane, such as those derived from the AprilTag markers on a building facade. By fitting a plane to these points, we can correct for small inaccuracies in their positioning, leading to a more accurate representation of the facade geometry.

In the context of our research, using a best-fitting plane can significantly enhance the quality of the 3D model by refining the locations of the AprilTags as shown in Figure 1. This correction process helps mitigate errors caused by slight misalignments or measurement noise, ensuring that the reconstructed facade closely matches the actual structure. Such improvements are particularly valuable when the facade's geometry is needed for precise applications like solar panel placement, where even minor discrepancies in the 3D model could affect performance. By leveraging the best-fitting plane approach, this method can contribute to more reliable and accurate models for various real-world applications, including energy optimization, urban planning, and architectural analysis.

3.2 Distance Estimation via Effective Focal Length

Accurate distance estimation to an AprilTag from a camera is a critical aspect of the proposed facade modeling method. This process relies on key camera parameters, including the *effective focal length* and the *pixel size* of the camera's sensor. These parameters enable the calculation of real-world distances by relating the apparent size of the AprilTag in the captured image to its known physical dimensions.

3.3 Effective Focal Length and Camera Geometry

The effective focal length (f) of a camera [4] is the distance from the lens to the sensor at which parallel rays of light converge to form a sharp image. It is typically expressed in millimeters and is a fundamental property of the camera's optics. Combined with the sensor's pixel size (p), the focal length determines the *field of view (FOV)* and directly impacts the perceived size of objects in the image.

For a camera with a focal length f and a sensor with pixel size p, the physical distance between the camera and an object, such as an AprilTag, can be determined using the principle of similar triangles. The relationship between the tag's real-world dimensions and its image size on the sensor provides the basis for this calculation.

3.4 Calculating Distance Using Image Data

When an AprilTag is captured in an image, its size in pixels (s_{pixels}) can be measured. The real-world width of the tag (W) is known a priori. Using these values, the distance to the tag (D) can be calculated using the following formula:

$$D = \frac{f \cdot W}{s_{\text{pixels}} \cdot p}$$

Here:

- f is the effective focal length of the camera (in millimeters),
- W is the real-world width of the AprilTag (in millimeters),
- s_{pixels} is the width of the AprilTag in the image (in pixels),
- p is the physical size of a single pixel on the sensor (in millimeters).

3.5 Practical Considerations

1. **Camera Calibration:** Accurate distance estimation requires precise knowledge of the camera's intrinsic parameters, including the focal length and pixel size. Camera calibration processes, such as those using checkerboard patterns, are essential for determining these values accurately.
2. **Tag Orientation and Perspective:** If the AprilTag is not perpendicular to the camera's optical axis, its apparent size in the image will be distorted. Perspective correction algorithms may be required to account for this distortion and ensure accurate measurements.
3. **Resolution and Image Quality:** The resolution of the captured image affects the precision of s_{pixels}. High-resolution images provide more accurate measurements of the tag's size, improving distance estimation accuracy.
4. **Lighting and Focus:** The clarity and visibility of the AprilTag in the image depend on proper lighting and focus. Poor lighting conditions or out-of-focus images can introduce errors in size estimation and, consequently, distance calculations.

3.6 Applications in Facade Modeling

By determining the distances to multiple AprilTags placed on a building's facade, a 3D reconstruction of the surface can be performed. The calculated distances form the basis for triangulating the relative positions of the tags, enabling the generation of a detailed geometric model of the structure. This approach, leveraging the effective focal length and pixel size, provides a robust and scalable solution for facade modeling with high accuracy.

4 The Main Pipeline

The proposed pipeline for facade modeling leverages AprilTags and digital imaging techniques to accurately reconstruct the geometry of building facades. The steps involved in the pipeline are outlined below:

4.1 Placement of AprilTags

Several AprilTags are placed in a grid-like formation on the facade like Figure 2.

This systematic arrangement ensures that the tags are evenly distributed across the surface, enabling accurate reconstruction. Care is taken to maintain consistent spacing between the tags and to cover the entire area of interest.

4.2 Image Capture

Multiple images of the tagged facade are captured from different distances. To minimize distortion and maintain accuracy, the camera angles are kept relatively shallow, avoiding overly steep perspectives that could negatively affect the results. These images serve as the input for the subsequent processing and analysis stages. For the

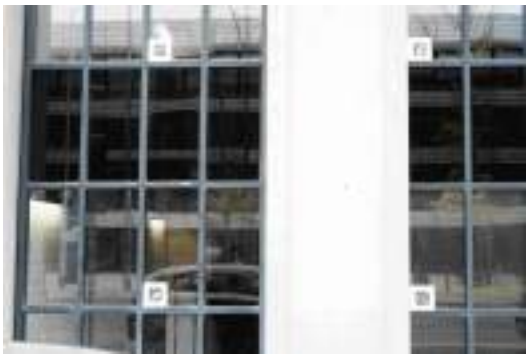


Figure 2. The setting of the AprilTags' outline



Figure 3. Contrast Adjusted Image

experiments themselves, we used the Sony Alpha 7R IV 35-mm-Vollbildkamera with 61,0 MP [5]. Based on our experiments, we found that this particular camera has an effective focal length of approximately 13257.6 (in the appropriate units).

4.3 Pre-Processing

Before detecting the AprilTags, the captured images undergo pre-processing to enhance their quality. This step includes modifying parameters such as gamma and contrast to ensure the tags are clearly visible and detectable in the images as shown in Figure 3.

These adjustments help mitigate the effects of varying lighting conditions and improve the reliability of the detection process.

4.4 AprilTag Detection

The AprilTags in the pre-processed images are detected shown in Figure 4 using the Python library `AprilTag` [6], which is available at `https://github.com/Tinker-Twins/AprilTag`. This library identifies the tags and determines their pixel dimensions (s_{pixels}) in the images. These pixel measurements are essential for calculating the real-world distances to the tags.

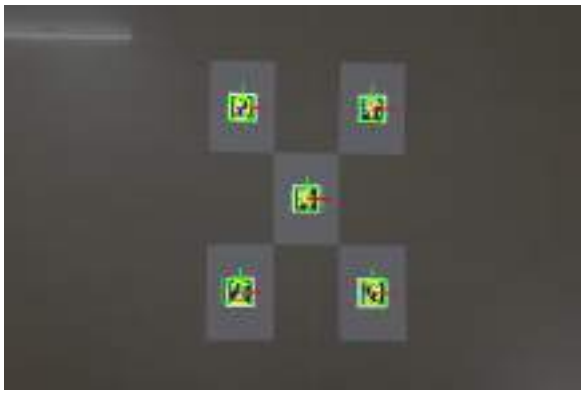


Figure 4. Detection results

4.5 Calculating the Distance Between Two AprilTags

Once the distances from the camera to individual AprilTags and their positions in the 3D space are determined, the distance between two AprilTags can be calculated by analyzing their relative spatial coordinates.

4.6 Steps for Calculation

1. **Determine the 3D Coordinates of Each AprilTag:** Using the camera as the origin of the coordinate system, the 3D position of each AprilTag (x, y, z) can be determined from the camera parameters and the pixel coordinates of the tag in the image. For a tag at distance D from the camera:

$$x = D \cdot \frac{u - u_0}{f_x}, \quad y = D \cdot \frac{v - v_0}{f_y}, \quad z = D$$

where:

- D is the distance from the camera to the AprilTag (calculated using the focal length, tag size, and pixel size),
- (u, v) are the pixel coordinates of the AprilTag in the image,
- (u_0, v_0) are the principal point coordinates (typically the center of the image),
- f_x and f_y are the focal lengths of the camera in the x- and y-axes (in pixel units).

2. **Final Correction via the Best-Fitting Plane** Since we know (approximately), that the AprilTags all lie in a Plane, we can use the best fitting plane for the 3D coordinates obtained in the previous section and project the points on this plane to obtain the corrected points (x', y', z'). Note that we used the Python library `scikit-spacial` [7] for the Best-Fitting plane compuations.

3. **Calculate the Euclidean Distance Between Two AprilTags:** Let the coordinates of the first AprilTag be (x_1, y_1, z_1) and the second AprilTag be (x_2, y_2, z_2). The distance d between the two tags is given by the Euclidean distance formula:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

4.7 Practical Considerations

- **Accuracy of Coordinates:** The accuracy of the calculated coordinates depends on the precision of the distance measurements (D) and the camera's intrinsic parameters (f_x, f_y, u_0, v_0).
- **Perspective Distortion:** If the AprilTags are not aligned perpendicularly to the camera's optical axis, corrections may be required to ensure accurate coordinate estimation.
- **Tag Identification:** Each tag must be uniquely identified in the image to correctly associate its 3D position with its real-world counterpart.

By calculating the 3D positions of multiple AprilTags and using the Euclidean distance formula, the relative distances between tags can be computed with high precision. This process enables the reconstruction of the facade's geometry with reliable spatial relationships between the tags.

5 Experimental Results

Several experiments were performed on one of the TUM buildings in Arcisstrasse by considering a section of the windows as the Facade. Six AprilTags, with different tag IDs were placed on the windows as shown in Figure. In this particular experiment, AprilTags with the following IDs were used: $1, 2, 3, 1617, 18, 20, 22$. Here we describe the results of once such experiment. The results of this experiment are described in two tables. Table 1 shows the estimated distance between the tags Tag A and Tag B (A and B being the Tag IDs) for three specific cases.

On the other hand, Tables 2 and 3 show the the 3D coordinates of the Tags with ID 2 and 16. Here, the Tag 1 is the reference Tag. This means that the coordinates have been changed in such a way so that the Tag 1 becomes the origin (with coordinates $(0, 0, 0)$).

6 Conclusion and Future Developments

In this work, we presented a robust pipeline for the accurate modeling of building facades using AprilTags and digital imaging techniques. By leveraging fiducial markers arranged in a systematic grid pattern, we demonstrated how the real-world distances and relative positions of these tags can be calculated to reconstruct the 3D geometry of facades. The process incorporates critical steps such as camera calibration, pre-processing of images to enhance quality, and the use of the `AprilTag` Python library for precise tag detection. The geometric data obtained from this pipeline can be applied to a wide range of fields, including renewable energy optimization, heritage conservation, and urban planning. One more thing to note is that this pipeline does not involve complicated (and rather archaic) tools such as a Total Station. Further, this pipeline can reduce the time it takes to estimate the facade geometry in comparison to other traditional methods.

We also discussed the mathematical foundations underpinning the reconstruction process, particularly the use of the camera's effective focal length, pixel size, and image data to determine the spatial coordinates of each AprilTag. Furthermore, the ability to calculate the distances between tags using their 3D positions allows for the creation of highly accurate and detailed facade models.

Looking ahead, we plan to enhance the accuracy and scalability of our approach by incorporating stereo vision techniques [8, 9, 10, 11]. Specifically, future experiments will utilize the FUJIFILM GFX 100 II camera [12], a high-resolution imaging system, in a dual-camera setup(the other camera being the current on). By capturing synchronized images from two cameras at different perspectives, stereo vision will enable precise depth estimation and reduce potential errors caused by distortion or

Distance Between Tags(mm)			
Tag A and Tag B	3-18	16-22	20 - 22
1	1235.3	2106.9	1621.0
2	1248.7	2127.7	1653.5
3	1278.7	2210.4	1605.1
4	1270.9	2183.5	1734.3
Ground Truth	1250.0	2150.0	1650.0

Table 1. Estimated distance between the tags

Tag 2 Coordinates (mm)			
Axis	X	Y	Z
1	-4979.87	61.68	63.04
2	-4979.50	53.34	55.96
3	-3345.30	71.55	86.37

Table 2. Estimated Coordinates of a Tag

Tag 16 Coordinates (mm)			
Axis	X	Y	Z
1	135.63	3812.57	63.04
2	127.26	3808.09	55.96
3	136.23	3725.82	86.37

Table 3. Estimated Coordinates of a Tag

steep angles. This will significantly improve the reliability of 3D reconstructions and make the method more suitable for complex facade geometries.

The integration of advanced camera systems and stereo vision techniques represents a promising direction for this work, offering increased accuracy and flexibility for facade modeling. These advancements will further expand the applicability of our approach to a broader range of architectural, engineering, and urban development challenges.

References

[1] Yifan Zhang, Jian Wang, and Xinyu Li. Analysis and improvements in apriltag based state estimation. *Sensors*, 23(1):123, 2023.

[2] Zhen Li, Jian Zhang, Ming Li, and Jun Wang. Automatic reconstruction of building façade model from photogrammetric mesh model. *Remote Sensing*, 13 (19):3801, 2021.

[3] Kepa Iturralde, Samanti Das, Aravind Srinivasaragavan, Thomas Bock, and Christoph Holst. An automated prefabricated facade layout definition for residential building renovation. *Buildings*, 13(12), 2023. ISSN 2075-5309. URL `https://www.mdpi.com/2075-5309/13/12/2981`.

[4] Canon Europe. *Understanding Focal Length*, 2025. URL `https://www.canon.ge/pro/infobank/understanding-focal-length/`. Accessed: 2025-01-15.

[5] Sony Europe B.V. Alpha 7R IV 35-mm-Vollformatkamera mit 61,0 MP description, 1999. URL `https://www.sony.de/electronics/wechselobjektivkameras/ilce-7rm4`.

[6] Tinker-Twins. Apriltag python library. `https://github.com/Tinker-Twins/AprilTag`, 2025.

[7] Andrew Hynes. *scikit-spatial: Spatial objects and computations based on NumPy arrays*, 2023. URL `https://scikit-spatial.readthedocs.io/en/stable/api_reference/Plane/methods/skspatial.objects.Plane.best_fit.html`. Accessed: 2025-01-15.

[8] Liang Tang and Gerard Medioni. Building facade detection, segmentation, and parameter estimation for mobile robot stereo vision applications. *Image and Vision Computing*, 31(10):759–773, 2013.

[9] Graham Fyffe, Andrew Jones, Oleg Alexander, and Paul Debevec. Multi-view stereo on consistent face topology. *ACM Transactions on Graphics (TOG)*, 36 (4):1–10, 2017.

[10] Hao Wang, Wei Li, and Yiming Zhang. Landmark-based localization using stereo vision and deep learning in gps-denied environments. *arXiv preprint arXiv:2402.12551*, 2023.

[11] Jeffrey A Delmerico, Jason J Corso, and Philip David. Boosting with stereo features for building facade detection on mobile platforms. In *Proceedings of Western New York Image Processing Workshop*, pages 1–4. IEEE, 2010.

[12] FUJIFILM Corporation. *FUJIFILM GFX100 II Product Page*, 2025. URL `https://eshop.fujifilm-x.com/de/fujifilm-gfx100-ii.html`. Accessed: 2025-01-15.

Acknowledgement



The research presented in this paper was originally carried out in the context of the H2020 project **ENSNARE**. **ENSNARE** has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. **958445**.

Automated Inspection Methodology of Wall Rebar Spacing and Diameters Using Laser Scanning

Philgu Kim[1], Jungyeon Kim[1], Jeongwoon Choi[1], Seokho Chi[1]

[1]Department of Civil and Environmental Engineering, Seoul National University, Republic of Korea
leophilgu@snu.ac.kr, janykjy@snu.ac.kr, cjw5428@snu.ac.kr, shchi@snu.ac.kr

Abstract

This study presents a methodology for the automated inspection of rebar diameters and spacing in wall structures using a laser scanner. Previous studies conducted the methods in controlled laboratory environments, and primarily focused on predicting rebar diameters larger than 10 mm. However, these methods have difficulties applying to real-world construction sites which have rebars on bent, and limited scanning positions. This research was performed on an actual construction site where small-diameter rebars (10 mm) were arranged in double layers. To address the curvature presented in rebar arrangement surfaces, a new methodology was introduced; linearity was analyzed to detect the complex rebar arrangements. The spacing of the rebars was calculated by projecting the point clouds onto an axis aligned with the rebar arrangement. Rebar diameters were estimated through RANdom Sample Consensus (RANSAC) circle fitting. The results demonstrate that the proposed methodology successfully modeled distorted rebar structures presented in construction sites, enabling accurate extraction and prediction of rebar objects. This approach can replace labor-intensive manual inspections, ensuring accurate and reliable verification against design specifications.

Keywords

Laser scanner; Point cloud; Automated inspection, Rebar spacing, Rebar diameter

1 Introduction

To ensure structural safety in reinforced concrete construction, rebars must be placed precisely at the designated locations and with the appropriate sizes specified in the design drawings. The placement and size of rebars directly play a critical role in determining the load-bearing capacity of reinforced concrete [6]. Improperly arranged rebars during concrete casting can result in structural failures. Therefore, rigorous inspections of rebar placement are essential.

Traditionally, rebar placement inspections in construction sites are performed visually by inspectors who physically traverse the site, using tape measures to check spacings and sizes. As the work area expands, these inspections become increasingly time-consuming, labor-intensive, and hazardous, as inspectors must move across all rebar-equipped areas [1].

To overcome the challenges of manual inspections, numerous studies have aimed to automate the rebar inspection process. Among these approaches, laser scanners have gained significant attention for their ability to capture point cloud data of rebar arrangements and to subsequently compute rebar diameters and spacing [2]. However, previous studies have conducted experiments under laboratory conditions. These studies primarily focused on predicting rebar diameters and spacing under optimal scanning angles, with most cases involving rebar diameters larger than 10 mm. In contrast, actual construction sites feature highly complex rebar arrangements. Thus, scanning positions are often limited and conditions, which differ significantly from laboratory settings, make it challenging to accurately detect rebars on bent or irregular surfaces. Consequently, these methods are unable to recognize rebars in diverse positions or with small diameters. To address these limitations, this study focused on scanning rebars arranged at real construction sites, segmenting each rebar, and automatically predicting their diameters and spacing. Furthermore, the proposed method can be utilized in Scan-to-BIM (Building Information Modeling) workflows by enabling the automated generation of 3D as-built rebar models. This approach ultimately contributes to improving construction quality control, structural verification, and digital lifecycle management in reinforced concrete construction.

2 Literature Review

Point clouds are widely used owing to their high efficiency and accuracy [10, 11, 12, 13, 14, 15]. Previous studies have created and analyzed point clouds generated by Terrestrial Laser Scanners (TLS) to automate rebar inspection [2, 3, 4, 5, 6]. These studies typically focused

on separating vertical and horizontal rebars to estimate both the diameter of each segment and the spacing between segments. For instance, Kim et al. [2] employed RANSAC [9] to segment individual rebars and applied Density-based Modeling (DBM) to enhance the classification accuracy for rebar diameter estimation. Li et al. [17] presented a scan planning method to select the optimal scan position of laser scanner for improving the rebar inspection accuracy. However, these methods are conducted under laboratory conditions. Most of previous studies focused on predicting rebar diameters and spacing based on optimal scanning angles, primarily focusing on rebars larger than 10 mm in diameter. In real-world scenarios, rebar arrangements are far complex, and scanning positions are frequently constrained. Likewise, the environmental differences from laboratory settings pose significant challenges in accurately detecting rebars on curved of uneven surfaces.

To address these gaps, this study proposes methodology for two purposes: (1) rebar detection using laser scans in complex construction sites, and (2) diameter and spacing prediction for small-diameter rebars.

3 Methodology

3.1 Research Framework

Figure 1 illustrates the framework of this study: rebar scanning and 3D reconstruction, data preprocessing, rebar detection, and diameter and spacing prediction. During the rebar detection stage, planarity, linearity, and iterative RANSAC are utilized to remove formwork and base surfaces. Subsequently, double-layered rebar arrangements are distinguished through Otsu thresholding. For spacing prediction, projection algorithm and cubic spline interpolation are used to predict the location of vertical rebars. For diameter prediction, rebars are divided into individual instances and projected onto the surface perpendicular to the first axis of Principal Component Analysis (PCA). The diameters are then predicted using circle fitting based on the RANSAC method.

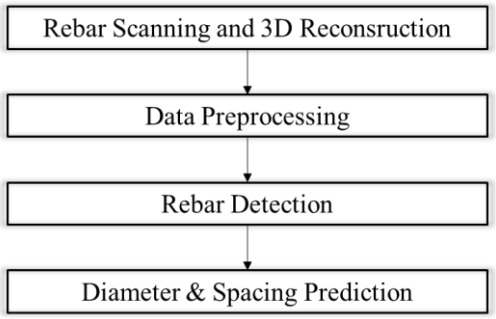
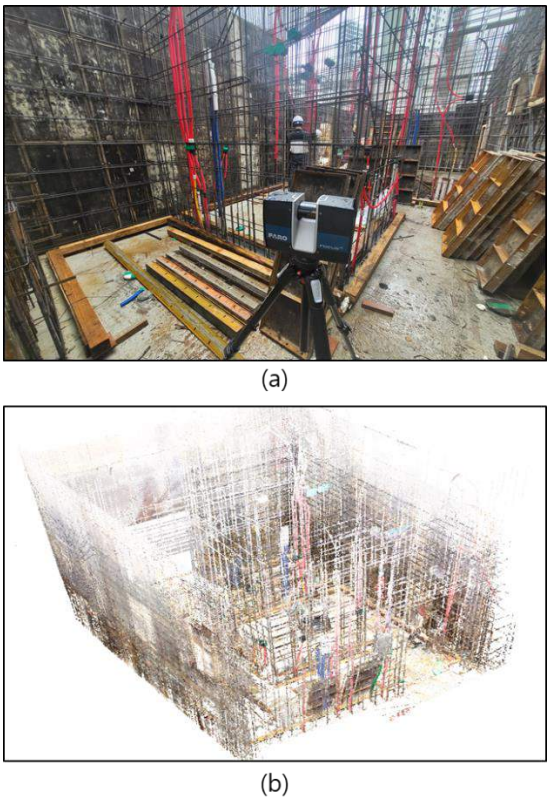


Figure 1 Research framework

3.2 Rebar Scanning and 3D Reconstruction

At an actual construction site, a wall rebar structure with 10 mm diameter rebars arranged in double layers at 300 mm intervals beside the formwork was scanned. The scanning was performed from five different positions using a FARO Focus M laser scanner. The point clouds collected from each position were subsequently combined into one unified point cloud using FARO Scene software, as shown in Figure 2.



(a)

(b)

Figure 2 (a) Scanning of wall rebar arrangement; (b) 3D point cloud reconstruction of wall rebar arrangement

3.3 Data Preprocessing

Prior to analyzing the rebars, noise was removed using the Statistical Outlier Removal (SOR) method [16]. In this process, points whose average distance to their 10 nearest neighbors exceeds twice the standard deviation threshold were considered outliers and removed.

3.4 Rebar Detection

The preprocessed point cloud contains various elements, including rebars, formwork surfaces, materials, and floor surfaces. To extract the rebar point cloud, unnecessary information must be removed. As shown in Figure 2(b), rebars exhibit an elongated linear shape in the point cloud, whereas other elements do not. To distinguish these, the linearity and planarity of the neighboring point cloud data were analyzed. For this analysis, the Radius Nearest-Neighbor Covariance (RNCC) method was applied [8]. RNCC computes the eigenvalues of the covariance matrix for points within a given radius *r*, representing the degree of data dispersion (see Equation (1)).

$$Cov = \frac{1}{N}\sum_{i=1}^{N}(p_i - \bar{p})(p_i - \bar{p})^T \quad (1)$$

In this content, N denotes the number of neighboring points within radius r, p_i is the coordinate vector of each neighboring point, and $\bar{p}$ is the centroid of the neighboring points. The first eigenvalue λ_1 represents the variance in the direction where the data is most dispersed. The second and third eigenvalues, λ_2 and λ_3, correspond to the variances in the second and third most dispersed directions, respectively.

By calculating the eigenvalues of the covariance matrix, the linearity (see Equation (2)) and planarity (see Equation (3)) of the point cloud can be evaluated. For linear data, λ_1 is large and λ_2 is small; therefore, a larger $\lambda_1 - \lambda_2$ value indicates greater linearity. For planar data, both $\lambda_1, and\ \lambda_2$ are large while $\lambda_3$is very small; thus, a larger $\lambda_2 - \lambda_3$ value indicates higher planarity.

$$L_\lambda = \frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2 + \lambda_3} \quad (2)$$

$$P_\lambda = \frac{\lambda_2 - \lambda_3}{\lambda_1 + \lambda_2 + \lambda_3} \quad (3)$$

Using the obtained linearity and planarity data, formwork surfaces with high planarity can be removed, while rebar point cloud data with high linearity can be segmented.

Figure 3 visually represents the distribution of linearity and planarity. As shown in Figure 3(c), the formwork appears in blue, indicating low linearity, while the rebars appear in red, indicating high linearity. In contrast, Figure 3(d) shows that the formwork has high planarity, whereas the rebars exhibit low planarity. By setting thresholds based on the distribution of linearity and planarity, filtering was applied to extract the rebar point cloud.

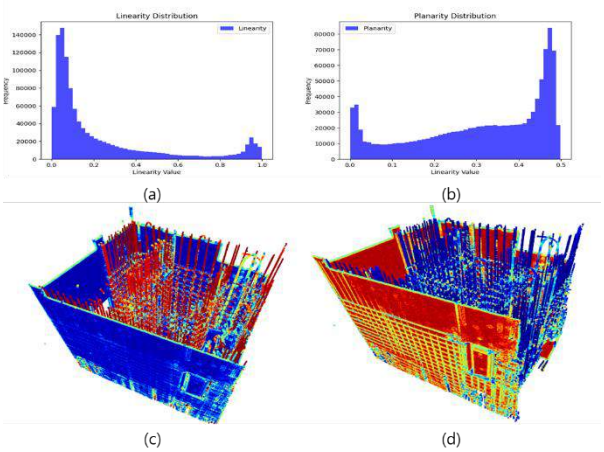


Figure 3 (a) Linearity distribution of point cloud; (b) Planarity distribution of point cloud; (c) Linearity expression of point cloud (red: high linearity, blue: low linearity); (d) Planarity expression of point cloud (red: high planarity, blue: low planarity)

A filtering threshold of 0.3 was applied to the linearity distribution, resulting in a point cloud that excludes the formwork and the floor surface, as shown in Figure 4. However, unnecessary noise still remained, such as construction materials.

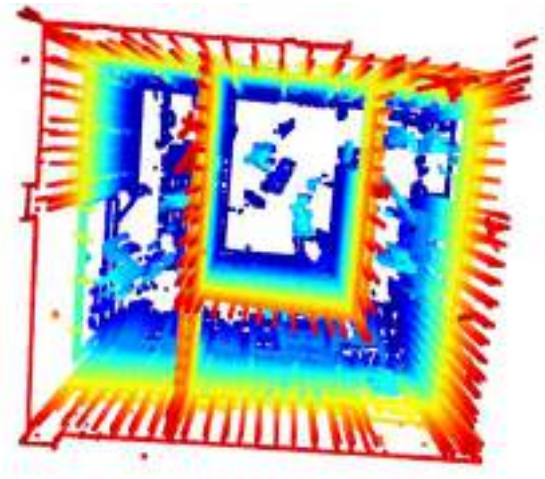


Figure 4 Filtered point cloud by linearity threshold

Such remaining noise in the point cloud needs to be removed, and the rebars arranged at various angles must be separated to calculate their spacing. To achieve this,

the RANSAC method was employed. RANSAC fits a specific model to the input data and identifies inliers that best conform to the model. Since rebar-arranged surfaces typically have a planar shape, fitting a plane model in a 3D space using RANSAC enables the extraction of the rebar-arranged surface, which is identified as inliers. By removing the extracted points and reapplying RANSAC to the remaining data, other rebar surfaces can be identified. This iterative process continues until no surface contains more than a predefined number of inliers, at which point RANSAC terminates. This method allows for the identification of all rebar surfaces while simultaneously removing noise, as illustrated in Figure 5.

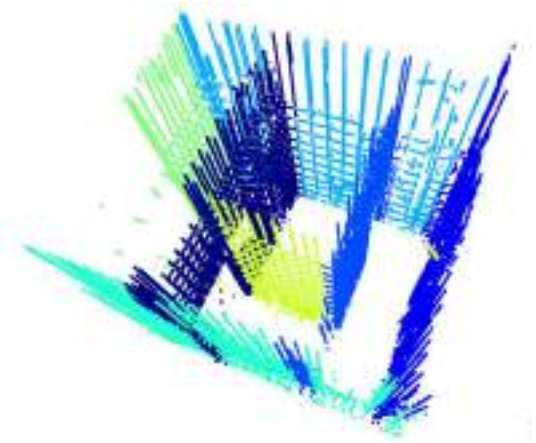


Figure 5 Iterative RANSAC of filtered point cloud

The rebar surface segments obtained through iterative RANSAC consist of rebars arranged in double layers. To further distinguish the two rebar surfaces, a plane is recalculated using RANSAC, and the point distribution by distance is analyzed, revealing two distinct peaks at different locations, as shown in Figure 6(a). To divide the data into two groups, Otsu Thresholding was applied. This method determines an optimal threshold by maximizing the between-class variance in a histogram, allowing for the separation of the double-layered rebar data (Figure 6(b)). The results are visualized in Figure 7.

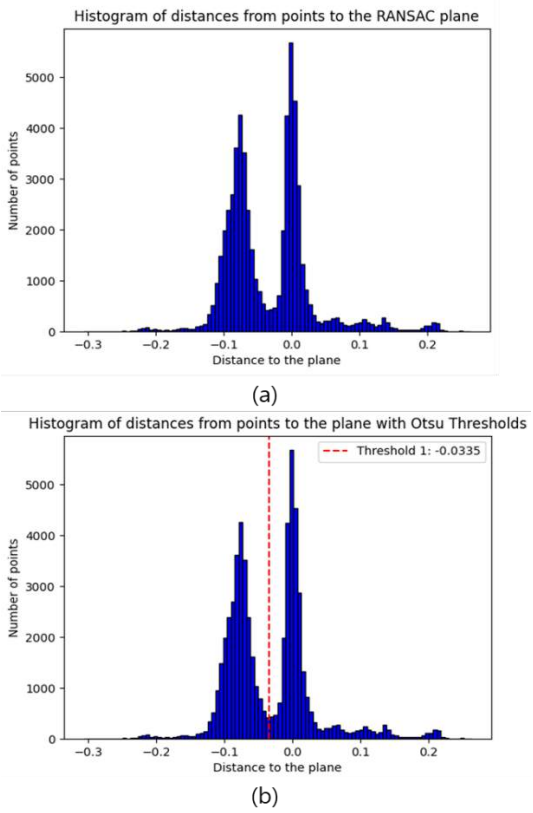


Figure 6 (a) Histogram of distances from points to the RANSAC plane; (b) Histogram of distances from points to the plane with Otsu Thresholds

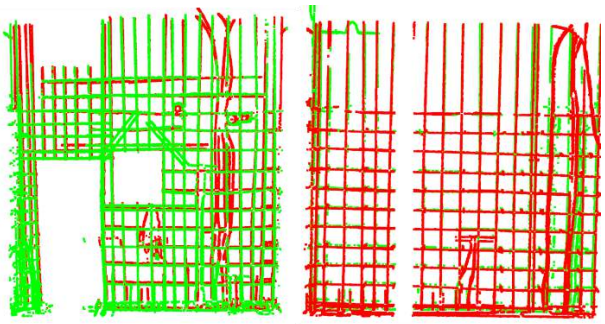


Figure 7 Visualization of double-layered rebar separation results

The study aimed to predict the rebar arrangement (i.e., spacing and diameter) of the actual double-layered rebar surfaces measured. Therefore, one-layered rebar surfaces separated in Figure 7 was used for further research.

3.5 Diameter & Spacing Prediction

Using the rebar point cloud obtained through the previous process, the rebar diameter and spacing could be estimated. In previous studies, rebar surfaces were

assumed to be perfectly planar, a feasible assumption in controlled laboratory environments. Under these conditions, RANSAC was used to separate horizontal and vertical rebars. However, as shown in Figure 8, the distance between the rebar point cloud and the RANSAC plane is visualized using color variations, with red indicating greater distances and blue representing shorter distances. This exhibits that the actual rebar surfaces in the field poses curvature rather than a perfect planar shape. This presents a limitation: in the presence of curvature, RANSAC can fit incorrect planes, leading to the spurious-plane problem [7]. This issue occurs in 3D point cloud segmentation, where the algorithm incorrectly detects planes that do not actually exist in the data. This limitation makes it difficult to apply the rebar segmentation methods proposed in previous studies to real-world construction sites. Accordingly, this study proposes a novel method for predicting the spacing of rebars that can be applied even when the rebars are bent in actual construction sites. Based on the rebar spacing data, we further suggest a methodology for estimating the rebar diameter.

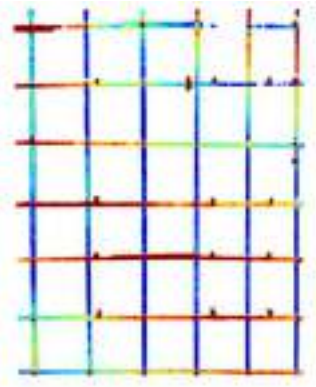


Figure 8 Visualization of distance from RANSAC plane (red: greater distance, blue: shorter distance)

3.5.1 Rebar Spacing Prediction

To estimate the spacing between rebars, the 3D points were projected onto a 1D axis aligned parallel to the rebar arrangement direction. Figure 9 shows histograms of the point distributions when the points are projected onto the horizontal and vertical axes. These histograms were normalized into curves using cubic spline interpolation, and the local maxima exceeding a specific threshold were identified, as shown in Figure 10. The locations of these local maxima on the normalized point density curve correspond to the centers of rebars aligned perpendicular to the projection axis. The spacing between horizontal and vertical rebars can thus be determined by calculating the distances between adjacent local maxima.

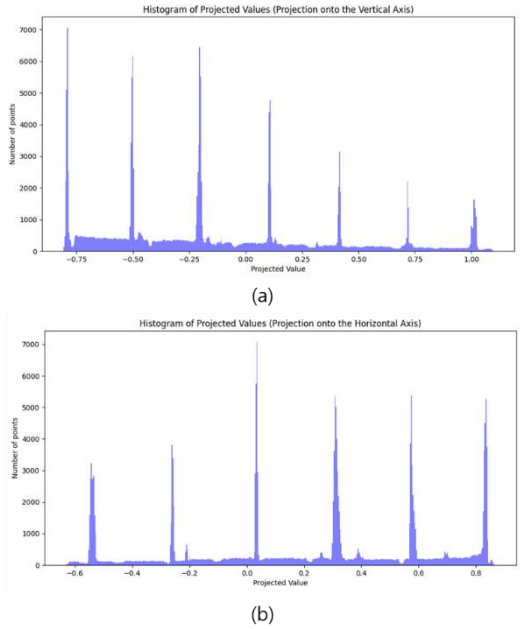


Figure 9 (a) Histogram of projected values (projection onto the vertical axis); (b) Histogram of projected values (projection onto the horizontal axis)

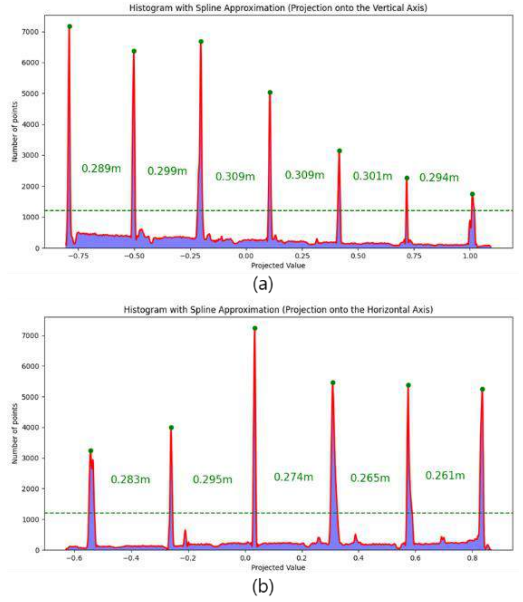


Figure 10 (a) Histogram with spline approximation and horizontal spacing prediction (projection onto the vertical axis); (b) Histogram with spline approximation and vertical spacing prediction (projection onto the horizontal axis)

The comparison between the predicted rebar spacing and the manually measured spacing is shown in Table 1.

Table 1 The comparison between the predicted rebar spacing and the manually measured spacing (mm)

		Spacing No.						Average error (mm)	Average accuracy (%)
		1	2	3	4	5	6		
Horizontal rebar spacing	Predicted	289	299	309	309	301	294		
	Measured	291	305	308	317	295	282		
	Error	2	6	1	8	6	12	6	98
Vertical rebar spacing	Predicted	283	295	274	265	261			
	Measured	284	287	282	268	252			
	Error	1	8	8	3	9		6	98

The measured spacing values refer to the manually measured spacing using a measuring tape. As shown in Table 1, the average spacing error for both horizontal and vertical rebars was 6 mm, which indicates an average error rate of 2%. This demonstrates that the proposed automated rebar spacing prediction method achieved an average accuracy of 98%.

3.5.2 Rebar Diameter Prediction

To estimate the diameter of the rebars, the cross-sectional data of each rebar was extracted. As shown in Figure 11, the segments between rebars were cropped based on the previously obtained rebar spacing information. These segments were then divided using Density-Based Spatial Clustering of Applications with Noise (DBSCAN). For each cluster, PCA was applied to calculate the first principal axis, which corresponds to the direction with the largest variance. The points were then projected onto a plane perpendicular to this principal axis. Since scanned rebars will show a circular shape in the projected points, a circle fitting algorithm could be used for diameter prediction.

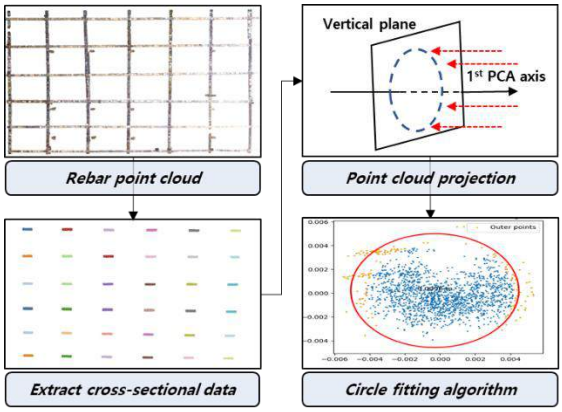


Figure 10 Processing steps for rebar diameter prediction

As the deformed rebars used at the construction site have ribs and lugs, the projected point clouds were generally distributed in circular or semi-circular shapes. However, in order to accurately estimate the diameter of each rebar segment, it is essential to use the outermost distributed points. This is because each segment contains almost no external noise, and the outermost points are likely to represent the outer surface of the rebar. Therefore, circle fitting based on outer points was applied to estimate the rebar diameters. Specifically, for each segment, a center point was determined, and RANSAC-based circle fitting was performed using the top 10% of points with the largest distances from the center. The result example is presented in Figure 12.

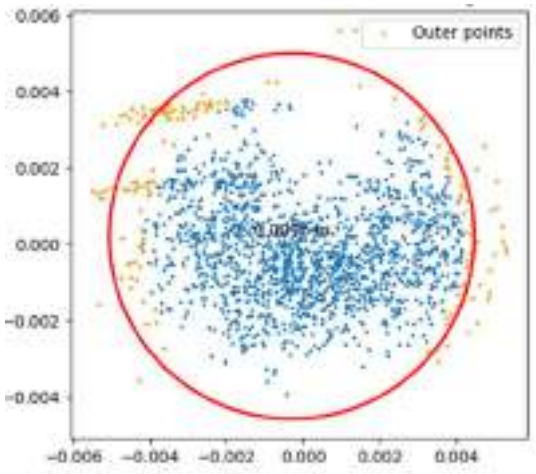


Figure 11 Segment diameter prediction example by extraction of outermost points and RANSAC circle fitting

The predicted rebar diameters were categorized into designations based on their nominal diameters, as shown in Table 2. Each segment was classified by matching its estimated diameter to the closest nominal diameter. These nominal diameters were used as the categories. The predicted diameters for each segment were classified into the nearest nominal diameter among these categories.

Table 2 Type of designation

Designation	Nominal diameter (mm)
D4	4.23
D5	5.29
D6	6.35
D8	7.94
D10	9.53
D13	12.7
D16	15.9
D19	19.1

Since each rebar was divided into multiple segments, the diameter of the rebar was determined by identifying the mode (i.e., most frequent value) of the class within the segment group. This approach helps mitigate prediction deviations caused by local noise in individual segments. It also provides a more effective basis for assessing the classification accuracy of each rebar into its correct nominal diameter, which is important for subsequent use in Scan-to-BIM applications. As shown in Figure 13, the predicted rebar diameters were 6 mm, 8 mm, 10 mm, or 13mm. At the segment level, the accuracy was 72.2%, while at the rebar level, the accuracy was 100%. In comparison, a previous study reported a segment level accuracy of 35.1% and a rebar level accuracy of 56% in estimating the 10 mm rebar diameters [2]. Therefore, the proposed method achieved significantly improved performance.

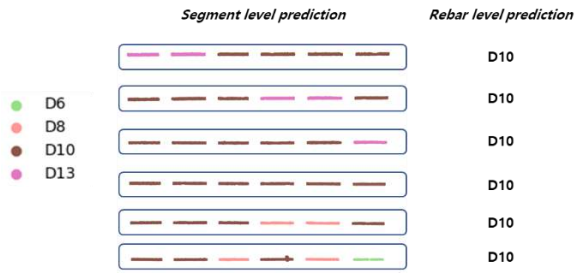


Figure 12 Rebar diameter prediction

4 Conclusion

4.1 Summary & Contribution

This study proposed a methodology for automating rebar arrangement inspection at construction sites. Unlike laboratory conditions, the research was conducted on an actual construction site where small-diameter rebars (10 mm) were arranged in double layers within a wall structure. First, a method for recognizing rebars in complex construction sites was proposed. The rebar was separated from non-rebar structural elements in the point cloud by utilizing the planar and linear characteristics of the objects. Subsequently, the rebar surfaces arranged in various directions were individually segmented using the RANSAC algorithm. A novel approach was introduced to estimate rebar diameters and spacing, even when the rebar surfaces were not perfectly planar. The accuracy of the proposed automated rebar spacing prediction was 98%, and the accuracy of rebar diameter prediction was 72.2% at the segment level and 100% at the rebar level.

Compared to thicker rebars, thinner rebars are more prone to bending or deformation. When the reinforcement surface is distorted, existing methods proposed in previous studies may no longer be applicable. Therefore, the proposed method can be particularly effective for detecting and analyzing small diameter rebars. Also, the proposed methodology can automate the complex and labor-intensive task of inspecting densely arranged rebars using laser scanners on real-world construction sites. Additionally, comparison and analysis with design specifications can ensure safe and accurate rebar arrangement inspection.

The proposed methodology can be further integrated with Scan-to-BIM technologies to support the automated generation of 3D models from scanned point cloud data. Through this process, rebar arrangement results can be accurately reflected in the digital model, enabling the classification of rebar types and the recognition of individual rebar elements within the 3D environment. Such integration allows for more precise modeling of the construction status and facilitates the identification of various components on-site.

4.2 Limitations

Since there were no rebars smaller than 10 mm in the scanned site, the current study was limited to 10 mm-diameter rebars. Further research will be conducted on sites with rebars of smaller or larger diameters.

When extracting rebar information from the point cloud using its linear or planar characteristics, information loss can occur in overlapping regions of rebars. This is due to the reduced linearity caused by distribution of adjacent points in both vertical and horizontal directions. Thus, further work will focus on constructing datasets that include rebar intersection points to enhance the applicability of the proposed method. Additionally, the method for calculating rebar spacing assumes that rebars are arranged vertically; thus, the proposed approach has limitations when rebars are placed diagonally.

Diameter and spacing estimation were performed only for a subset of the separated rebar segments, rather than for the entire dataset. This is because the current method could not effectively handle noise-included data, where other materials were not fully removed from the segmentation process. Future research will focus on addressing this issue.

5 Acknowledgement

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. RS-2023-00241758).

References

[1] G. A. Washer and C. A. Chang. Guideline for implementing quality control and quality assurance for bridge inspection, *National Cooperative Highway Research Program*, Project 20-07, Task, 252, 2009.

[2] M.-K. Kim, J. P. P. Thedja, H. L. Chi, and D. E. Lee. Automated rebar diameter classification using point cloud data based machine learning. *Automation in Construction*, 122:103476, 2021.

[3] M.-K. Kim, J. P. P. Thedja, and Q. Wang. Automated dimensional quality assessment for formwork and rebar of reinforced concrete components using 3D point cloud data. *Automation in Construction*, 112:103077, 2020.

[4] H. L. Chi, M. K. Kim, K. Z. Liu, J. P. P. Thedja, J. Seo, and D. E. Lee. Rebar inspection integrating augmented reality and laser scanning. *Automation in Construction*, 136:104183, 2022.

[5] X. Yuan, A. Smith, R. Sarlo, C. D. Lippitt, and F. Moreu. Automatic evaluation of rebar spacing using LIDAR data. *Automation in Construction*, 131:103890, 2020.

[6] X. Yuan, A. Smith, F. Moreu, R. Sarlo, C. D. Lippitt, M. Hojati, S. Alampalli, and S. Zhang. Automatic evaluation of rebar spacing and quality using LiDAR data: Field application for bridge structural assessment. *Automation in Construction*, 146:104708, 2023.

[7] L. Li, F. Yang, H. Zhu, D. Li, Y. Li, and L. Tang. An Improved RANSAC for 3D Point Cloud Plane Segmentation Based on Normal Distribution Transformation Cells. *Remote Sens*, 9(5):433, 2017.

[8] W. Zhao, Y. Jiang, Y. Liu, and J. Shu. Automated recognition and measurement based on three-dimensional point clouds to connect precast concrete components. *Automation in Construction*, 133:104000, 2022.

[9] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, *Commun*, ACM 24 (6):381-395, 1981.

[10] M.-K. Kim, Q. Wang, H. Li. Non-contact sensing based geometric quality assessment of buildings and civil structures: a review, *Automation in Construction*, 100:163–179, 2019.

[11] Q. Wang, M.-K. Kim. Applications of 3D point cloud data in the construction industry: A fifteen-year review from 2004 to 2018. *Advanced Engineering Informatics*, 39:306-319, 2019.

[12] Q. Wang, Y. Tan, and Z. Mei, Computational methods of acquisition and processing of 3D point cloud data for construction applications, *Archives of Computational Methods in Engineering*. 27:479–499, 2020.

[13] V. Kasireddy and B. Akinci. Assessing the impact of 3D point neighborhood size selection on unsupervised spall classification with 3D bridge point clouds. *Archives of Computational Methods in Engineering*, 52:101624, 2022.

[14] C. Zhang, J. Shu, H. Zhang, Y. Ning, and Y. Yu. Estimation of load-carrying capacity of cracked RC beams using 3D digital twin model integrated with point clouds and images. *Engineering Structures*, 310:118126, 2024.

[15] J. Shu, C. Zhang, K. Yu, M. Shooshtarian, and P. Liang. IFC-based semantic modeling of damaged RC beams using 3D point clouds. *Structural Concrete*, 24(1):389-401, 2022.

[16] R.B. Rusu. Semantic 3D object maps for everyday manipulation in human living environments. *KI - Künstliche Intelligenz*, 24:345-348, 2010.

[17] F. Li, M.-K. Kim, and D.-E. Lee. Geometrical model based scan planning approach for the classification of rebar diameters. *Automation in Construction*, 130:103848, 2021.

Automated defect detection in clay printing

Patricia Peralta*, Thamer Al-Zuriqat, Mahmoud Noufal, and Kay Smarsly

Institute of Digital and Autonomous Construction, Hamburg University of Technology, Germany
patricia.peralta.abadia@tuhh.de

Abstract

Additive manufacturing (AM) of eco-friendly materials has the potential to decarbonize the construction industry by enabling the creation of complex structures with minimal waste. Clay has been integrated into AM processes as a building material, giving rise to an emerging research field referred to as "clay printing". Defects, such as tearing and sagging, are common in clay printing and could affect the structural integrity, load-bearing capacity, and overall durability of the structures. However, limited research on defect detection in clay printing and lack of datasets restrict the development of defect detection models. This paper presents a tool – the automated defect detection (ADD) preprocessor – developed to generate a dataset for defect detection models in clay printing. The tool uses images and videos as input for preprocessing and labeling images required to build the dataset, meeting the requirements of defect detection models based on convolutional neural networks. The ADD preprocessor is implemented and validated as a proof of concept for clay printing processes. The results demonstrate the capability of the ADD preprocessor to successfully build a dataset for the deployment of defect detection models in clay printing.

Keywords –

Clay printing; additive manufacturing; defect detection; dataset; convolutional neural networks

1 Introduction

Additive Manufacturing (AM) has become a significant innovation in the construction industry due to its numerous advantages that surpass "traditional" construction methods [1]. One key aspect, compared to traditional construction methods is the ability to create complex and customized structures without additional cost or material waste [2]. The layer-by-layer construction process enables the realization of geometrically intricate designs and freeform structures that are challenging or impossible to achieve with conventional formwork or machining methods [3]. Furthermore, additive manufacturing of eco-friendly materials has the potential to decarbonize the construction industry [4].

Clay, a traditional building material, has been adapted for AM processes and has emerged as a new research field, referred to as "clay printing" [5]. Clay printing often encounters defects during printing, such as tearing and sagging, that could affect the structural integrity, load-bearing capacity, and overall durability of the structures. By exploring synergies with concrete printing, defects in clay printing may be investigated. Defects in concrete printing, including tearing (breakpoints due to low workability of the material) [6], sagging/buckling (due to deformation under self-weight) [6] and shrinkage cracking [7], can also be observed in clay printing. However, research on defect detection in clay printing remains limited, which hinders the development of defect detection models due to insufficient data.

Efforts towards automated defect detection approaches for AM processes have been reported, where most research has been focusing on fused filament fabrication (FFF) and, more recently and to a lesser extent, on concrete printing. For FFF processes, recent defect detection approaches for offline and real-time detection have been developed based on computer vision techniques, including comparing point clouds [8] and images [9] as well as on convolutional neural network (CNN) architectures [10, 11]. Similarly, approaches based on computer vision, supplemented by machine learning and deep learning, have been developed for concrete printing to detect layer deformations [12] and to assess geometries of additively manufactured structures [13]. Further approaches devised for real time defect detection in concrete printing have been developed based on detection transformers to detect and measure tearing with high accuracy [14]. Yet current defect detection approaches present several limitations. On the one hand, vision-based systems are used in real-time defect detection to monitor printing processes and detect defects, but are limited to the view point of the vision-based systems. On the other hand, defect detection models based on deep learning, such as CNN classification models, may be trained to detect and classify specific defects, where the scope and accuracy of the defect detection models are dependent of the datasets available for training [11]. Consequently, the need for datasets arises, due to the limited availability of data regarding

defects in clay printing.

In this paper, the development of a tool for automated defect detection (ADD) is presented. The tool, named "ADD preprocessor", is designed to build a dataset aiming to support defect detection models in clay printing using images and videos as input data. The image data is preprocessed and labeled to build the dataset, which meets the requirements of defect detection models based on convolutional neural networks. The ADD preprocessor is implemented and validated as a proof of concept for clay printing processes. The remainder of the paper is structured as follows. First, the methodology for deep-learning-based defect detection and the implementation of the ADD preprocessor is presented. Next, the proof of concept is described, followed by a discussion of the conclusions and potential future research.

2 Deep-learning-based defect detection model

In this section, the methodology for generating deep-learning-based defect detection models is described. The methodology is based on a previous study of the authors (described in [15]), and it is adapted for clay printing. Considering the lack of available data for clay printing, the ADD preprocessor is designed and implemented as a tool for generating datasets for defect detection models. Below, a short description of the research methodology and the implementation for the ADD preprocessor is presented.

2.1 Research methodology

The research methodology is proposed to generate a deep-learning-based defect detection model using a CNN architecture. As described in Figure 1, the creation of deep-learning-based defect detection models usually comprises two phases, the (i) data preparation phase and (ii) classification model development phase. The ADD preprocessor focuses on the data preparation phase (highlighted in orange in Figure 1) and generates a dataset D_n as an output using images and videos as input data.

In the **data preparation phase**, data is collected, preprocessed, labeled, augmented, and split. During *data collection*, *n* images of additively manufactured clay structures with and without defects are part of the dataset D_n. The images included in the dataset D_n may be collected from different sources, where the images may vary in size, color, and viewing angle. Next, during *data preprocessing*, the images are cropped, resized to a predefined size, transformed into a gray scale, and corrected for orientation to a consistent horizontal view. The gray-scale images are then *labeled* with the corresponding defect class: Tearing, sagging, buckling, and shrinkage cracking. Finally, for *data augmentation*, augmentation techniques, such as mirroring, noise, light exposure ("brightness"), rotating and random image cutouts, are applied to increase the size of the dataset D_n. Mirroring and rotating are used to improve robustness to random camera angles, while noise, cutouts, and variable exposure increase resilience to varying lighting conditions. For the next phase, dataset D_n is *split* into three sub sets, a training dataset ($D_{training}$), a validation dataset ($D_{validation}$), and a testing dataset ($D_{testing}$).

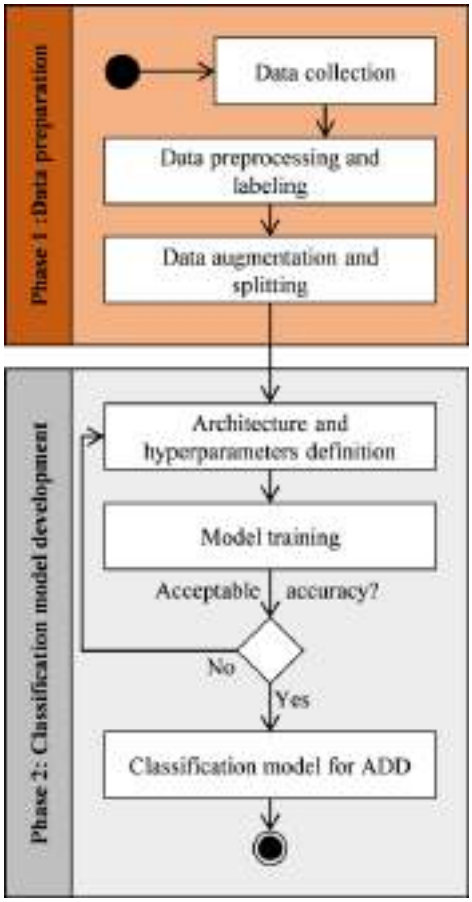


Figure 1. Methodology for generating deep-learning-based defect detection models [15]

In the **classification model development phase**, the classification model is developed in three steps, architecture definition, model training, and model testing. For *architecture definition*, the architecture of the classification model is selected and hyperparameters are defined. An architecture commonly used is the "You Only Look Once" (YOLO) architecture, which prioritizes speed, accuracy, and user-friendliness while preserving computational efficiency [14]. Upon defining the CNN architecture, the training dataset $D_{training}$ is deployed to train the CNN classification model. During

model training, the model "learns" from existing relationships between known inputs (images) and known outputs (defect classes, e.g. "sagging"). Furthermore, loss metrics are used to guide the optimization of the CNN classification model and to evaluate the error between the predictions of the model and the actual target outcomes. In general, a decreasing loss value trending toward zero indicates that the model is effectively "learning" from the training dataset. Meanwhile, and to prevent overfitting during training process, the validation dataset $D_{validation}$ is applied to fine-tune and evaluate the performance of the models during training. Upon completing training, the testing dataset $D_{testing}$ is used to test the performance of the classification model until achieving an acceptable accuracy. During *model testing*, evaluation metrics are used, including precision, recall and the mean average precision at a single intersection-over-union threshold, to test the capacity of the model aiming to avoid false positive and false negative classifications and to expresses the prediction accuracy of the model.

2.2 Implementation

Acquiring large datasets for clay printing is challenging because of to the lack of research on defects in clay printing in the context of the construction industry. Existing labeling tools are limited to only annotate images, requiring already standardized images as inputs and additional tools for data augmentation. Therefore, the ADD preprocessor aids in collecting and standardizing images stemming from various sources, as well as to labeling and augmenting the image data to generate a dataset. To satisfy the requirements for deep-learning-based defect detection models in the context of clay printing, the following requirements are considered:

- *Image size*: CNN classification models require standardized image sizes to improve accuracy and reduce complexity [12]. Images are resized, either by downscaling or upscaling, to an optimal image size that minimizes computational cost while maximizing feature resolution. To ensure efficient batch training and adequate feature resolution, an image size of 640x640 pixels is predefined as the optimal image size for this study.
- *Colorblind model*: Variations in the material color may affect the accuracy of the defect detection models [12, 14]. By using gray-scaled images, it is possible to prevent correlating material color with defect classes, as well as to increase the computational speed of the CNN model. In gray-scale images, each pixel represents a specific light intensity, ranging from black to white, with no color information.
- *Defect labeling*: Inconsistent rheological and mechanical properties of the material, dimensions of the extruded layers, and the number of stacked layers dictate the characteristics of defects, as discussed in [14]. However, details that might resemble defects may have been intentionally designed for esthetics and function, such as water protection, shading, and ventilation. Therefore, design intentions of additively manufactured structures are considered when identifying defects during the labeling process. The labeled images are accompanied by annotation files with the same name, containing the coordinates of bounding boxes that describe the location of defects and labels of the corresponding defect types.
- *Data augmentation*: Augmentation techniques are employed to increase the size of the dataset and to improve the robustness of the CNN classification models. Random camera angles are introduced to the dataset by mirroring and rotating images in the range of [-20°, 20°]. Varying lighting conditions are introduced to the dataset by adding noise to 4% of the gray-scale pixel intensity values and by adjusting brightness in the range of [-35%, 35%].

A labeling strategy is defined for the ADD preprocessor. For each defect type, a corresponding label and index number is assigned (Table 1). Defects are manually annotated with a bounding box and labeled with the corresponding index number. The annotation is then saved in a CSV file.

Table 1. Defect labeling for clay printing

Label	Index number
No defect	0
Tearing	1
Sagging	2
Buckling	3
Shrinkage cracking	4

The ADD preprocessor is implemented using Python programming language, together with a graphical user interface (GUI) for manual user inputs and processing algorithms for resizing input images, adjusting view angles, converting the images to gray scale, and augmenting the data. The GUI, as shown in Figure 2, facilitates generating datasets by importing and saving images into user-defined file directory ("source folder" and "save folder"), as well as automating the preprocessing steps with three main trigger buttons ("crop", "process", and "save").

When importing an image into the ADD preprocessor, areas of interest of the additively manufactured structures can be manually selected using the GUI and cropped with the crop button. When importing videos, a previous step is performed to capture a frame every half a second. With the process button, processing algorithms are triggered.

First, the areas of interest are centered and resized to the predefined optimal image size. Next, the viewing angle of the printed product is adjusted to ensure horizontal orientation, and the image is converted to gray-scale. Then, the generated gray-scale images are labeled with the corresponding defect class (tearing, sagging, buckling, and shrinkage cracking). Finally, the generated gray-scale images are augmented by mirroring, adding noise (randomizing 4% of the gray-scale pixel intensity values), adjusting brightness (± 35%), and rotating (± 20°) the images. With the save button, the output images and corresponding labels and annotation files are saved for the classification model development phase.



Figure 2. ADD preprocessor interface

3 Proof of concept

In this section, the ADD preprocessor functionality for generating datasets for clay printing is validated through a proof of concept. The proof of concept involves generating a dataset based on images collected from experimental tests conducted under laboratory conditions with a constant temperature of 20 °C and consistent lighting. The robotic system used for clay printing is a Potterbot SCARA v4 with a linear ram extruder and 3.5-liter extrusion tubes capable of printing highly plastic clays (Figure 3). Experimental tests for evaluating the extrudability and buildability of six clay mixtures are used to collect image data for defect detection. The clay mixtures are designed with clay contents varying between 20% and 30%, and various defects that might affect extrudability and buildability are expected during the printing process.

The images are captured with a camera with a maximum resolution of 6000x4000 pixels (Sony Alpha 6000) that is equipped with a lens with a focal length of 24-105 mm (Sony FE 24-105mm F4 G OSS) and fixed to a tripod. The focus and lens aperture are manually adjusted to ensure stable and clear image quality. Images are taken from various viewpoints, specifically highlighting defects observed during the printing process. From the images collected for validation, a total of 16 images are selected as input for the ADD preprocessor as proof of concept, from which 5 images display buckling defects, 3 images display sagging defects, 1 image displays tearing defects, and 7 images present no defects on the additively manufactured structures.

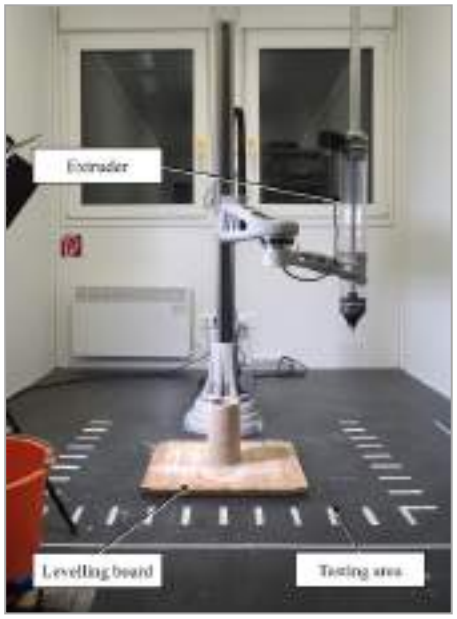


Figure 3. Robotic system for clay printing

As a result, the ADD preprocessor allows selecting areas of interest from the input images and label defects (Figure 4). The area of interests, marked exemplarily with a yellow box in Figure 4, are selected manually for cropping, focusing on the defects. Upon cropping, the defects are labeled and annotated with the corresponding defect types. As shown in Figure 4, a red bounding box annotates a sagging defect. The predefined image size of 640x640 pixels ensures adequate feature resolution to

identify defects by using the context given by the pixels in the neighborhood.

The output of the data augmentation has increased the dataset size from 16 images to 64 images, providing variability in the dataset, including various viewing angles, lighting conditions, and noise. An example of the output of data augmentation for a sagging defect is shown in Figure 5. By preprocessing the images for generating the dataset, including cropping, resizing, gray-scaling and data augmentation, the ADD preprocessor not only standardizes the dataset but also prepares it for efficient training of the CNN classification model. Owing to the data augmentation techniques employed in the ADD preprocessor, the CNN classification model can be trained to have a suitable performance under random camera angles and varying lighting conditions, while avoiding overfitting.

The dataset presents positive instances that are classifiable according to defect type (i.e. images with defects) and negative instances with no detectable defects (i.e. images without defects). By including negative instances into the dataset, the defect detection model is enabled to learn patterns to not lead to false positive predictions. Even though the proof of concept has resulted in a dataset with a relatively small size, the dataset can be enlarged by collecting more images, e.g. obtained from open-access sources and from further experimental test and printings conducted in research.

By addressing the current lack of datasets for clay printing, the ADD preprocessor enables the development of advanced defect detection models that can improve the structural quality and reliability of additively manufactured clay structures. Furthermore, the ADD preprocessor can be easily employed for concrete printing due to the synergies detected among defects in clay printing and defects in concrete printing.

4 Summary and conclusions

Automated defect detection in clay printing has been hindered due to insufficient data on defects in clay printing, driving the development and validation of the ADD preprocessor, a tool designed to generate datasets to support defect detection models in clay printing. By leveraging images and videos, the ADD preprocessor processes and labels image data to create standardized

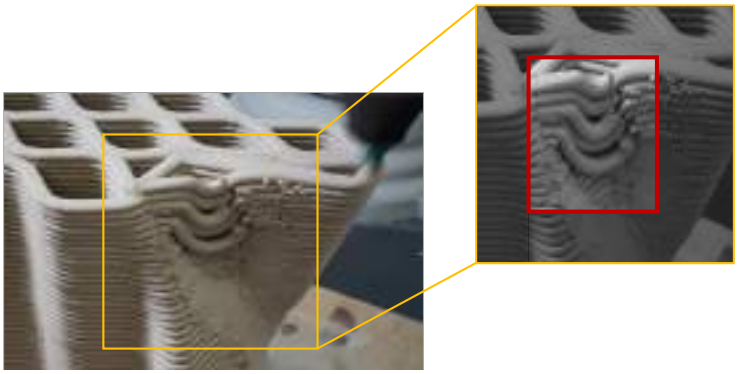


Figure 4. Identification of the area of interest (yellow box) for a “sagging” defect (red box)

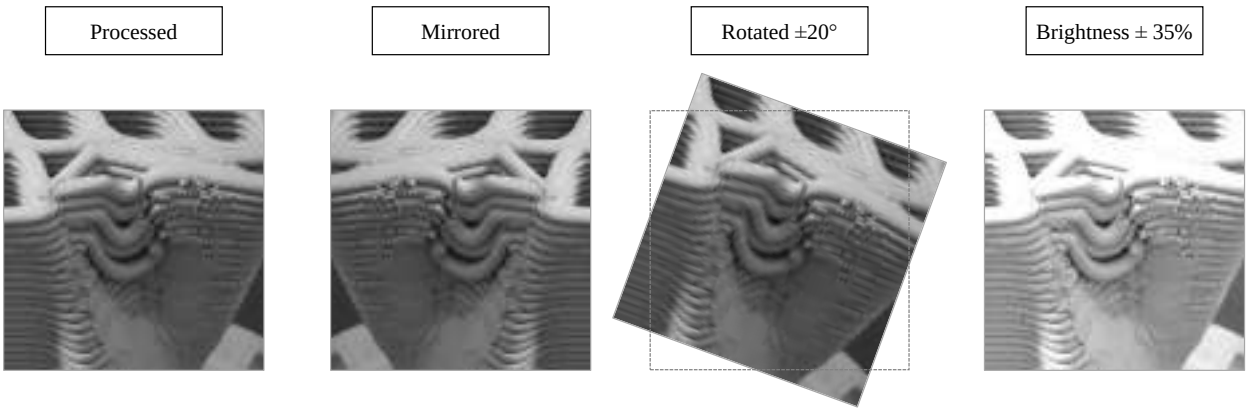


Figure 5. Outcome of the data augmentation for a “sagging” defect

datasets that meet the requirements of deep-learning-based defect detection models. The proof of concept has demonstrated the effectiveness of the ADD preprocessor in preparing positive instances that facilitate the identification and the categorization of defects (i.e., sagging, tearing, and buckling) as well as the inclusion of negative instances for more robust model training.

The ADD preprocessor possesses the ability to generate a structured dataset for defect detection in clay printing. The results showcase the potential of the ADD preprocessor to address the lack of datasets for defect detection, which has hindered the development of autonomous defect detection in clay printing. By providing a streamlined workflow for data preparation (i.e. *data preprocessing* and *data augmentation*), the ADD preprocessor supports the creation of defect detection models that are better equipped to ensure the structural integrity and quality of additively manufactured clay structures. Compared to existing labeling tools, the ADD preprocessor provides control over image standardization and data augmentation while implementing a similar labeling strategy to existing labeling tools. Yet, the manual inputs required by the ADD preprocessor, such as selecting areas of interest, could introduce subjectivity into the dataset.

To further drive the research towards automated defect detection in clay printing, future efforts could be directed towards real-time defect monitoring and towards testing the compatibility of the datasets generated by the ADD preprocessor with advanced neural network architectures, such as transformer-based models. Automated defect detection has the potential to advance the broader adoption of clay printing as a sustainable and reliable construction method.

Acknowledgements

The authors gratefully acknowledge the financial support provided by the German Research Foundation (DFG) through grant SM 281/22-1, by Hamburg University of Technology through the I3 Program under the junior project “Digitalization of earth printing to advance climate-informed engineering” as well as by the Federal Ministry for Digital and Transport (BMDV) within the mFUND program under grants 01FV2013B, and 01FV2059C. Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the sponsors mentioned above.

References

[1] P. Peralta, and K. Smarsly. Requirements analysis of additive manufacturing for concrete printing – A systematic review. In *Proceedings of the* 39*th International Symposium on Automation and Robotics in Construction* (*ISARC*). 336–343, Bogota, Colombia, 2022.

[2] K. Smarsly, P. Peralta, D. Luckey, S. Heine, and G.-M. Ludwig. BIM-based concrete printing. In *Proceedings of the International ICCCBE and CIB W*78 *Joint Conference on Computing in Civil and Building Engineering*, Sao Paolo, Brazil, 2020.

[3] A. Wolf, P.L. Rosendahl, and U. Knaack. Additive manufacturing of clay and ceramic building components. *Automation in Construction*, 133(2022):103956, 2022.

[4] J. Vollmert, P. Peralta, A. Tandon, T.L. Harder, H. Al-Nasser, and K. Smarsly. Digital-twin-based monitoring and simulation of robot kinematics for clay printing. In *Proceedings of the* 15*th European Conference on Product and Process Modeling* (*ECPPM*), Dresden, Germany, 2024.

[5] M. Gomma, W. Jabi, V. Soebarto, and Y.M. Xie. Digital manufacturing for earth construction: A critical review. *Journal of Cleaner Production*, 338(2022):130630, 2022.

[6] R.A. Buswell, W.L. De Silva, S.Z. Jones, and J. Dirrenberger. 3D printing using concrete extrusion: A roadmap for research. *Cement and Concrete Research*, 112(2018):37–49, 2018.

[7] G.M. Moelich, J. Kruger, and R. Combrinck. Plastic shrinkage cracking in 3D printed concrete. *Composites Part B*: *Engineering*, 200(2020): 108313, 2020.

[8] X. Zhao, Q. Li, M. Xiao, and Z. He. Defect detection of 3D printing surface based on geometric local domain features. *The International Journal of Advanced Manufacturing Technology*, 125(1):183–194, 2023.

[9] J. Xie, A. Saluja, A. Rahimizadeh, and K. Fayazbakhsh. Development of automated feature extraction and convolutional neural network optimization for real-time warping monitoring in 3D printing. *International Journal of Computer Integrated Manufacturing*, 35(8): 813–830, 2022.

[10] M.F. Khan, A. Alam, M.A. Siddiqui, M.S. Alam, Y. Rafat, N. Salik, and I. Al-Saidan. Real-time defect detection in 3D printing using machine learning. *Materials Today*: *Proceedings*, 42(2021): 521-528, 2021.

[11] L. Xu, X. Zhang, F. Ma, G. Chang, C. Zhang, J. Li, S. Wang, and Y. Huang. Detecting defects in fused deposition modeling based on improved YOLO v4. *Materials Research Express*, 10(9):095304, 2023.

[12] O. Davtalab, A. Kazemian, X. Yuan, and B. Khoshevis. Automated inspection in robotic additive manufacturing using deep learning for layer deformation detection. *Journal of Intelligent Manufacturing*, 33(3):771–784, 2022.

[13] S.A. Nair, G. Sant, and N. Neithalath. Mathematical morphology-based point cloud analysis techniques for geometry assessment of 3D printed concrete elements. *Additive Manufacturing*, 49(2022):102499, 2022.

[14] H. Zhao, J. Sun, X. Wang, Y. Wang, Y. Su, J. Wang, and L. Wang. Real-time and high-accuracy defect monitoring for 3D concrete printing using transformer networks. *Automation in Construction*, 170(2025): 105925, 2025.

[15] T. Al-Zuriqat, M. Noufal, P. Peralta, K. Dragos, and K. Smarsly. Automated defect detection in fused filament fabrication coupling deep learning and computer vision. In *Proceedings of the 2025 European Conference on Computing in Construction* (*EC*3). Porto, Portugal, 2025 (submitted).

Improved Information Extraction from Bridge Inspection Reports using Fine-tuned Generative Pre-trained Transformers

Abdelhady Omar[1] and Osama Moselhi[2]

[1] Ph.D. Candidate, Department of Building, Civil and Environmental Engineering, Concordia University, Montréal, QC, Canada.
[2] Professor and Director, Centre for Innovation in Construction and Infrastructure Engineering and Management (CICIEM), Department of Building, Civil and Environmental Engineering, Concordia University, Montréal, QC, Canada.
Email: abdelhady.omar@mail.concordia.ca, moselhi@encs.concordia.ca

Abstract –

Bridge inspection reports contain a wealth of crucial data on bridge components and their related structural defects. This study introduces a novel method that harnesses the power of Generative Pre-trained Transformers (GPT) for improved information extraction from bridge inspection reports. While most studies in this domain focus solely on data extraction, this study transforms inspection data into a ready-to-use format for better utilization in condition assessment and predictive modeling, enabling better-informed budget allocation and decision-making. It employs (1) a baseline GPT model (BL-GPT), which leverages OpenAI's large language models to process textual inspection data through optimized prompt engineering, and (2) a fine-tuned GPT model (FT-GPT), which enhances the baseline by incorporating domain-specific training to improve performance. These models capture and evaluate the severity levels of reinforced concrete bridge defects based on textual inspection data. The models are validated on data extracted from 2,255 inspection reports—spanning a period of five years (2018–2022)—for a set of bridges in Québec, Canada. The FT-GPT is found to significantly improve performance, stability, and reliability in detecting and standardizing the severity of different types of concrete defects in bridge decks. In specific, it achieves accuracy rates of 98.79% for rebar corrosion, 99.09% for delamination, and 98.64% for cracking, scaling, and spalling of concrete. This study demonstrates the potential integration of generative AI in asset management, an application that has yet to be realized.

Keywords –

Bridge Inspection Reports; Infrastructure Asset Management; Concrete Bridge Decks; Natural Language Processing (NLP); Large Language Models (LLMs); Generative AI (GenAI); Generative Pre-trained Transformers (GPTs); Prompt Engineering; Fine-tuning.

1 Introduction

Bridges are a critical civil infrastructure enabling mobility, fostering social connectivity, and facilitating economic growth. However, their structural integrity is compromised by several factors, including harsh environmental conditions, freeze–thaw cycles, exposure to de-icing chemicals, deferred maintenance, increasing traffic volumes, and aging [1]. Preventive measures are essential to ensure these structures do not deteriorate to the point of compromising safety or requiring closures. Bridge deterioration is manifest in defects such as rebar corrosion, cracking, delamination, scaling, and spalling of concrete, with data on the type, location, and severity of these defects collected during bridge inspections. However, there are significant challenges concerning the effective utilization of bridge inspection data, as it is often scattered across unstructured textual formats in inspection reports [2, 3]. The extraction, structure, and characterization of this data are further complicated by the need for automated, data-driven analyses to support condition assessments, predictive modeling, and decision-making. These barriers hinder the development of reliable asset management tools, delaying preventive and corrective actions and jeopardizing bridge safety [4].

Natural Language Processing (NLP) offers a promising solution by enabling the extraction of critical information from unstructured textual data [5, 6]. Advances in Large Language Models (LLMs) and Generative Artificial Intelligence (GenAI) have further enhanced NLP's capabilities, making it increasingly effective in processing and analyzing complex technical information [7, 8]. These cutting-edge AI techniques can automate the extraction and "standardization" of bridge inspection data. In the present study, "standardization" refers to the systematic transformation of unstructured and heterogeneous bridge inspection data into a consistent, ready-to-use, machine-readable format, enabling better-

informed condition assessments, predictive modeling, and maintenance planning, ultimately improving the reliability and efficiency of bridge asset management. While numerous studies have proposed methods to extract and structure data from bridge inspection reports, e.g., [2, 9-13], a significant gap remains with respect to transforming this structured data into formats ready for direct use in condition assessment and predictive modeling.

Liu and El-Gohary [14] proposed a hybrid data fusion method combining named entity normalization and entropy-based numerical fusion to unify bridge deficiency data. Similarly, Liu and El-Gohary [15] introduced a spectral clustering-based approach to link records describing the same bridge components or deficiencies. Li et al. [10] developed a context-aware deep-learning framework using bi-directional long short-term memory and conditional random fields to extract defect-related information, although it is noteworthy that their approach requires further standardization to align the extracted data with predictive modeling requirements. Zhang et al. [13] combined rule-based methods with generation-based LLMs to structure bridge inspection data, but noted that, without standardization, the extracted information remains unsuitable for advanced analytics. The present study takes the next step by converting extracted bridge inspection data into fully standardized formats that directly support more accurate condition assessment and predictive modeling.

LLMs, which are trained on vast datasets, excel in tasks such as translation, summarization, and question-answering by capturing nuanced language patterns and long-range dependencies [16-18]. GenAI extends LLM capabilities to generate novel content across different modalities, including text, images, and videos, using generative adversarial networks and advanced transformer architectures [8, 19]. OpenAI's GPT models, including GPT-3.5 and GPT-4, among others, exemplify these capabilities. GPT-3.5, with 175 billion parameters [17], is capable of efficiently performing a wide range of tasks [20], while GPT-4, with significantly more parameters (estimated to be in the trillions), offers enhanced accuracy and versatility [8, 21]. These models have been applied across diverse domains, such as integrating ChatGPT with augmented reality for operation and maintenance guidance [7], and enabling BIM-based virtual assistants [22].

Nevertheless, the following gaps are identified: (1) *lack of standardization in inspection data*—existing studies focus on data extraction but fail to standardize it for applications such as condition assessment and predictive modeling; (2) *underutilized potential of GenAI models, such as GPTs, beyond their role as chatbots*—GenAI's capabilities in asset management remain untapped, with conventional NLP approaches struggling to handle the complexity and variability of inspection data; and (3) *need for fine-tuned domain-specific models*. In this regard, general-purpose GenAI models lack precision to interpret technical language in inspection reports, whereas fine-tuned, systematic models could significantly improve this process.

2 Objectives

This study introduces a novel method leveraging GPT to standardize inspection data for improved information extraction, enabling automated condition assessments and predictive modeling. This is achieved through the following sub-objectives: (1) developing unified guidelines for defect identification and classification to ensure consistency; (2) harnessing a baseline GPT using optimized prompt engineering for data standardization; (3) fine-tuning the baseline with a large volume of domain-specific bridge inspection data; and (4) conducting a comprehensive comparative analysis between the two models.

3 Developed Method

An overview of the method developed in the present study is depicted in Figure 1. Details of each step are provided in the following subsections.

3.1 Case Study and Data

This study focused on bridge decks and their structural elements for a set of concrete bridges in Québec that are among the oldest in Canada [23] and are managed by the Ministry of Transport and Sustainable Mobility of Québec (MTQ). The analysis was concentrated on principal deck elements (i.e., structural slab and exterior sides) [24]. Concrete defects, i.e., corrosion, cracking, delamination, scaling, and spalling, were evaluated. In this respect, two sets of data were collected: the first set consisting of inspection comments extracted from bridge inspection reports and the latter consisting of defect severity thresholds extracted from the MTQ *Structures Inspection Manual* [24].

The dataset, extracted from 2,255 inspection reports and comprising 9,430 data points, contained detailed attributes such as structure number, inspected elements, condition ratings, and inspector comments (which in turn contained detailed descriptions of the location, type, and intensity of concrete defects in different bridge deck elements). The data extraction process, illustrated in Figure 1, involved systematically retrieving and preparing the bridge inspection report data so that it could be transformed from unstructured textual data into a structured format for further analysis. The reports, typically stored across multiple sources, were automatically retrieved and prepared for data extraction. These reports then underwent automated data extraction using a rule-based approach integrating Docparser and Python via a REST API. This process involved dividing reports into manageable segments, annotating samples, creating extraction rules, and iteratively verifying and

refining the extracted data for accuracy. For further details on the data collection and extraction processes—along with examples of the unstructured inspection reports, the structure of condition data stored in these reports, and the structure of the data input/output as part of the standardization process—the interested reader may refer to the authors' previous study [12]. Data concerning defect severity thresholds (Table 1) was extracted from the aforementioned MTQ *Structures Inspection Manual* [24]. This manual categorizes severity levels from slight to severe based on standardized criteria, guiding the classification of defects. All visualizations and examples provided in this paper were translated from the original language of the collected data (i.e., French) to English.

To illustrate the developed method, the following inspection comment can be considered: "*Delamination and spalling, cracks less than 0.8 mm, moderate scaling, and rust traces and deposits*." According to Table 1, the following classifications can be made:

- *Corrosion*: rust traces are classified as severity level S_2.
- *Cracking*: cracking < 0.8 mm in width is classified as S_2.
- *Delamination*: delamination is classified as S_3.
- *Scaling*: moderate scaling is classified as S_2.
- *Spalling*: spalling is classified as S_3.

This example highlights the process of standardizing inspection comments by mapping the identified defects to their respective severity levels, ensuring systematic characterization for future utilization in condition assessment and predictive modeling, among other uses. However, applying this process to a large dataset poses significant challenges due to the variability in language styles and terminology. For example, the same given observation could be expressed differently (i.e., in different formats or phrasing) by different inspectors. As such, the models must effectively manage linguistic variability while ensuring consistent classification across the dataset.

The dataset, comprising 9,430 records, was divided as follows. A subset of 10% (947 records) was manually annotated and standardized to create the training/testing set, which was used to fine-tune the base model and evaluate model performance. This training/testing set was then further divided, with 30% (285 records) allocated to fine-tuning (focusing on refining the prompt engineering and optimizing the model's responses), and the remaining 70% (662 records) forming the testing subset for assessing the comparative performance of the models. The remaining of the dataset (90%: 8,483 records) was used as the generalization set to validate the best-performing model on a broader scale, ensuring its robustness, generalizability, and practical applicability in real-world scenarios.

3.2 Prompt Engineering

This study leveraged the power of GPT-3.5 for processing and standardizing unstructured bridge inspection data. It should be noted that this specific model was selected due to its cost-effectiveness and its having been recommended by OpenAI as the most suitable for fine-tuning, at the time of conducting the experiments in this study [25]. Prompt engineering, it should be noted, is a foundational step in guiding LLMs to produce accurate and contextually relevant responses. The design of prompts is essential when dealing with complex datasets such as bridge inspection comments. In this study, the prompt engineering process, illustrated in Figure 2(a) and Figure 2(d), integrated the following components:

- *Defect types*: defining defect categories such as corrosion, cracking, delamination, scaling, and spalling ensures that the model effectively identifies and categorizes the specific issues to be analyzed.
- *Severity levels*: standardizing defect severity levels into categories, i.e., slight, moderate, significant, and severe, enables consistent evaluation across inspection records.
- *Classification rules and thresholds*: establishing clear rules and thresholds provides a framework for consistent defect classification, ensuring the model adheres to predefined criteria.
- *Examples*: providing practical examples demonstrates how rules and thresholds are applied, enhancing the model's understanding and performance.
- *Goal definition*: specifying the purpose of the analysis (e.g., standardizing defect severity classification) makes the model's output focused on actionable insights.
- *Input and output formats*: clearly defining input and output structures ensures that the model correctly interprets raw data and generates structured, consistent outputs for downstream analysis.

These components were carefully combined to create the "*prompt base*," a foundational structure containing essential instructions and context. Inspection batch records were dynamically appended to the prompt base, creating a complete set of prompts to guide the model's analysis. This modular design, it should be noted, ensures flexibility and facilitates updates, as text files containing prompt components can be modified directly without requiring programming expertise. In this manner, the structured prompts streamline the workflow and maintain consistency.

3.3 Fine-tuning

Fine-tuning is a crucial process that customizes the LLM to enhance its performance for domain-specific tasks. In this study, the fine-tuning process was conducted using *GPT-3.5-turbo-0125*, a model optimized for such applications. The fine-tuning workflow, presented in Figure 2(b), involved the following steps:

- *Dataset preparation*: fine-tuning datasets were prepared with a training set (90%) and a validation set

(10%) to ensure effective learning and generalization; each dataset contained examples reflecting expected inputs and responses, formatted to align with the fine-tuning requirements and designed prompts.

- *Training configuration*: the model was trained (fine-tuned) using 1,399,893 tokens over three epochs, with hyperparameters such as a batch size of 1, a learning rate multiplier of 2, and a constant seed value to ensure reproducibility.
- *Performance metrics*: training and validation losses, along with training accuracy, were tracked to evaluate the model's learning and generalization capabilities.
- *Checkpoints*: intermediate checkpoints were created during training to monitor performance and prevent overfitting. These checkpoints enabled flexibility in evaluating the model at various stages of training.
- *Evaluation and iteration*: the model's outputs were evaluated against ground truth data.

Following prompt engineering and fine-tuning, the performance of both the baseline model GPT-3.5-turbo (BL-GPT) and the fine-tuned model (FT-GPT) was validated through an iterative testing process. Figure 2(c) illustrates the workflow, which involved dividing datasets into samples and batches, incorporating delays to comply with API rate limits, and analyzing performance against ground truth data. The standardized outputs were then evaluated, and the process was refined iteratively until the desired performance criteria were met.

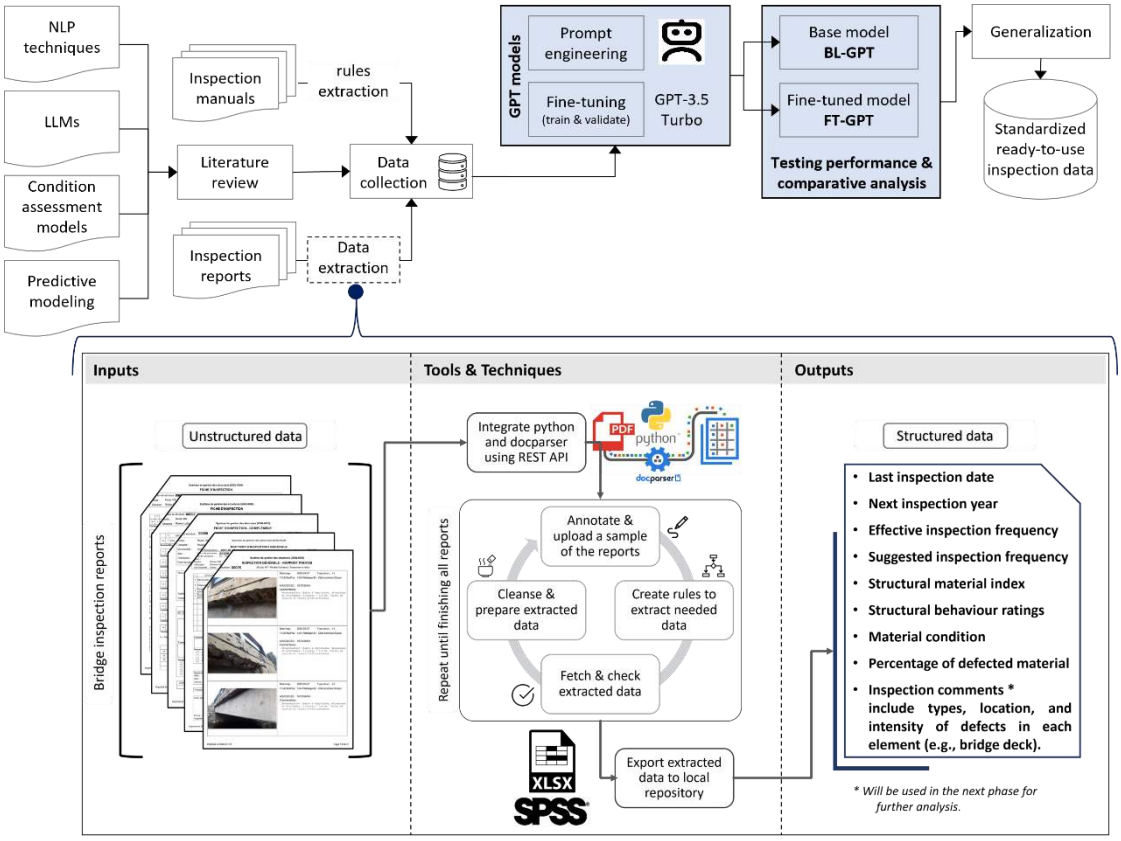


Figure 1. Overview of the developed method

Table 1. Thresholds for defect severity in bridge inspections [24]

	Severity level			
Defect	Slight (S_1)	Moderate (S_2)	Significant (S_3)	Severe (S_4)
Corrosion	Rust-free	Rusty concrete surface or reinforcement without cross-section reduction	Visible rusty reinforcement and reduced cross-section up to 30%	Very rusty reinforcement with greater than 30% reduction in cross-section
Cracking	—	< 0.8 mm	0.8–3.0 mm	> 3.0 mm
Delamination	—	—	Presence	—
Scaling	< 25 mm	25–50 mm	50–100 mm	> 100 mm
Spalling	—	—	Presence	—

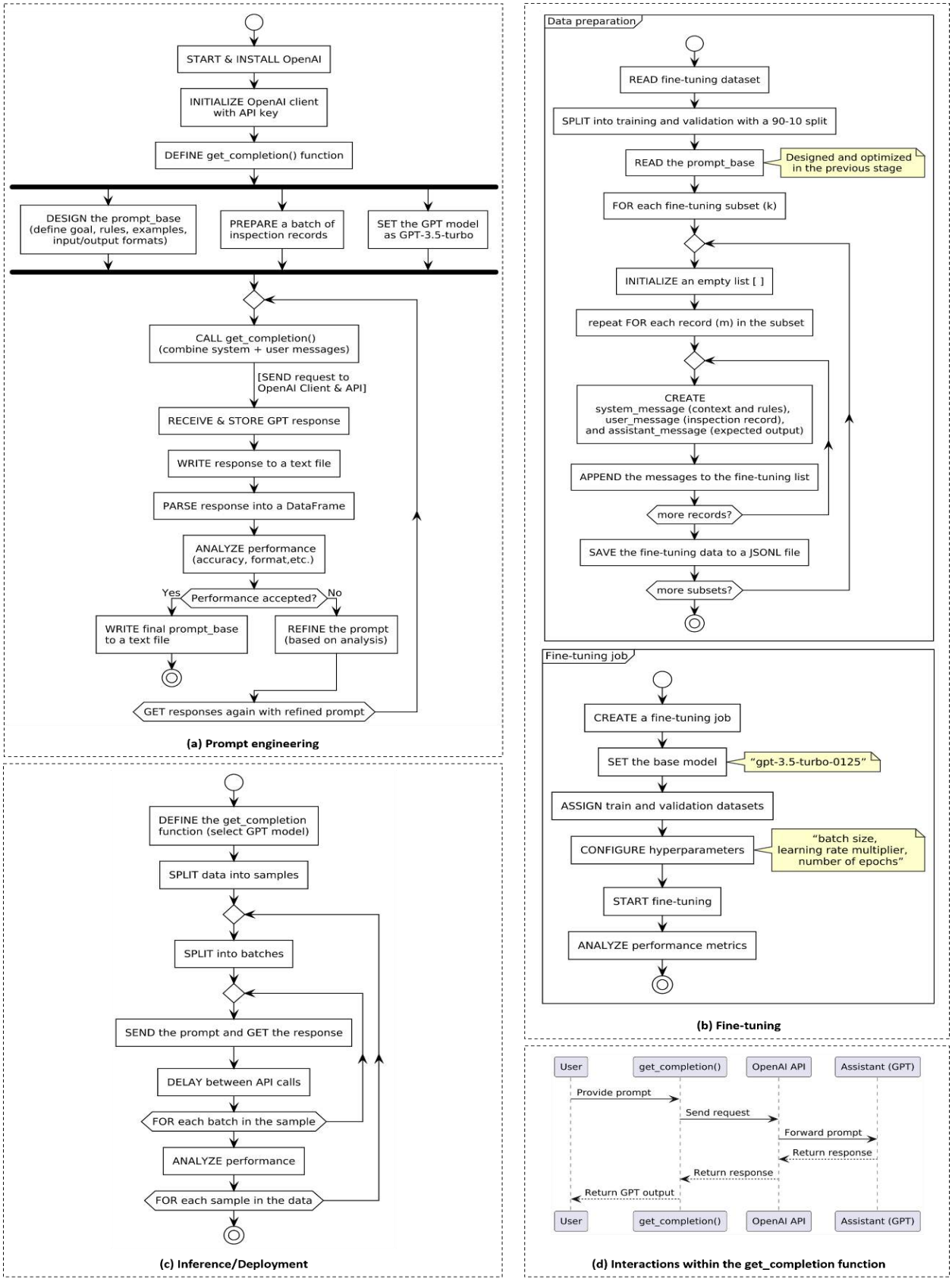


Figure 2. Workflow for GPT-based bridge data processing

4 Results and Discussion

The analysis of the 9,430 inspection records identified frequently occurring terms such as "cracking," "spalling," and "delamination." Figure 3 illustrates the structure of the input data to the BL-GPT and FT-GPT, along with the structure of the output data from the fine-tuned model. The BL-GPT and FT-GPT were tested on the 662-record test dataset, followed by a comprehensive comparative analysis of their performance in identifying and categorizing concrete bridge deck defects. The evaluation focused on accuracy, precision, recall, and F1 score across defect types, as represented by Eqs. (1) – (4), respectively, where TP, TN, FP, and FN are defined as True Positive, True Negative, False Positive, and False Negative, respectively. For the multi-class problem at hand, precision, recall, and F1 score were computed individually for each class and then aggregated using the weighted average method.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1\ score = \frac{2TP}{2TP + FP + FN} \quad (4)$$

As shown in Table 2, FT-GPT consistently outperformed BL-GPT in the detection and classification of all defect types, achieving exceptional performance and reliability. BL-GPT exhibited particularly poor performance in detecting and characterizing spalling, achieving only 61.63% accuracy and a 54.50% F1 score, substantially lower than FT-GPT's 98.64% in both metrics. The lower performance metrics of the base model is attributable to several factors:

- *Complexity of inspection comments*: Inspection comments often include multiple defect descriptions and measurements, introducing complexity and ambiguity that result in misinterpretation of context.
- *Insufficient training data and generalized knowledge*: The base model lacks exposure to a diverse and specialized dataset, limiting its ability to generalize effectively to complex inspection comments. Although it is pre-trained on extensive general-purpose data, its generalized knowledge is insufficient for the technical and domain-specific language used in bridge inspection reports. As such, without fine-tuning on domain-specific data, the model is unable to fully adapt to the terminology, phrasing, and contextual nuances unique to bridge inspection comments, resulting in less accurate and consistent standardization.
- *Stochastic nature*: Operating probabilistically with billions of parameters, the base model may produce varying outputs even with the same input, leading to inconsistencies in defect classification and severity assessment.
- *Language mismatch and translation nuances*: The process of translating between languages—in this study, the original data was in French, while the prompts and interactions with the model were designed in English—can result in the loss of subtle meanings and technical nuances.

The results of this study demonstrate that fine-tuning is as an efficient solution to overcome the limitations outlined above. By adapting the model to domain-specific data, fine-tuning enhances its ability to accurately interpret terminology, phrasing, and contextual details unique to bridge inspections. This process mitigates issues related to contextual misinterpretation, complexity, and insufficient domain knowledge. Furthermore, fine-tuning improves output reliability and stability, ensuring accurate defect classification and severity assessment. It also strengthens the model's capacity to handle multilingual inputs, maintaining the integrity of technical terms during translation and standardization. The fine-tuned model offers several advantages over conventional NLP techniques, particularly in handling complex and unstructured textual data. It reduces the need for extensive programming and manual feature crafting, allowing engineers to interact with the model and scale it to their specific tasks using their linguistic patterns. Through the use of a well-designed prompt, the model eliminates the need for complex coding expertise, making it accessible and adaptable to various practical applications.

The best-performing model, FT-GPT, was applied to the generalization dataset, which comprised 8,483 records (i.e., 90% of the overall dataset). This step involved evaluating the model's robustness, generalization capabilities, and practical applicability. The dataset was enriched (i.e., through generalization) to include standardized inspection narratives, the severities of different types of reinforced concrete bridge deck defects presented in a consistent, standardized format, and associated condition ratings for each element in these bridge decks. This data asset can be utilized in future studies on the asset management of bridges. As a proof of concept, the data has already been used in developing data-driven condition assessment models to evaluate the condition of bridge deck elements and overall deck condition based on the identified defect severities. (For further details, the interested reader may refer to the authors' previous work [12].)

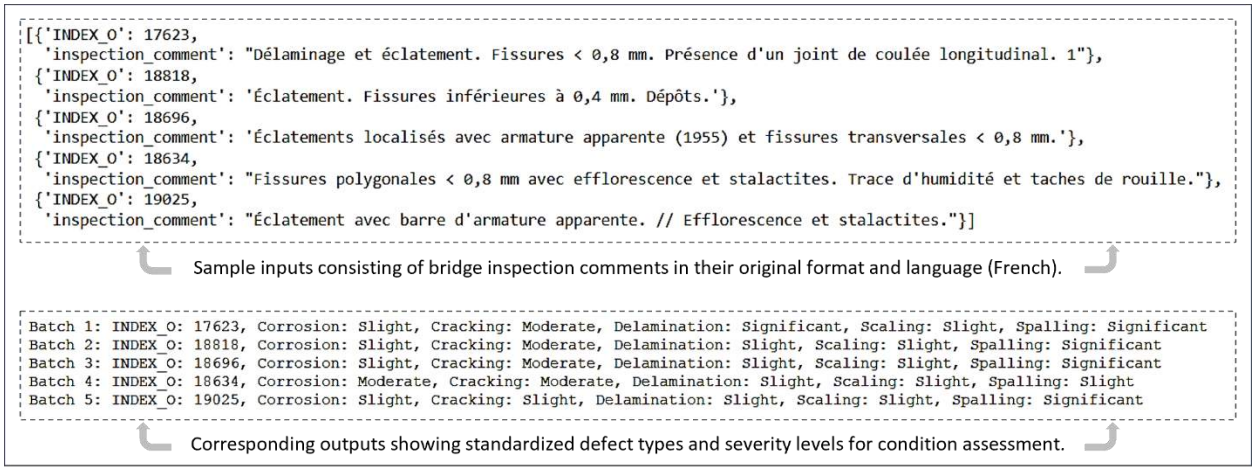


Figure 3. Structure of inputs and outputs of the fine-tuned model

Table 2. Performance comparison of BL-GPT and FT-GPT models in detecting and standardizing concrete defects

	Model	BL-GPT				FT-GPT			
	Performance	Accuracy	Precision	Recall	F1 score	Accuracy	Precision	Recall	F1 score
Defect	Corrosion	88.22%	88.43%	88.22%	87.44%	98.79%	98.99%	98.79%	98.84%
	Cracking	92.90%	94.15%	92.90%	93.35%	98.64%	98.68%	98.64%	98.58%
	Delamination	96.37%	97.41%	96.37%	96.88%	99.09%	99.10%	99.09%	99.09%
	Scaling	91.24%	88.53%	91.24%	88.37%	98.64%	98.78%	98.64%	98.64%
	Spalling	61.63%	78.09%	61.63%	54.50%	98.64%	98.64%	98.64%	98.64%

As of May, 2024, i.e., at the time of conducting the experiments for this study, the implementation cost and processing time for analyzing 100 batches of inspection comments were evaluated. The BL-GPT model incurred a total cost of ~$0.09, processing ~167,238 tokens, consisting of ~161,560 input tokens and ~5,678 output tokens, with an average processing time of approximately 7 min. In contrast, the fine-tuned GPT model had a higher total cost of ~$0.52, with ~3,449 output tokens and a slightly reduced processing time of approximately 6.5 min. This highlights that the fine-tuned model generates more concise responses while incurring a lower processing time. The fine-tuning process itself (with the number of tokens and parameters discussed in Section 3.3) incurred a total cost of ~$11.19.

5 Summary and Concluding Remarks

This study represents a significant step toward leveraging GenAI models for asset management and enhancing the utilization of bridge inspection data. The process focuses on detecting and categorizing defects in reinforced concrete bridge decks—including corrosion, cracking, delamination, scaling, and spalling—into defined severity levels. LLM-based GPT-3.5 was harnessed as a baseline model. This model was fine-tuned with domain-specific data to enhance its ability to process complex technical information from bridge inspection reports. A robust prompt engineering process was developed to effectively capture and standardize defect severity levels. The fine-tuned model demonstrated superior performance and was found to be reliable in standardizing inspection comments. The contributions of this study include (1) introducing a GPT-based method for improved information extraction from and standardizing bridge inspection data to support data-driven decision-making; (2) demonstrating the effective use of GenAI in infrastructure management; (3) fine-tuning LLMs to improve performance, stability, and reliability; (4) employing robust prompt engineering in asset management; and (5) providing a standardized dataset for utilization in future work. While this study focused on defects in concrete bridge decks, the methods are extendable to other infrastructure systems, paving the way for improved infrastructure management more broadly. It should be noted that further investigations are needed to evaluate the implementation costs associated with these models. Future research should also explore the performance of newer models, such as GPT-4o and o1, as well as open-source alternatives, such as Llama and Gemini.

References

[1] A. Omar and O. Moselhi. Condition monitoring of reinforced concrete bridge decks: Current practices and future perspectives. *Current Trends in Civil & Structural Engineering*, 8(4), 2022.

[2] K. Liu and N. El-Gohary. Ontology-based semi-supervised conditional random fields for automated information extraction from bridge inspection reports. *Automation in Construction*, 81:313–327, 2017.

[3] R. Li, T. Mo, J. Yang, D. Li, S. Jiang, and D. Wang. Bridge inspection named entity recognition via BERT and lexicon augmented machine reading comprehension neural model. *Advanced Engineering Informatics*, 50:101416, 2021.

[4] A. Omar and O. Moselhi. An integrated approach for automated acquisition of bridge data and deficiency evaluation, In *Proceedings of the 40th International Symposium on Automation and Robotics in Construction (ISARC)*, Vol. 40, IAARC, Chennai, India, 2023.

[5] X. Chen, H. Xie, and X. Tao. Vision, status, and research topics of Natural Language Processing. *Natural Language Processing Journal*, 1:100001, 2022.

[6] J. Just. Natural language processing for innovation search – Reviewing an emerging non-human innovation intermediary. *Technovation*, 129:102883, 2024.

[7] F. Xu, T. Nguyen, and J. Du. Augmented reality for maintenance tasks with ChatGPT for automated text-to-action. *Journal of Construction Engineering and Management*, 150(4):04024015, 2024.

[8] P. Ghimire, K. Kim, and M. Acharya. Opportunities and challenges of Generative AI in construction industry: Focusing on adoption of text-based models. *Buildings*, 14(1):220, 2024.

[9] K. Liu and N. El-Gohary. Semantic neural network ensemble for automated dependency relation extraction from bridge inspection reports. *Journal of Computing in Civil Engineering*, 35(4):04021007, 2021.

[10] T. Li, M. Alipour, and D. K. Harris. Context-aware sequence labeling for condition information extraction from historical bridge inspection reports. *Advanced Engineering Informatics*, 49:101333, 2021.

[11] S. Moon, S. Chung, and S. Chi. Bridge damage recognition from inspection reports using NER based on recurrent neural network with active learning. *Journal of Performance of Constructed Facilities*, 34(6):04020119, 2020.

[12] A. Omar and O. Moselhi. Automated data-driven condition assessment method for concrete bridges. *Automation in Construction*, 167:105706, 2024.

[13] C. Zhang, X. Lei, Y. Xia, and L. Sun. Automatic bridge inspection database construction through hybrid information extraction and large language models. *Developments in the Built Environment*, 20:100549, 2024.

[14] K. Liu and N. El-Gohary. Fusing data extracted from bridge inspection reports for enhanced data-driven bridge deterioration prediction: A hybrid data fusion method. *Journal of Computing in Civil Engineering*, 34(6):04020047, 2020.

[15] K. Liu and N. El-Gohary. Improved similarity assessment and spectral clustering for unsupervised linking of data extracted from bridge inspection reports. *Advanced Engineering Informatics*, 51:101496, 2022.

[16] S. Banerjee, C. M. Potts, A. H. Jhala, and E. J. Jaselskis. Developing a construction domain–specific artificial intelligence language model for NCDOT's CLEAR Program to promote organizational innovation and institutional knowledge. *Journal of Computing in Civil Engineering*, 37(3):04023007, 2023.

[17] T. Nguyen-Mau, A.-C. Le, D.-H. Pham, and V.-N. Huynh. An information fusion based approach to context-based fine-tuning of GPT models. *Information Fusion*, 104:102202, 2024.

[18] S. Moon, S. Chi, and S.-B. Im. Automated detection of contractual risk clauses from construction specifications using bidirectional encoder representations from transformers (BERT). *Automation in Construction*, 142:104465, 2022.

[19] S. Chung, S. Moon, J. Kim, J. Kim, S. Lim, and S. Chi. Comparing natural language processing (NLP) applications in construction and computer science using preferred reporting items for systematic reviews (PRISMA). *Automation in Construction*, 154:105020, 2023.

[20] OpenAi. OpenAI platform documentation: Models. On-line: https://platform.openai.com/docs/models, Accessed: 11 June 2024.

[21] OpenAi. GPT-4. On-line: https://openai.com/gpt-4/, Accessed: 11 June 2024.

[22] J. Zheng and M. Fischer. Dynamic prompt-based virtual assistant framework for BIM information search. *Automation in Construction*, 155:105067, 2023.

[23] Statistics Canada. Age of public infrastructure: a provincial perspective. On-line: https://www150.statcan.gc.ca/n1/pub/11-621-m/11-621-m2008067-eng.htm, Accessed: 8 March 2022.

[24] Ministry of Transport and Sustainable Mobility of Québec. *Structures Inspection Manual*. Government of Québec, National Library and Archives of Québec, 2017.

[25] OpenAi. Fine-tuning guide. On-line: https://platform.openai.com/docs/guides/fine-tuning, Accessed: 11 June 2024.

Practical Considerations and Case Study for Implementing Autonomous UAV Video Collection of Construction Activities

Yusheng Huang[1], Amin Hammad[1]*

[1]Concordia University, Canada
yusheng.huang@mail.concordia.ca; amin.hammad@concordia.ca (*corresponding author)

Abstract –

Various methods have been proposed in recent years to extract valuable insights from images and videos of construction activities to improve project management. Previous studies have proposed automated video collection using camera-equipped multirotor unmanned aerial vehicles (CE-MUAVs) to enable the collection of construction videos with the ability to fit the camera's field of view (FOV) automatically to the specific areas where the activities are happening. However, multiple practical considerations should be addressed before implementing autonomous UAV video collection of construction activities to fulfill the technical, safety, and data quality requirements, while respecting the legal regulations. This paper aims to identify these considerations and to develop a workflow for implementing an effective video collection of construction activities using autonomous CE-MUAVs. A test on an actual construction site is conducted to validate the applicability of the workflow.

Keywords –
Construction activity; Video Collection; Path planning; 4D Simulation; UAV.

1 Introduction

With the emergence of advanced computer vision (CV) techniques, it is possible to automatically extract rich semantics and high-level knowledge from digital images and videos, enabling more comprehensive and efficient analysis for different construction management applications [1], including activity monitoring [2], progress monitoring [3, 4], and health and safety monitoring [5–7]. However, the complexity and high costs associated with video collection have significantly constrained the widespread adoption of these methods. Collecting high-quality videos of specific areas of interest on construction sites, where activities happen, according to the project schedule remains challenging with conventional approaches, such as stationary cameras, handheld devices, equipment-mounted cameras, and wearable systems [1].

Over the past decade, considerable attention has been paid to automated video collection using camera-equipped multirotor unmanned aerial vehicles (CE-MUAV), which provides mobility to the camera to fit its field of view (FOV) automatically to the important parts of the construction site while ensuring high efficiency and reasonable cost [8–10]. Previous studies have proposed generating collision-free UAV paths based on a BIM model to support visual inspection tasks of buildings [11, 12]. Other studies proposed methods for sampling the viewpoints to achieve full visual coverage of the structure considering the sensor (i.e., camera or laser scanner) specifications, such as FOV, focal length, and sensor size [13, 14]. Then, routing algorithms are applied to generate the optimal path through the sampled viewpoints. In addition, Ibrahim et al. [15] proposed a method of UAV path planning for progress monitoring during the construction phase considering the schedule using waypoints. However, current implementations of CE-MUAVs primarily focus on scanning static objects in outdoor and indoor environments for purposes like 3D reconstruction and inspection. On the other hand, limited research is available about collecting videos of construction activities using autonomous UAVs.

To address this research gap, the authors previously proposed a method to perform simulation-based optimization of path planning for CE-MUAVs considering multiple construction activities according to the as-planned micro-schedule [16]. This method identifies the top-ranking viewpoints (VPs) to collect high-quality videos about the construction activities in the micro-schedule with good coverage of the workspace while minimizing the distance to the workspace. In addition, the method generates optimized paths that minimize the travel duration between the top-ranking VPs while avoiding collisions. While effective in a virtual environment, this approach has yet to be tested in real-world scenarios. However, multiple practical considerations should be addressed before implementing autonomous UAV video collection of construction activities to fulfill the technical, safety, and data quality requirements, while respecting the legal regulations. This paper aims to identify these considerations and to develop a workflow for implementing an effective video collection of construction activities using autonomous

CE-MUAVs. A test on an actual construction site is conducted to validate the applicability of the workflow.

2 Simulation-Based Optimization of Path Planning for CE-MUAV

Figure 1 shows the method of simulation-based optimization of path planning for CE-MUAVs proposed in the previous work of the authors [16]. This method identifies the top-ranking viewpoints (VPs) to collect high-quality videos of construction activities based on the micro-schedule with good coverage of the workspaces, while minimizing the distance to these workspaces. In addition, the method generates an optimized path that enables the UAV to collect videos based on the identified VPs while minimizing travel time and avoiding collisions. It has been demonstrated that aerial videos captured while the UAV is hovering offer distinct advantages, as they enable the capture of finer scene details and intricacies [17]. It is assumed that the CE-MUAV will collect videos during traveling and hovering states. However, the focus is on spending more time to collect videos while hovering. There are three main modules in the simulation-based path planning: (1) Preparation of the simulation environment; (2) VPs optimization; and (3) Path planning for the UAV to minimize the travel duration.

In the *simulation environment preparation module*, the workspaces for a certain period are identified from the as-planned 4D simulation of the project. Then, the solution variables of the UAV are identified according to the selected model of the UAV. The search spaces for different workspaces are then defined considering the safety risks in the construction site, the CV requirements (e.g., range of camera, tilt angle), and the compliance with legal regulations, which will be further explained in Section 3. At last, cubic cells are generated within the workspaces for calculating the camera coverage.

In the *VPs optimization module*, the top-ranking VPs for each workspace, where the FOV of the hovering camera can provide good coverage while minimizing the distance, are identified by integrating a multi-objective genetic algorithm (MOGA) with simulation. Two objective functions are defined: maximizing coverage and minimizing the camera-to-workspace distance. For each workspace, multiple top-ranking VPs will be identified. With these VPs, the *path optimization module* solves the generalized traveling salesman problem (GTSP) to optimize the cycle path with the least travel duration to collect videos from one of the top-ranking VPs for each workspace. It is worth noting that in this paper, the Theta* algorithm is utilized to generate a more efficient collision-free path to travel between the VPs compared with the A* algorithm used in our previous paper [16]. This optimized cycle path consists of multiple waypoints, each specifying the location and pan angle of the UAV and the tilt angle of the camera. At last, the hovering duration at each VP of the optimized cycle path is defined according to the importance level of the task in different workspaces.

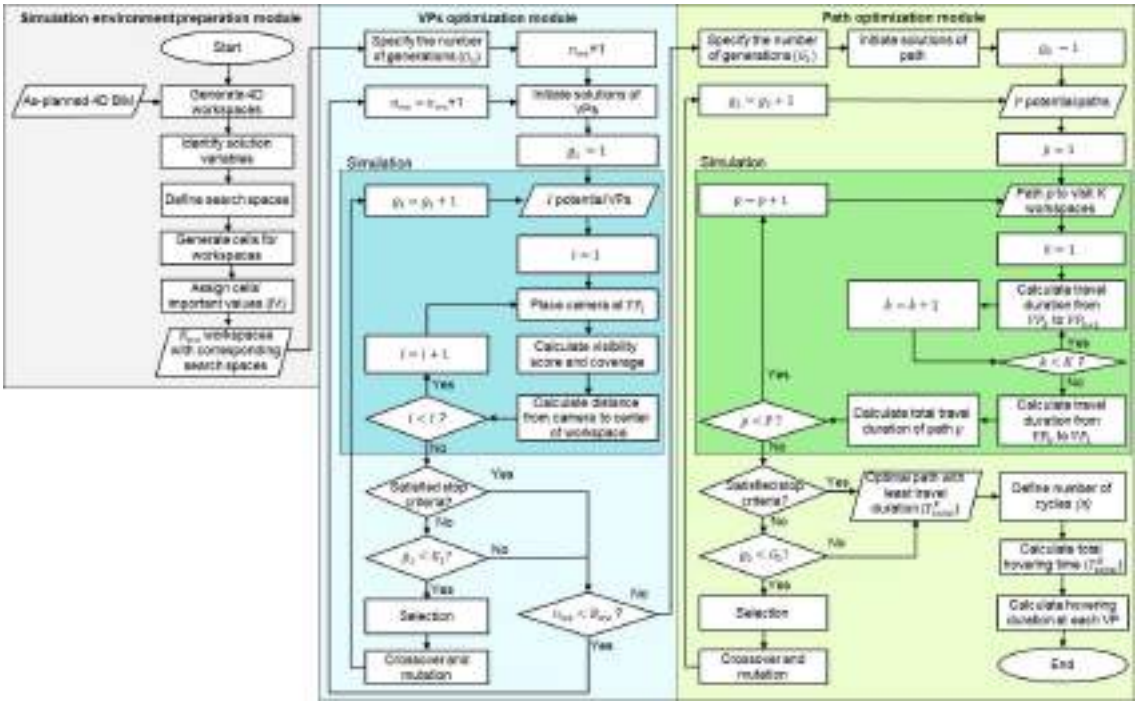


Figure 1. Simulation-Based Optimization of Path Planning for CE-MUAV

3 Practical Considerations for Implementation

Multiple practical considerations must be addressed before implementing autonomous UAV video collection of outdoor construction activities in actual projects using the optimized path generated by the proposed method in Section 2.

3.1 Technical requirements

To ensure effective implementation, several technical requirements must be addressed. Here, the technical requirements can be considered in two aspects, the hardware requirements and software requirements. The hardware requirements mainly refer to the requirements in selecting an appropriate UAV that is capable of performing the task. First, the UAV should support pre-planned flight operations and offer an application programming interface (API) to enable importing the optimized flight path from the simulation console. The UAV should also have a reliable antenna for Global Navigation Satellite System (GNSS) or alternative systems to maintain accurate flight paths. The operational range and battery life must be sufficient to complete the video collection tasks. In addition, the UAV should be equipped with a camera capable of collecting high-quality video to meet the data requirements for CV, which are explained in Section 3.4. Moreover, sufficient onboard storage for videos is essential. Automated data transmission using 5G communication to a remote console or database is highly preferable, as it enhances efficiency by eliminating manual data retrieval.

On the other hand, appropriate software tools should be selected. For example, to develop a detailed construction 4D simulation model, Naviswork [18] was used in the author's previous work [16]. This model was then imported into Unity 3D game engine [19], which has a realistic physics engine for simulating the UAV operations. In addition, converting waypoints from simulation coordinates into UAV-readable data is necessary. The implementation process involves three coordinate systems: (a) coordinate system of the construction 4D simulation, (b) coordinate system of the simulation for path planning, and (c) coordinate system for UAV navigation. Typically, models are developed in a local coordinate system specific to the software tools [20]. To incorporate the geospatial context of the coordinate system of (a) and (b), a geographic information system (GIS) can be used to show the positions in a projected coordinate system (PCS) [21]. However, UAV navigation uses a geographic coordinate system (GCS). For example, DJI UAV navigation [22] uses a geodetic coordinate system. Therefore, to navigate the UAV, it is necessary to have a tool to translate the coordinates of the waypoints of the optimized path from the PCS to GCS.

3.2 Safety requirements

The integration of UAVs in construction may encounter several safety risks [23–25], highlighting the need for careful operation planning and UAV selection to ensure safe and effective operations. The major risk is collisions with workers or objects on construction sites. During the planning phase, risk assessment should be performed to identify all potential collision risks based on the site layout. In the *simulation environment preparation module*, safety buffers should be assigned to workspaces and high-risk areas. In addition, selecting UAV models equipped with active collision avoidance can further mitigate these risks [29]. UAVs may also encounter motor malfunctions or communication loss, necessitating a detailed emergency contingency plan to address such scenarios before deployment. Moreover, the selection of a UAV model should consider risk mitigation features, such as pre-flight fault detection, automatic return-to-home functionality in case of loss of connection, and a reliable and secure communication system to prevent unauthorized access or interference to the UAV. Furthermore, UAVs may pose safety risks to workers on-site due to distractions [24]. Selecting light-weight UAV models that generate less noise can help reduce this issue.

3.3 Compliance with legal regulations

UAV deployment must comply with the legal requirements. The UAV should be properly registered in accordance with local laws and regulations. The operators must undergo training and obtain the necessary licenses or certifications before conducting operations. If the operation takes place in controlled airspace, the necessary certificates and permits must be obtained as required by regulations. In addition, detailed documentation of the operational risk assessment, precautions for risk mitigation, and an emergency contingency plan may be required by the responsible department (e.g., Transport Canada) when applying for special flight operations [26].

In *simulation environment preparation module*, zones where the project team does not have permission to operate the UAV, such as areas near surrounding buildings, should be clearly defined to prevent unauthorized intrusion. The compliance of such regulations should be verified through virtual verification after generating the optimized path and confirmed again during on-site verification before deployment.

Moreover, the selection of UAV model should take into account the ease of obtaining operational permissions. For instance, basic operations that do not require a special flight operations certificate apply to UAVs weighing less than 25 kg in Canada [27]. It is

worth noting that exceptions may apply to micro-UAVs weighing under 250 g in some countries [26], which may reduce the complex process of obtaining permission to fly. Operating a micro-UAV does not require a pilot license or certificate if the operator ensures safe operations with good practices and compliance with all applicable requirements and regulations.

Furthermore, during the operation, the UAV use must respect the privacy of workers and other individuals captured in the video. The project team should obtain consent from workers before deployment. In addition, the team is responsible for developing secure data communication and storage methods to protect the privacy of the collected data.

3.4 CV requirements

The video collection process must meet specific CV requirements to ensure the captured frames are suitable. A proper tilt angle should be maintained to clearly show the pose of workers and equipment. In addition, CV methods (either object detection or activity recognition) are usually susceptible to aerial imagery with the small size of objects and variable tilt angle of view [17, 28, 29]. Therefore, the distance from the camera to the objects should be less than a certain maximum value to avoid the low resolution of objects in the frame while maintaining the coverage of the designated workspace. This CV requirement is fulfilled through the *VPs optimization module*, as explained in Section 2. In addition, a stable FOV with minor vibration and motion blur improves the quality of CV results. It has been suggested that aerial videos captured while the UAV is hovering offer distinct advantages, as they enable the capture of finer scene details and intricacies [17]. Moreover, hovering time should be sufficient for gathering essential data from the optimal VPs.

Table 1 lists the considerations of the requirements in different processes during the implementation. It should be mentioned that these requirements are often interconnected.

4 Implementation Workflow

The proposed method for implementing optimized video collection of construction activities using CE-MUAVs based on optimized path planning is shown in Figure 2. The implementation can be divided into three phases under three different coordinate systems as mentioned in Section 3.1. In the model preparation phase, a high LOD as-planned 4D simulation is developed. Then, the as-planned 4D model is integrated with the GIS data to show the geospatial context of the project. By reviewing the 4D simulation, risk assessment is performed and requirements are identified according to the key considerations as mentioned in Section 3. Based on these requirements, the type of UAV is selected. Then, the 4D simulation model is transferred to the simulation platform for UAV path planning.

Table 1. Practical considerations during implementation

Implementation step	Technical requirements	Safety requirements	Legal regulations compliance	CV requirements
UAV model selection	-UAV and sensor specifications -Pre-planned flight function -Reliable antenna and navigation system	- Robust accident prevention functionalities - Malfunctions handling - Light-weight UAV to avoid distractions - Secure communication method	- Weight of the UAV	- UAV performance to avoid vibrations and motion blur - Camera specifications
Simulation modelling and path optimization	-Simulation tool -Tool to translate coordinates of waypoints -Operational range and battery life	-Safety buffer to avoid collision risks	-Unauthorized flight zones	-Defining reasonable search space
Pre-flight planning		-Risk assessment -Contingency plan -Precautionary measures	-Operation certificate and flight permit -Privacy protection	

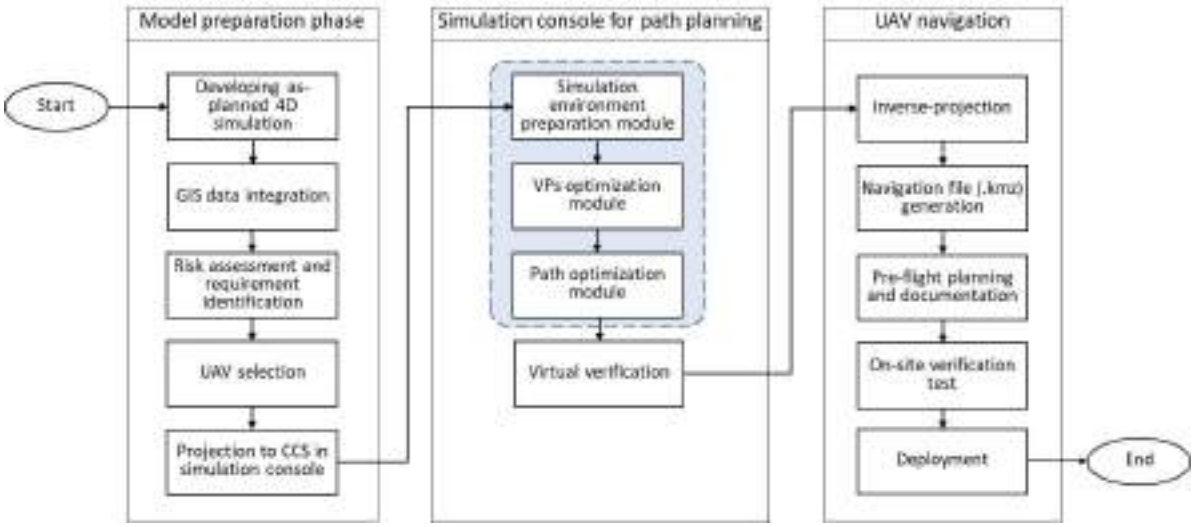


Figure 2. Implementation workflow of Autonomous UAV Video collection based on optimized path planning

Simulation-based path planning of CE-MUAVs is performed to find the optimized path for automated and effective data collection of construction activities using 4D as-planned simulation as explained in Section 2. In the *simulation environment preparation module*, the search spaces, safety buffer, and unauthorized flight zones should be defined considering the safety requirements, compliance with legal regulations, and the CV requirements as mentioned in Section 3. At the end of this phase, virtual verification is conducted to verify the efficiency and effectiveness of video collection along the optimized path. This is achieved by running a simulation in which the virtual UAV follows the same path within the virtual environment. The verification process involves reviewing the UAV's trajectory to eliminate the safety risks and analyzing the camera's FOV during hovering to ensure comprehensive coverage of the workspace.

To navigate the UAV, the waypoints of the optimized path in the simulation system are inverse-projected from PCS to the GCS as mentioned in Section 3.1. Using the waypoints in the GCS, a navigation file for the UAV is generated. One commonly used format for the navigation file is the .kmz format. Before deploying the UAV with the navigation file, pre-flight planning is carried out to outline precautionary measures and emergency contingency procedures. Once these plans are documented, an on-site verification test is conducted to ensure the UAV follows the optimized path and hovers VPs to collect videos of the expected workspaces of tasks. The on-site test should also verify that the operation is free of safety risks. In addition, it should ensure compliance with legal regulations during the deployment (e.g., preventing the collection of videos that infringe on workers' privacy). To minimize disruptions, this test should be conducted during off-peak hours, with the UAV moving slowly to allow immediate stops if any risks arise. Once verification is complete, the operation plan with the optimized path will be deployed to collect videos.

5 Case study

A field test was conducted on a construction site to validate the applicability and effectiveness of the proposed workflow. The test focused on activities carried out during a short period of a park construction project, which includes an outdoor skating rink and a pavilion. Figure 3 shows the workflow implementation in this case study.

As shown in the bird-view of the GIS integrated site layout (Figure 3(a)), the construction site is divided into two parts by a road. The left part is the main construction area, where two tasks were underway: the workspace for pouring concrete for the pavilion (WS3) and excavating and leveling the area designated for the skating rink (WS2). The right part contains a loading area used for storing the excavated soil (WS1). As shown in Figure 3(a), this construction site is surrounded by three buildings, each with a height of more than 25 meters, resulting in potential collision risks. Therefore, safety buffers are assigned to these buildings. Additionally, a safety buffer was maintained along the road, creating a danger zone with a height of 10 meters.

5.1 CE-MUAV selection and variable identification

The UAV used in this case study is the DJI Mini 4

Pro [30], which features a maximum speed of 12 m/s and a maximum ascent/descent speed of 5 m/s in "Normal Mode". Weighing 249 g with its standard battery, this UAV simplifies regulatory compliance in most North American jurisdictions as mentioned in Section 3.3, making it ideal for construction projects. Its compact and lightweight design minimizes safety risks while maintaining excellent maneuverability. Furthermore, its omnidirectional obstacle sensing system significantly reduces the risk of collisions, enhancing operational safety. The maximum flight time is 30 minutes with the standard battery.

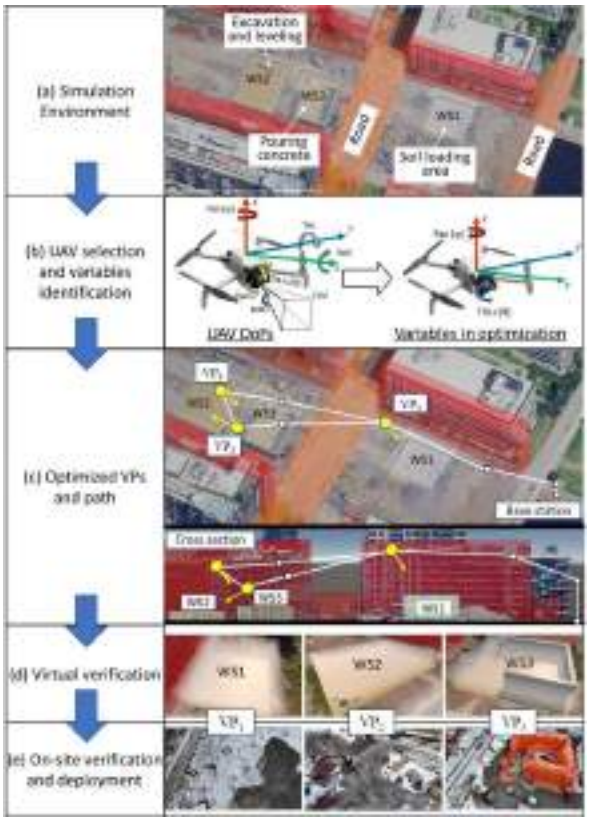


Figure 3. Workflow implementation in the case study

The UAV supports pre-planned flight operations under GNSS and provides an API [22] for importing mission files from a computer. Its camera captures 4K videos in MP4 format at up to 100 fps and offers a diagonal field of view (FOV) of 82.1° and a horizontal FOV of 74.3°, ensuring high-quality video collection for construction monitoring.

The variables for the viewpoint (VP) optimization module are defined based on this UAV model. As illustrated in Figure 3(b), the UAV provides ten degrees of freedom (DoFs). Six of these are associated with the UAV's pose (x, y, z, pan, tilt, and roll), while the camera, mounted on a 3-axis mechanical gimbal, introduces four additional DoFs (pan-c, tilt-c, roll-c, and FOV). However, the API does not allow the control of the tilt and roll angles of the UAV, as well as the pan and roll angles of the camera. In addition, the camera's FOV is fixed at its default value. Thus, the variables considered in the optimization process are the UAV's position (x, y, z), its pan angle (φ), and the tilt angle of the camera (θ).

5.2 Prototype development

A prototype system was developed using Unity3D game engine [19] to create a simulation platform for VPs optimization and path planning. As mentioned in Section 3.1, a method for translating the waypoints of the optimized path in the local coordinate of the path simulation platform to the UAV-readable waypoints under GCS is required. The 3D geospatial platform add-in for Unity named Cesium [31] is used for this purpose. The prototype system is run on a computer with i7-6800k CPU, 64 GB of random access memory (RAM), and two GTX 1070 GPUs.

5.3 Results

The results from the simulation-based path planning platform are multiple top-ranking VPs identified for each workspace. Using these top-ranking VPs, the path optimization module generates an optimized flight path as shown in Figure 3(c). The time required to travel between different waypoints is calculated. In this case study, the expected number of cycles per operation is 30, and the operation safety factor α for calculating the total operation time is set to 0.9. The traveling time between VPs (T^F_{total}), the total hovering time(T^H_{total}), and the total duration (T_{CP}) of one cycle path are calculated as shown in Table 2, using the method outlined in [16].

Table 2 Path durations with 30 expected cycles

Operation	Duration (s)
Base station → VP_1	9.55
Total cycle path (T_{CP})	53.36
- Total flight time (T^F_{total})	13.22
- VP_1→ VP_2	5.64
- VP_2→ VP_3	1.70
- VP_3→ VP_1	5.88
- Total hovering time (T^H_{total})	40.14
- Hovering at VP_1	10.04
- Hovering at VP_2	10.03
- Hovering at VP_3	20.07
VP_1→ Base station	9.55

Due to the higher level of interest associated with the task of pouring concrete, a higher importance level weight (W_{WP}) of 2 is assigned, while a weight of 1 is given to other workspaces. The distributions of the hovering durations for each VP are calculated as shown

in Table 2. Virtual verification is performed as shown in Figure 3(c) to assess the efficiency and effectiveness of video collection along the optimized path. Figure 3(d) illustrates the simulated camera views when hovering at three VPs along the optimized path.

Finally, an on-site test was conducted to validate the optimized path, as shown in Figure 3(e). The results demonstrate that the UAV successfully navigated to the designated hovering waypoints and adjusted its FOV to the intended angles, capturing videos of the areas of interest where construction activities were happening according to the schedule.

6 Conclusions and Future Work

This paper highlights the key practical considerations for implementing autonomous UAV-based video collection of outdoor construction activities. It also proposes a workflow for achieving optimized and efficient video collection using autonomous UAVs. A case study on a construction site is conducted for validation. The results of the case study demonstrate that the workflow is effective in generating an operational plan with optimized waypoints, enabling the UAV to capture videos of construction activities from areas where the main tasks are taking place.

Despite its effectiveness, several areas of the proposed workflow require further refinement. The proposed workflow focuses on collecting videos of progress-related construction activities within predefined workspaces according to the as-planned schedule. However, the location of the workspaces may change, but these changes are not reflected in the schedule. In addition, other construction activities may occur outside the predefined workspaces. Future studies will develop methods for dynamically adjusting the UAV path during the operation whenever workspace locations change or new workspaces are identified. In addition, the case study presented in this research is based on a medium-scale construction site. Given the diverse and complex nature of construction sites, further testing is needed across various scenarios to evaluate the workflow's robustness, such as indoor environments.

References

[1] Paneru, S. and Jeelani, I., "Computer vision applications in construction: Current state, opportunities & challenges," *Autom. Constr.*, 2021, vol. 132.

[2] Chen, C., Zhu, Z., and Hammad, A., "Automated excavators activity recognition and productivity analysis from construction site surveillance videos," Autom. Constr., 2020, vol. 110.

[3] Wang, Z., Zhang, Q., Yang, B., et al., "Vision-Based Framework for Automatic Progress Monitoring of Precast Walls by Using Surveillance Videos during the Construction Phase," J. Comput. Civ. Eng., 2021, vol. 35, no. 1.

[4] Arif, F. and Khan, W. A., "Smart Progress Monitoring Framework for Building Construction Elements Using Videography–MATLAB–BIM Integration," Int. J. Civ. Eng., 2021, vol. 19, no. 6, pp. 717–732.

[5] Tran, S. V.-T., Nguyen, T. L., Chi, H.-L., et al., "Generative planning for construction safety surveillance camera installation in 4D BIM environment," Autom. Constr., 2022, vol. 134.

[6] Han, S. and Lee, S., "A vision-based motion capture and recognition framework for behavior-based safety management," Autom. Constr., 2013, vol. 35, pp. 131–141.

[7] Soltani, M. M., Zhu, Z., and Hammad, A., "Framework for Location Data Fusion and Pose Estimation of Excavators Using Stereo Vision," J. Comput. Civ. Eng., 2018, vol. 32, no. 6.

[8] Molina, A. A., Huang, Y., and Jiang, Y., "A Review of Unmanned Aerial Vehicle Applications in Construction Management: 2016–2021," Standards, 2023, vol. 3, no. 2.

[9] Guan, S., Zhu, Z., and Wang, G., "A Review on UAV-Based Remote Sensing Technologies for Construction and Civil Applications," Drones, 2022, vol. 6, no. 5.

[10] Li, Y. and Liu, C., "Applications of multirotor drone technologies in construction management," Int. J. Constr. Manag., 2019, vol. 19, no. 5, pp. 401–412.

[11] Freimuth, H. and König, M., "Planning and executing construction inspections with unmanned aerial vehicles," Autom. Constr., 2018, vol. 96, pp. 540–553.

[12] Li, F., Zlatanova, S., Koopman, M., et al., "Universal path planning for an indoor drone," Autom. Constr., 2018, vol. 95, pp. 275–283.

[13] Bircher, A., Kamel, M., Alexis, K., et al., "Three-dimensional coverage path planning via viewpoint resampling and tour optimization for aerial robots," Auton. Robots, 2016, vol. 40, no. 6, pp. 1059–1078.

[14] Phung, M. D., Quach, C. H., Dinh, T. H., et al., "Enhanced discrete particle swarm optimization path planning for UAV vision-based surface inspection," Autom. Constr., 2017, vol. 81, pp. 25–33.

[15] Ibrahim, A., Golparvar-Fard, M., and El-Rayes, K., "Multiobjective Optimization of Reality Capture Plans for Computer Vision–Driven Construction Monitoring with Camera-Equipped UAVs," J. Comput. Civ. Eng., Sep. 2022, vol. 36, no. 5.

[16] Huang, Y. and Hammad, A., "Simulation-Based Optimization of Path Planning for Camera-Equipped UAVs That Considers the Location and

Time of Construction Activities," Remote Sens., 2024, 16(13).
[17] Othman, N. A. and Aydin, I., "Challenges and Limitations in Human Action Recognition on Unmanned Aerial Vehicles: A Comprehensive Survey," Trait. Signal, 2021, 38(5):1403–1411.
[18] "Navisworks | Autodesk," https://www.autodesk.com/ca-en/products/navisworks/overview.
[19] "Unity 2021.3," https://docs.unity3d.com/Manual/index.html.
[20] Diakite, A. A. and Zlatanova, S., "Automatic geo-referencing of BIM in GIS environments using building footprints," Comput. Environ. Urban Syst., Mar. 2020, vol. 80, p. 101453.
[21] Jaud, Š., Donaubauer, A., Heunecke, O., et al., "Georeferencing in the context of building information modelling," Autom. Constr., 2020, 118.
[22] "Cloud API," https://developer.dji.com/doc/cloud-api-tutorial/en/api-reference/dji-wpml/waylines-wpml.html.
[23] Kakaletsis, E., Symeonidis, C., Tzelepi, M., et al., "Computer Vision for Autonomous UAV Flight Safety: An Overview and a Vision-based Safe Landing Pipeline Example," ACM Comput Surv, Oct. 2021, vol. 54, no. 9, p. 181:1-181:37.
[24] Jeelani, I. and Gheisari, M., "Safety challenges of UAV integration in construction: Conceptual analysis and future research roadmap," Saf. Sci., 2021, vol. 144.
[25] Balestrieri, E., Daponte, P., De Vito, L., et al., "Sensors and Measurements for UAV Safety: An Overview," Sensors, 2021, vol. 21, no. 24.
[26] Canada, T., "Find your category of drone operation," AARV 13921205, Feb. 19, 2021, https://tc.canada.ca/en/aviation/drone-safety/.
[27] Canada, T., "Drone safety," AEAD 14507934, 2023, https://tc.canada.ca/en/aviation/drone-safety.
[28] Peng, H. and Razi, A., "Fully Autonomous UAV-Based Action Recognition System Using Aerial Imagery," in Advances in Visual Computing, Cham, 2020, pp. 276–290.
[29] Aldahoul, N., Karim, H. A., Sabri, A. Q. Md., et al., "A Comparison Between Various Human Detectors and CNN-Based Feature Extractors for Human Activity Recognition via Aerial Captured Video Sequences," IEEE Access, 2022, 10: 3532–63553.
[30] "DJI Mini 4 Pro - Specs - DJI," DJI Official, https://www.dji.com/ca/mini-4-pro/specs.
[31] "Cesium for Unity," Cesium, Oct. 08, 2024, https://cesium.com/learn/unity/.

Drywall finishing with collaborative robot arm in off-site construction

Philippe D'Amours, Samuel Faucher, François Ferland and Alexandre Girard

Interdisciplinary Institute for Technological Innovation (3IT), University of Sherbrooke, Quebec, Canada

philippe.damours@usherbrooke.ca, samuel.faucher2@usherbrooke.ca,
francois.ferland@usherbrooke.ca,alexandre.girard2@usherbrooke.ca

Abstract -

The automation of tasks in construction industries is a solution for countries experiencing a labour shortage in those industries. The task of drywall finishing in an off-site construction environment is particularly interesting for automatization because of the large amount of labour, the strain on the worker's body, the falling risk and the large quantity of dust particles. This article explores assisting workers with drywall finishing using an arm-mounted orbital sander with a collaborative robot arm placed on a mobile platform, combined with a drywall joint compound machine vision detection algorithm. Preliminary results suggest the robot can finish the same depth of material as a human worker, which is typically less than 0.19 mm.

Keywords -

Drywall Finishing; Machine Vision; Construction Robot; Collaborative robot

1 Introduction

Gypsum board, also known as gypsum drywall or plaster board, offers a cost-effective and popular [1] option to build the interior wall of modern-day construction in Canada. Unfortunately, while gypsum boards are easy to install, the seams and screws need to be covered by multiple layers of compound, also known as mud or plaster, to form a uniformly flat surface ready for paint. While drying, this compound shrinks and leave an uneven surface full of bumps and cracks. To finish the drywall, those defects need to be sanded down before applying paint. The general surface is also sanded down to make seams between boards less visible and more aesthetically pleasing.

Unfortunately, drywall finishing is quite laborious and inconvenient for workers [2]. To finish the compound, workers use power tools to remove most of the uneven cracks and bumps. Subsequently, workers use handheld tools like sanding blocks to smooth out the finer details. Sanding the compound produces a lot of irritable dust particles [2], especially when using power tools. To reach elevated walls, workers work with their arms over the shoulders for long hours, which can lead to pain and injuries [3].

Assisting workers with this task represents an opportunity for automation. Fortunately, off-site construction offers a stable environment for automatization and contributes to the reproducibility of the presented results. This paper presents the design of a sanding mobile robot capable of performing the power-assisted part of the drywall finishing task. Our key contributions are : 1) The integration of a 7-DOF collaborative manipulator arm-mounted orbital sander with an autonomous mobile platform; 2) An automatic computer vision pipeline capable of detecting and localizing sections of drywall to sand; and 3) A control scheme to perform the first pass of compound sanding on vertical walls.

This paper is structured as follows. Section 1 presents an overview of the requirements and chosen hardware for the robot. Section 2 presents the machine vision algorithm used to localize the compound joint. Section 3 presents the control strategies used to finish the drywall. Section 4 presents our measurement methodology and discusses the results.

2 Sanding robot hardware

The robot hardware has been designed in accordance with our partners' requirements, which can be summarized as follows: 1) The ability to navigate the construction plant and overcome ground obstacles such as cable covers and ground conveyor chain tracks; 2) A vertical sanding reach of 10 feet to cover the majority of their wall production output; and 3) An orbital sander compatible with vacuum dust control. Figure 1 presents the overall robot and Table 1 the major hardware components, which are described below.

The mobile base is a Clearpath's Husky[1] with the high-torque motorization option for high payload. This mobile platform let us move the robot using a wireless remote control in the off-site construction environment.

For the mobile arm, a collaborative robot is necessary to limit interaction forces, because mobile robots are not constrained to a traditionally fenced robot environment.

[1] https://clearpathrobotics.com/husky-a300-unmanned-ground-vehicle-robot/

The Fanuc CRX-20iA/L[2] has a reach of 1418 mm and a maximum payload of 20kg to manipulate the orbital sander. This arm also features internal force sensors that enables force control. An additional robot axis motor is necessary to reach 10' high walls, and an Igus linear axis of 1050 mm[3] was specified for this task.

The orbital sander mounted on the manipulator arm is a Festool PLANEX LHS-E 225 EQ[4]. The head of the orbital sander was adapted to the robot effector with a custom 3D printed part. Festool P320 ready-to-use abrasive pads for the 225 mm orbital sander are used and is the least abrasive option offered by Festool for this orbital sander.

For the vision system, the RGB-D (color + depth) Intel D435i[5] camera is common and easy to use for the design of the first prototype. After comparing multiple segmentation approaches based on other sensors, such as a Time-of-Flight camera, the combination of near-UV lights with the RGB-D camera was found to be the most efficient in maximizing the visual differences between the compound and the board's paper layer [4].

Commercially-available solutions, such as the CANVAS 1200CX[6], use similar hardware configuration. However, they usually used a sprayed compound, and our study focuses on performance with manually applied compound.

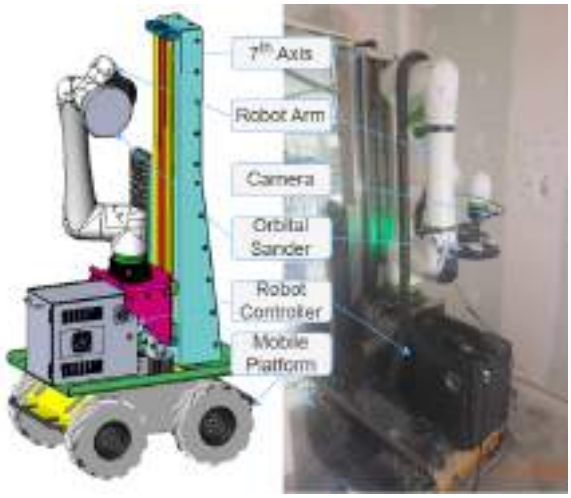


Figure 1. Left: CAD model. Right: Robot

[2]https://crx.fanuc.eu/wp-content/uploads/2022/06/MDS-04849-FR.20220329130728332.pdf
[3]https://www.igus.eu/product/20917?artNr=SLW-25120-xxx
[4]https://www.festoolcanada.com/produits/sableuses-lectriques/sableuses-autoportes--rallonges/571935---planex-lhs-e-225-eq-usa
[5]https://www.intelrealsense.com/depth-camera-d435i/
[6]https://www.canvas.build/products/canvas1200cx/

Table 1. Hardware Overview

Robot Arm	CRX-20iA/L
Robot Controller	R-30iB Mini Plus
Additional Axis Motor	βiS 4/4000-B
Linear axis	Igus SLW-25120
Camera	Intel RealSense D435i
UV Light	Everbeam 100W IP66
Orbital Sander	PLANEX LHS-E 225 EQ
Sander's abrasive	Festool GRANAT P320
Mobile Platform	Clearpath Husky

3 Perception pipeline

The perception pipeline is used to find the 3D position coordinate of the compound surface to finish in the 3D frame of the camera. The combination of a stereo depth camera with an RGB component creates a point cloud in the camera's frame with corresponding colour values for each point.

Every point in the cloud is segmented depending on whether they originate from the paper layer of the gypsum board or from the compound covering the seam of the boards and the screws head. First, near-ultraviolet spotlight helps differentiate the board and the compound because the fine layer of paper on the boards is fluorescent and emits blue light when exposed to near-UV light [4]. While the joint compound simply reflects the near UV light [4] as shown in figure 2.

To maximize this colour difference, empirical tests revealed that the red colour channel is best to differentiate the blue fluorescence from the paper layer and near-UV reflection from the compound. Thus, the compound surface corresponds to the highest value in the red channel. Then the image is equalized with a Contrast Limited Adaptive Histogram Equalization (CLAHE) and filtered with Otsu's threshold [4]. Then, a median filter is used with a kernel of 21 and a binary threshold of 250, both empirically determined. A segmentation example is shown in figure 2. Finally, the open-source computer vision algorithm minimal area rectangle from OpenCV[7] generates an oriented bounding box around the highlighted compound in the image [4].



Figure 2. Left: RGB vision. Middle: RGB vision with near-UV spotlight. Right: Segmentation result

[7]https://opencv.org/

4 Control Law

Depending on the orbital sander diameter, the 2D bounding box in the RGB image frame is segmented into smaller rectangles corresponding to the sanding passes necessary to cover the full compound surface to finish. Empirically, 2 or 3 passes are needed to cover most seams. The sequence of rectangles, from top-left to bottom-right, form a path for the orbital sender to follow. This 2D path is converted to 3D in the reference frame of the stationary robot by finding the corresponding points in the depth channel of the RGB-D camera.

The 3D path points are converted into robot's joint position with KDL inverse kinematics available with ROS's moveIT package[8] and sent to the robot with a modified version of ROS indsutrial's KAREL drivers[9]. The Fanuc's force control schedule is adjusted with a closed-loop force control of 1.1N (empirically determined) and with the parameters tuned to reduce vibration [5]. This make sure the arm applies a constant contact with a varying environment.

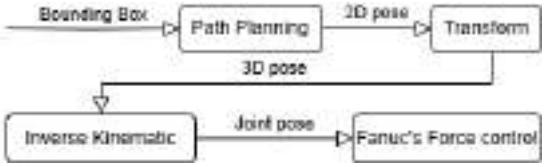


Figure 3. Block diagram of the control law

To make a good assisting robot for drywall finishing, the robot should at least remove the rougher bumps and cracks that workers would remove with similar power tools, without removing too much material so that workers may remove the finer details with handheld sanders. Also, removing too much material would leave visible traces under the paint. To leave the workers with enough material for the sanding block finish, the robot should remove slightly less material than the workers.

Three parameters determine the depth of material removed [6]: the grit of the sanding paper (Preston coefficient), the relative velocity between the compound and the orbital sander, and the pressure applied by the robot arm. Using the right equipment is key to control the first two parameters. Although the specific Preston coefficient between the compound and the sander abrasive was not examined in this study, using the finest grit abrasive reduces Preston coefficient, thus also the amount of material removed. In the same way, using the lowest orbital sander speed also reduces the amount of material removed. Reducing the commanded force in the Fanuc contouring controller would reduce the pressure, but it also reduce the performance of the algorithm to maintain contact on the compound surface.

[8] https://moveit.ai/
[9] https://wiki.ros.org/fanuc/

Although not all the sanding would be done automatically, this robot would reduce most of the workers' inconvenience with dust exposure and strain on the body, but also reduce human labour time while maintaining workers control over quality.

5 Experimental setup and results

The HandyScan 300 handheld external precision laser scanner, also known as the HandyScan SILVER[10] is used to measure the results of the finish. The difference between the scan before and the scan after the finish determines the depth of compound removed. To do so, scans are realigned using an Iterative Closest Point (ICP) algorithm with HandyScan's positioning targets to obtain the transformation between both frames of reference. Differences between point cloud positions in the Z axis serve as a basis for comparison.

The average depth in the Z-axis across the joint before and after the finishing, as shown in bottom figure 4, can help visualize the orientation of the two gypsum boards and the shape of the compound on the seam.

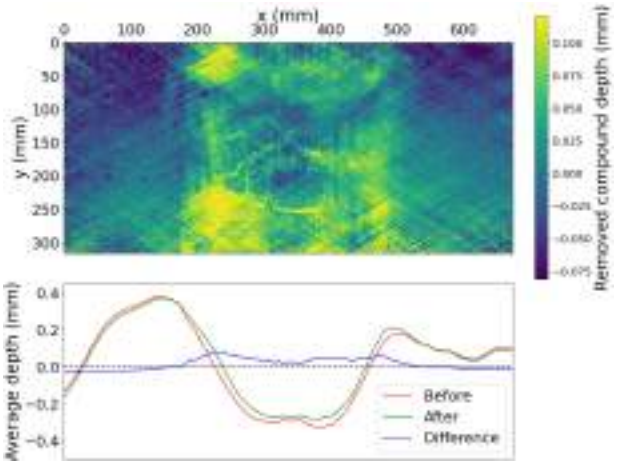


Figure 4. Top: Position difference in the camera Z axis before and after finishing. Bottom: Means position in the camera Z axis across the joint

5.1 Worker baseline

To investigate how much material is removed during workers' drywall finishing, 7 sample scans from different seams were taken before and after the complete finishing process (power tool-assisted and hand-finished). The baseline of what a typical drywall finish is the average of all depth removed plus twice the average standard deviation

[10] https://www.creaform3d.com/en/handyscan-3d-silver-series-professional-3d/technical-specifications

Table 2. Worker and robot depth removed (mm)

	Worker		Robot	
	Means	σ	Means	σ
1	0.10	0.04	0.08	0.03
2	0.11	0.07	0.09	0.06
3	0.04	0.04	0.10	0.03
4	0.04	0.04	0.14	0.05
5	0.08	0.05	0.09	0.03
6	0.13	0.08		
7	0.04	0.03		
8	0.16	0.04		
Means	0.09	0.05	0.10	0.04
Upper bound	0.19		0.18	

for each joint in Table 2. With a normal distribution, two standard deviations over the mean represent 95% of all measurements. This gives an approximate upper bound of how much material is removed from the general surface without measuring the max depth measured, which is where the large bump would skew the result.

On average, the upper bound of the worker's measured depth is 0.19 mm (average mean of 0.09 plus twice the average standard deviation of 0.05). Large bumps caused by uneven drying will have more material removed and larger cracks will have relatively less material removed.

5.2 Robot finishing

With adjusted orbital speed and commanded force, preliminary results presented in Table 2 show that the robot sanded an average amount of 0.10 mm with an average standard deviation of 0.04 mm. Thus, 95% of the robot's finish is expected to be below approximately 0.18 mm. With the upper bound of the workers estimated at 0.19 mm, it leaves enough material for the workers' finer handheld tool finish. The workers' upper bound is a useful benchmark for the robot's adjustment, but it is not a hard limit because many of the workers' and robot's individual scans had a mean and standard deviation higher than the expected workers' upper bound and they still did not damage the gypsum board. Furthermore, individual seams with a larger amount of irregular compound features, such as bumps, necessarily lead to a larger amount of material removed, even if the applied finish is the same. Thus, the proposed system performs as expected regardless of the amount of irregular compound features.

6 Conclusion

A construction robot has been tested to assist the finishing of drywall. The robot is composed of a machine vision algorithm to plan the trajectory of the robot arm, the integrated Fanuc force controller to regulate the pressure of the orbital sander on the wall, and a mobile platform to navigate autonomously inside the off-site construction environment. Preliminary results show that the robot can detect compound on gypsum boards and use the orbital sander mounted on the robot arm to do most of the drywall finishing. Compared to the workers, the robot sanded out a similar quantity of compound and left enough material to adjust the finish if necessary, without removing too much material as to irreparably damage the surface.

Future work includes confirming the repeatability of the preliminary test, introducing the robot to a production line for an assessment by the workers and investigating the automation of the mobile platform, enabling the whole system to navigate autonomously between drywall joint. We will also investigate the applicability of the approach to on-site construction settings.

Acknowledgements

The design of the system has been done in partnership with R.C.M. Solutions Modulaires Inc. and this work was supported by Mitacs Business Strategy Internship (BSI) grant IT34280 and the NSERC-CREATE program on collaborative robotics in manufacturing (CoRoM).

References

[1] John Burrows. *Canadian wood-frame house construction*. Canada Mortgage and Housing Corporation, Ottawa, 2nd combined imperial/metric ed. ; rev. ed edition, 2006.

[2] L. Sznelwar, F. Mascia, and U. Montedo (editors. Drywall Finishing Industry: Macroergonomic Evaluation and Intervention Design, 2008.

[3] Christopher S. Pan, Sharon S. Chiou, Hongwei Hsiao, Paul Becker, and Magdy Akladios. Assessment of perceived traumatic injury hazards during drywall taping and sanding. *International Journal of Industrial Ergonomics*, 25(6):621–631, July 2000. doi:10.1016/S0169-8141(99)00049-9.

[4] Samuel Faucher. Identification de surfaces de cloisons sèches par un système robotisé pour l'automatisation de tâches de poncage. *Université de Sherbrooke*, 2023. URL `http://hdl.handle.net/11143/21284`.

[5] Marko Švaco, Nikola Vitez, Bojan Šekoranja, and Filip Šuligoj. Tuning of Parameters for Robotic Contouring Based on the Evaluation of Force Deviation. *Transactions of FAMENA*, 42(3):33–45, October 2018. doi:10.21278/TOF.42302.

[6] Lukáš Veselý, Ondřej Matoušek, Tomáš Vít, and Jan Novosád. Optimization of the polishing process on a polishing machine with horizontal overarm. *EPJ Web of Conferences*, 264:01048, 2022. doi:10.1051/epjconf/202226401048.

BIM-FM interoperability: integrating existing FM platform with visualization of IFC models

Andressa Oliveira[1], José Granja[1], Pedro Machado[2], Ali Motamedi[3], and Miguel Azenha[1]

[1]University of Minho, ISISE, ARISE, Department of Civil Engineering, Portugal
[2]Matosinhos City Hall, Portugal
[3]Université du Québec à Montreal, École de Technologie Supérieure, Canada

Email Andressa Oliveira, Email José Granja, Email Pedro Machado, Email Ali Motamedi, Email Miguel Azenha

Abstract -

This article documents the developments from a collaboration with the Matosinhos City Hall in Portugal. The collaboration included the development of a web platform to integrate a management software already in use by the city hall, with 3D visualization capabilities, and the development of information requirements to guide the preparation of information models to be visualized on the platform. The platform enables the visualization of 3D models developed according to the IFC (Industry Foundation Classes) schema using the IFCjs library to manipulate, investigate and visualize IFC files. In addition, the web platform enables access to and manipulation of operational data through the 3D model. The paper demonstrates how the IFC model elements are linked to the database of an existing management system. This solution addresses current operational limitations and promotes broader BIM adoption in the FM sector.

Keywords -

Facility Management (FM); Building Information Modelling (BIM); Interoperability; Integration; Web Platform

1 Introduction

Facility Management (FM) plays a critical role in ensuring facilities' functionality, comfort, and safety [1]. The adoption of Building Information Modelling (BIM) by facility managers to enhance operational efficiency introduces unique challenges [2]. To promote a broader implementation of BIM in FM, it is essential to advance professional knowledge of information management processes that align with ISO 19650 series guidelines and have computational resources to support such digitalization. While computerized management platforms are widely utilized, the effectiveness of these systems depends on their ability to integrate diverse data sources [3, 4]. This includes incorporating three-dimensional models to enable comprehensive and efficient management. Within this context, a collaboration between Matosinhos City Hall (CMMatosinhos) in Portugal and the University of Minho (UMinho) aimed to enhance municipal asset management by integrating BIM methodology into the Intelligent Maintenance Management Platform (IMMP) already employed by CM-Matosinhos (Infraspeak). Currentlly the IMMP provides access to a web interface that lacks the capacity for visualizing buildings and assets.

2 Development methodology and functional requirements identification

The collaboration aimed to develop a web platform that integrates CMMatosinhos's IMMP with the 3D visualization of buildings, enabling users to consult, edit, and add information about the assets. It involved defining information requirements for IFC models to be used on this platform. The requested platform's features and functions included a building selection interface, 3D visualization of buildings and assets, displaying building's and assets' identification and open requests, and enabling managers to access the IMMP page for details. Users can also submit requests for specific assets. Two main pages have been identified for the web platform: the building selection page and the interaction page, supporting two user profiles: manager and common user. Both can visualize the building in 3D, view identification information, and submit requests. Managers can also access open requests and IMMP pages. The frontend uses HTML, CSS, and JavaScript, with the IFCjs library for the interactive viewer and investigation of the IFC file. The backend is built in Python with the Flask framework. The building was modelled using Autodesk Revit 2023 and exported in the IFC4 ADD2 TC1 schema, following the information requirements developed for this research (section 3.2).

3 Implementation

To develop the proposed solution, the 'IMMP database' must be integrated into a web platform presenting the IFC model visualization. This integration necessitates the web platform to establish a correlation between specific assets within the IFC model and their equivalent within the IMMP platform. Achieving this integration depends on the

unique identification of elements across both data sources and establishing a mapping between them. For the mapping, it is essential to comprehend how the assets are identified within the IMMP database (IMMP identifiers) and how CMMatosinhos identifies them in its inventory (CMMatosinhos identifiers). The 'CMMatosinhos identifiers' were organized as part of the alphanumeric information requirements for the model (as detailed in section 3.2). These identifiers became properties associated with various elements within the model, correlating them to corresponding items in the CMMatosinhos inventory. In contrast, the 'IMMP identifiers' were not included in the IFC model but were used to establish a one-to-one mapping between the elements in the IMMP database and those in the model.

3.1 Unique identifiers

Considering the functionalities to be implemented (section 2), three types of elements are essential: *locations*, *equipment* and *requests*. *Locations* represent spatial assets that comprise the buildings and are essential to the equipment's localization. *Equipment* is any physical asset from the CMMatosinhos' inventory (vehicles, chairs, ladders, etc.). Finally, *requests* represent the solicitation for an operation action to be carried out and are always associated with the other types of elements (*equipment* or *location*).

IMMP identifier : The IMMP database contains two categories for entering elements of the type *location*: LOCATION and ELEMENT. Within the LOCATION category, *locations* are organized hierarchically (from buildings to zones and to rooms) and identified by the attribute 'local_id'. Within the second category (ELEMENT), it is possible to enter elements of two types: EQUIPMENT and LOCAL. Therefore, the *locations* classified as ELEMENT are of the type LOCAL and represent only individual rooms (i.e. an office but never a building). Regarding the elements of type *equipment*, those are entries of the ELEMENT category and EQUIPMENT type, and the 'element_id' attribute uniquely identifies each element. Therefore, the 'local_id' and 'element_id' attributes are the IMMP identifiers for *locations* and *equipment*. For the *requests* elements, those designated as 'open' are relevant for this work since they still require actions. Each open *request* is uniquely identified by the 'failure_id' attribute and includes a 'local_id' attribute that specifies the *location* associated with the *request*. Contrarily, information regarding its connection to *equipment* is accessed through the relationships of the *equipment* elements, which contain the corresponding 'failure_id' codes. The 'IMMP identifiers' established here are generated automatically by the IMMP, being a sequence of numbers (e.g. 586188) that do not correspond to the CMMatosinhos inventory.

CMMatosinhos identifier : 'CMMatosinhos identifiers' refers to the coding used by CMMatosinhos to identify its assets within its inventory. Each asset within the IMMP has a unique identifier specific to the IMMP (explained previously) but also includes the CMMatosinhos inventory code associated with one of its attributes. Of the three elements (*equipment*, *location*, and *requests*), *requests* do not have a CMMatosinhos identifier as they do not represent physical assets, being not part of CMMatosinhos' inventory. As for *locations*, the value of the 'full_code' attribute represents the codification used by CMMatosinhos to identify spaces, and it contains a unique code for each location (e.g. 009.G1.P0.048). For *equipment*, the value of the 'code' attribute is a potential unique identifier. Within the CMMatosinhos inventory, the value of this attribute relates to the type of equipment (e.g. #CAD.ESCR.0056 for an office chair). However, considering that there are cases where multiple *equipment* share the same inventory code, a combination of the 'code' attribute of the equipment and the 'full_code' attribute of the location where it is situated has been defined for identifying elements within the equipment group.

3.2 Level of information need

The information requirements developed in this work include the level of information need [5]. Given the platform features (section 2), the purpose for the model is the visualization of the building and assets, incorporating only alphanumerical information for building identificaton and integration with the IMMP. Table 1 presents the level of information need for the complete project (*IfcBuilding*), where Code represents the 'CMMatosinhos identifier'. The requirements for individual architectural elements are presented together with those within its interior.

Table 1. Level of information need for the building

Objects:	**Building**	
IFC Class:	**IfcBuilding**	
Alphanumerical information		
Attributes		
	Attribute name	Content
	LongName	Name
Properties		
Group	Property name	Content
**	Local_CMMatosinhos	Code
Legend	** CMMatosinhos_Identification	

Regarding the elements composing the building's interior, (*locations*, *equipment*, and *requests*), *requests* were excluded as their visualization is not necessary. Therefore, three main groups were established for modelling:

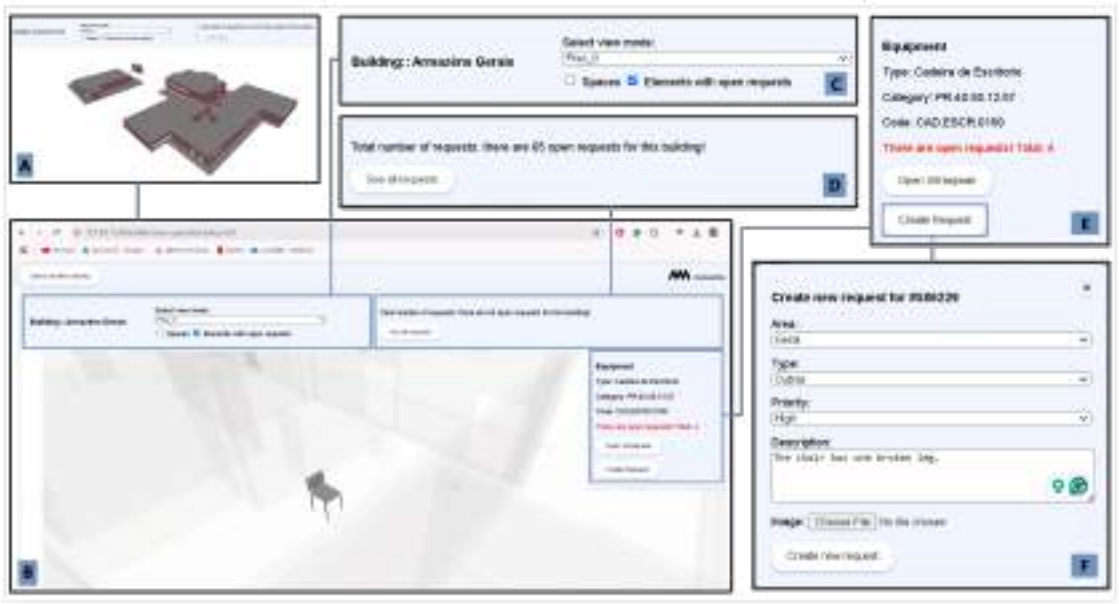


Figure 1. Web platform

architectural elements, equipment, and locations (table 2). Architectural elements serve for visualization and only require geometrical information. The alphanumerical information required for locations and equipment relates to the 'CMMatosinhos identifiers' and also some additional information for identifying and classifying the assets.

3.3 Web platform

Figure 1 shows screenshots of the web platform from the perspective of a user of the manager type. Screenshot A shows the interaction page after the model is loaded with a 3D view of the entire building. When an asset is selected, the platform appears as screenshot B. The page includes an upper section (screenshot C) with the building's identification and diverse options of visualization modes. The next section (screenshot D) warns about the presence and number of open requests. The side menu (screenshot E) displays information about the selected asset, and it appears on the user interface only when a piece of equipment or space is selected. In this menu, it is possible to visualize information identifying the asset and its number of open requests. In addition, two buttons allow the redirection to the asset's page within the IMMP and the creation of requests (screenshot F) associated with the asset. The request is registered within the IMMP platform, and the total number of open requests is updated automatically.

Figure 2 shows the information flow between the platform's frontend and backend and how the platform reacts to the manager's interactions. When a manager selects a building from a dropdown, a request retrieves the IFC file from the models' database and information from the IMMP database. Following, a 3D visualization of the building is displayed for the user, while the IFC file is investigated to retrieve the 'CMMatosinhos identifiers' of the building elements. The backend combines these 'CMMatosinhos identifiers' with 'IMMP identifiers' to map elements from both sources. This mapping allows for the identification of elements with open requests, enabling the visualization to filter and display only those elements. Additionally, the number of open requests related to the building is counted and presented to the user. Following this, users can interact with the model, view detailed asset information, access the asset's IMMP page, or create linked requests.

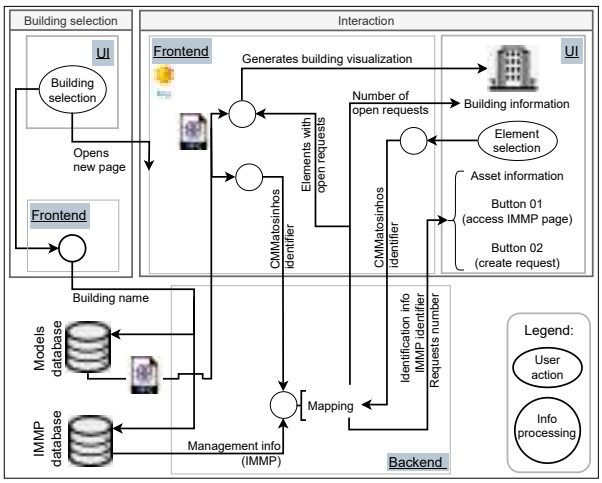


Figure 2. Platform's information flow diagram

Table 2. Level of information need for architectural elements, equipment and locations

Objects:		**Archit. elements**	**Equipment**	**Locations**
IFC Class:		**Variable**	**Variable**	**IfcSpace**
Geometrical information				
Detail		Real representation of external limits. Single element without layers or components.	Simplified outer shell, real volumes and dimensions. Single element without layers or components.	Real volume.
Dimensionality		3D	3D	3D
Location		Relative	Relative	Relative
Appearance		Colour must be similar to reality, without textures.	Colour must be similar to reality, without textures.	Not visible
Alphanumerical information				
Attributes				
	Attribute name	Content		
	LongName	X	X	Name
Properties				
Group	Property name	Content		
**	Codigo_CMMatosinhos	X	Code	X
	Local_CMMatosinhos	X	X	Code
	Categoria_CMMatosinhos	X	Category	X
	Tipo_CMMatosinhos	X	Category description	X
Legend	**Variable** IFC classes vary depending on the architectural element or equipment. ****** Name of group of properties: CMMatosinhos_Identification **X** Objects of this type must not contain this property or attribute.			

4 Conclusion

This work developed a BIM-based solution for FM, integrated with a previously non-BIM-supporting management platform. The outcomes of this work provide a detailed procedure for future integration processes. This collaboration demonstrates that the operational sector does not need to completely alter its management system or adopt a different management platform to implement BIM effectively. However, it does need to prepare its inventory organization to facilitate digitalization. During the collaboration, an iterative process with CMMatosinhos helped to address the limitations encountered, showing that the close participation of managers is crucial for the FM sector's digitalization.

5 Acknowledgements

This work was partly financed by FCT / MCTES through national funds (PIDDAC) under the R&D Unit Institute for Sustainability and Innovation in Structural Engineering (ISISE), under reference UIDB / 04029/2020 (doi.org/10.54499/UIDB/04029/2020), and under the Associate Laboratory Advanced Production and Intelligent Systems ARISE under reference LA/P/0112/2020. This work is also financed by national funds through FCT – Foundation for Science and Technology, under grant agreement PRT/BD/154416/2023 attributed to the 1st author, under the MIT Portugal Programme.

References

[1] International Facility Management Association. What is facility management, 2023. URL `https://www.ifma.org/about/what-is-fm/`.

[2] Lidia Pinti, Ricardo Codinhoto, and Serena Bonelli. A review of building information modelling (bim) for facility management (fm): Implementation in public organisations. *Applied Sciences (Switzerland)*, 12, 2022. ISSN 20763417. doi:10.3390/app12031540.

[3] Serdar Durdyev, Mojtaba Ashour, Scott Connelly, and Amir Mahdiyar. Barriers to the implementation of building information modelling (bim) for facility management. *Journal of Building Engineering*, 46:103736, 2022. ISSN 23527102. doi:10.1016/j.jobe.2021.103736.

[4] Stefania Siccardi and Valentina Villa. Trends in adopting bim, iot and dt for facility management: A scientometric analysis and keyword co-occurrence network review. *Buildings*, 13, 2023. ISSN 20755309. doi:10.3390/buildings13010015.

[5] European Commitee for Standardization. En 17412-1. building information modelling. level of information need. concepts and principles, 2020.

Efficient RAG(Retrieval-Augmented Generation) for Construction in Low-Resource Language

Kichang Choi[1], **Seungwon Baek**[1], **Jonewon Ma**[2], **Hongjo Kim**[1]

[1]School of Civil and Environmental Engineering, Yonsei University
[2]Department of Building, Civil, and Environmental Engineering, Concordia University

amki1027@yonsei.ac.kr, baeksw@yonsei.ac.kr, jongwon.ma@concordia.ca, hongjo@yonsei.ac.kr

Abstract -

The Retrieval-Augmented Generation (RAG) framework has gained attention as a fast and cost-effective method for enhancing the performance of large language models (LLMs). However, its performance remains limited in minority languages such as Korean, and this issue is exacerbated in specialized fields like construction. To address these limitations, this study proposes a dataset construction method that allows low-cost fine-tuning of embedding models originally trained on English-based data. By applying this method in the construction domain, we achieved a top-1 document retrieval accuracy of 58.65%, surpassing the performance of a commercial embedding model provided by OpenAI. We further analyzed how improvements in the embedding model influence the overall RAG pipeline and present both a dataset creation approach and an appropriate evaluation strategy for testing RAG's performance. Our findings suggest that this method can significantly enhance technical efficiency by providing a foundation for diverse language users to effectively utilize RAG in the construction domain.

Keywords -

Retrieval-Augmented-Generation; Embedding model; Construction; LLM; Retrieval; Fine-tuning

1 Introduction

Efforts to enhance productivity across various fields, including the construction industry, are continuously advancing. Successful examples include the application of deep learning models for optimizing schedules in large-scale construction projects [1] and detecting pavement cracks [2]. However, applying pretrained models to specialized domains such as construction requires domain adaptation through fine-tuning, which demands significant costs and resources [3, 4]. This is considered a major issue in utilizing deep learning technologies in the construction industry and is one of the challenges that must be addressed.

To mitigate this issue, many attempts have been made to adopt RAG [5]. RAG allows LLMs to generate answers by retrieving relevant information from external databases instead of relying solely on internally stored knowledge. This approach reduces the need for costly LLM fine-tuning. However, the embedding models and LLMs, which are core components of RAG, demonstrate high performance only in English or general-purpose language contexts [6, 7]. Their performance drops significantly in low-resource languages such as Korean or specialized domains like construction. Among these components, embedding models play a critical role in external database generation and retrieval. When domain adaptation is not applied, embedding models exhibit low retrieval accuracy, which negatively impacts the performance of the entire RAG pipeline, as confirmed through our experiments.

To address this challenge, we constructed a fine-tuning dataset specialized for the Korean construction domain and fine-tuned the embedding model to improve RAG performance, as shown in Figure 1. Publicly available construction standards were utilized to extract high-quality sentences, and we proposed a method to generate fine-tuning datasets using LLMs at a low cost. This approach enabled the creation of high-quality fine-tuning datasets for under $10. Using this dataset, we trained a small embedding model that outperformed OpenAI's commercial embedding model, "text-embedding-3-large" [8]. Our final approach achieved a top-1 document retrieval accuracy of 58.65%. This demonstrates that the proposed method makes LLMs and RAG more accessible and practical for low-resource construction domains.

2 Related Work

The development of LLMs has brought significant advancements across various fields [9]. However, as the size of these models increases, training costs rise substantially [10], and limitations such as hallucination remain prevalent [11]. To address these challenges, RAG has been introduced [12]. RAG enhances LLM responses by retrieving relevant information from external databases rather than relying solely on internal knowledge. This approach is particularly effective for tasks requiring domain-specific knowledge or up-to-date information [13].

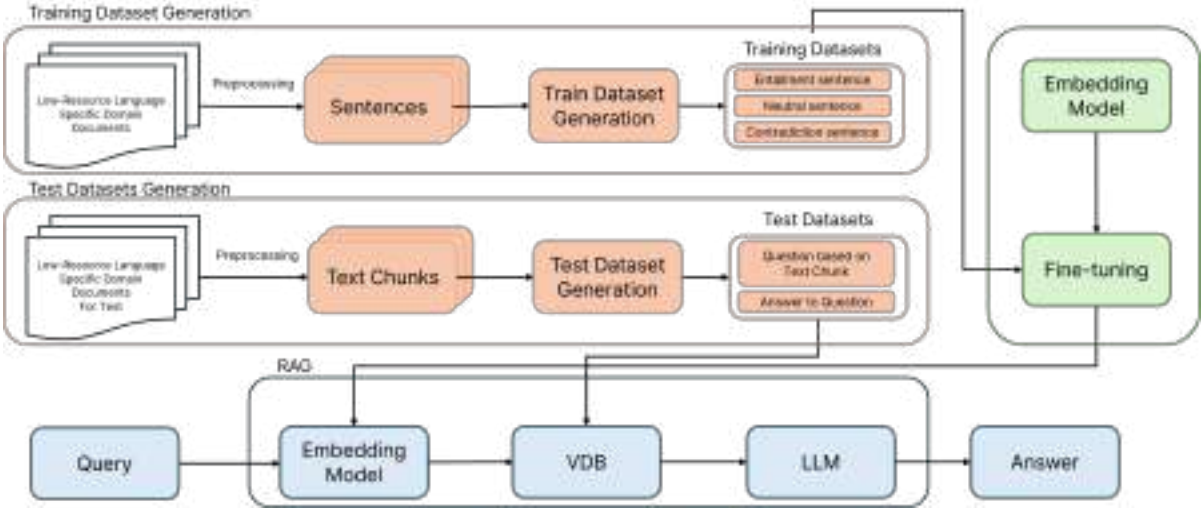


Figure 1. Workflow of embedding model fine-tuning and RAG pipeline construction. This figure illustrates the process of generating a fine-tuning dataset for the embedding model and integrating the fine-tuned model into the RAG pipeline for improved retrieval performance.

The performance of RAG heavily depends on the quality of its Retriever, which retrieves relevant external data. Dense retrieval methods, in particular, rely on embedding models to convert text into high-dimensional vectors and measure semantic similarity between queries and documents [14]. These embedding models are critical to the effectiveness of the RAG pipeline, as their ability to capture semantic meaning, contextual relationships, and syntactic nuances directly impacts retrieval accuracy. Earlier embedding techniques, such as Latent Semantic Analysis (LSA) [15] and Latent Dirichlet Allocation (LDA) [16], have been superseded by methods like Word2Vec [17] and Transformer-based architectures, including BERT [6]. Today, advanced embedding models are available as commercial APIs, such as OpenAI's high-performance models [8].

Recently, LLMs have also been applied in the construction industry. Examples include waste recognition [18], project schedule management [19], and generating captions for site images to monitor progress [20]. However, these applications remain in the early stages and primarily rely on general-purpose language processing capabilities. Discussions about embedding models tailored to construction-specific tasks are rare. To improve RAG performance in low-resource languages and specialized domains, embedding models fine-tuned on domain-specific data are essential.

Fine-tuning pre-trained models for specific tasks has become a common approach for achieving high performance in domain-specific applications [21]. In the construction domain, fine-tuning has been successfully applied to tasks such as automated compliance checking of building codes [22], pavement defect detection [23], and construction management system development [24]. However, constructing high-quality training datasets for fine-tuning in low-resource languages or niche domains is often costly, as it requires expert annotations [25].

Several methods have been proposed to address these challenges and improve the efficiency of dataset generation. One approach involves using LLMs, such as GPT-3 [26], to generate synthetic datasets automatically, which is particularly useful in low-resource language settings [27]. Another method is pseudo-labeling, where unlabeled data is annotated using pre-trained models, effectively expanding the dataset size and diversity [28]. Finally, small curated datasets have been shown to yield high performance with minimal data, as demonstrated by LIMA [29].

In this study, we propose an automatic dataset generation method using LLMs to improve RAG performance in the Korean construction domain. This approach enables the cost-effective creation of high-quality datasets while enhancing the efficiency of domain-specific embedding models.

3 Methodology

3.1 Dataset Generation

This study proposes a cost-efficient dataset generation method aimed at reducing the costs of data creation while improving the quality and applicability of the datasets to RAG systems. Figure 2 illustrates the proposed dataset generation pipeline.

To construct the training dataset, we first collected high-quality sentences written in Korean and relevant to the con-

Table 1. Training dataset structure for fine-tuning. This table presents the composition of the training dataset, where each original sentence S_1 is paired with three corresponding sentences S_2 representing Entailment, Neutral, and Contradiction relationships.

S_1	S_2 Entailment	S_2 Neutral	S_2 Contradiction
The manufacturer specializes in producing materials specified in the design documents. They evaluate production records, supply performance, and cases of product defects to select a suitable manufacturer.	They select manufacturers with a proven track record of high quality and consistently reliable deliveries.	They seek manufacturers with high production capacity, reliable supply performance, and competitive pricing.	The manufacturer only uses material types specified in the design documents, and no quality issues occur.
The concrete for the wall section is poured so that each part always maintains almost the same height.	The height is consistently maintained when pouring concrete for the wall section.	When pouring concrete for the wall section, they aim to keep the height uniform for each layer.	While pouring concrete for the wall section, some areas were made lower than others.
The allowable deviation of the target air content must be within ±1.5%.	The target air content must have an allowable deviation within ±1.5%.	The allowable deviation of the target air content should be maintained within ±1.5%.	The allowable deviation of the target air content may be ±2.0%.

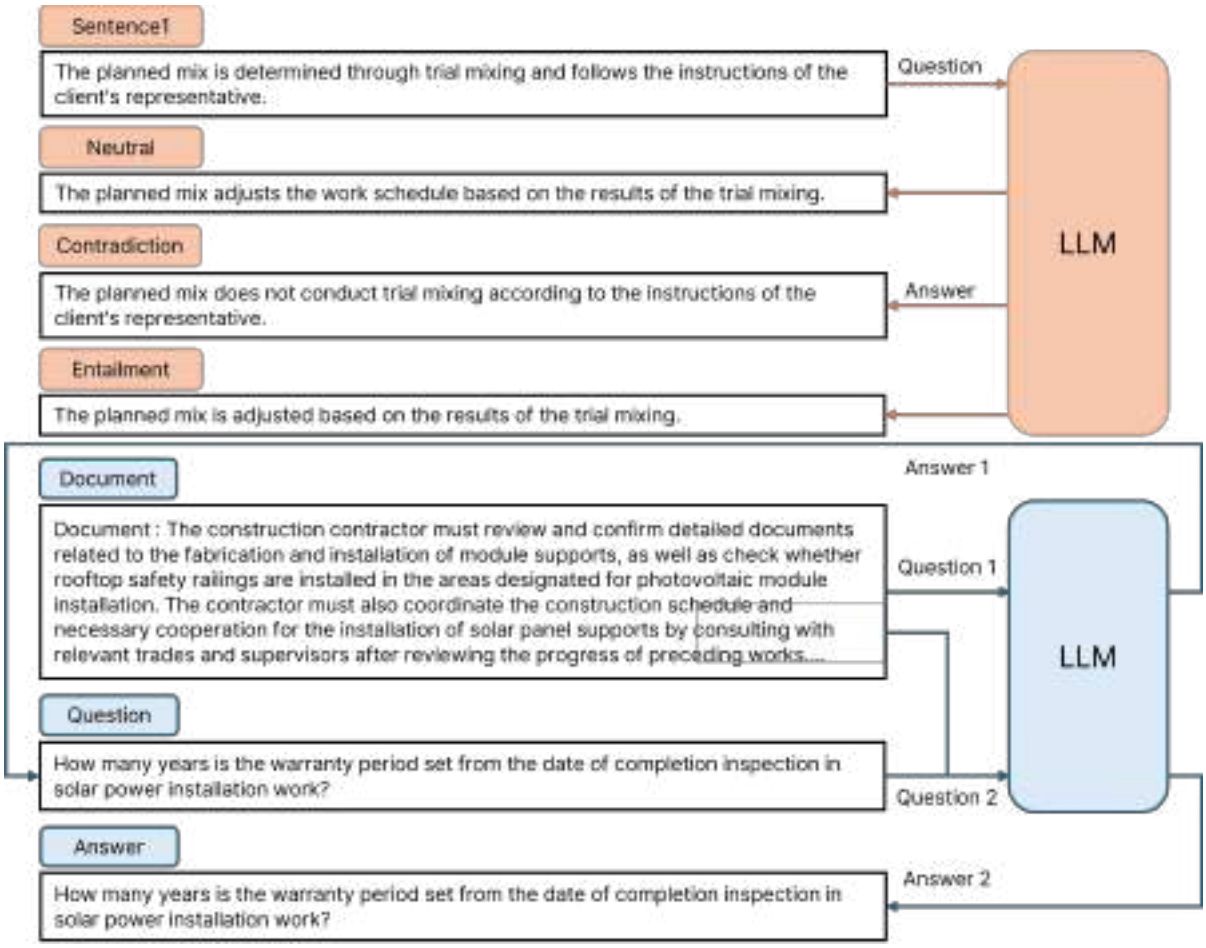


Figure 2. Process of training and test dataset generation. This figure visually represents the sequential generation of training and test datasets using an LLM. The training dataset is constructed to fine-tune the embedding model, while the test dataset is generated to evaluate retrieval performance.

struction domain. Publicly available construction specification documents were identified as appropriate sources for this purpose. Many non-English-speaking countries, including Korea, Japan, China, Vietnam, and France, maintain construction specifications in their native languages, which align well with the objectives of this study. Specifically, we extracted 8,499 sentences from the Korean Construction Specifications. Each of these sentences, referred to as S_1, was expanded into three additional sentences, S_2, representing entailment (positive), neutral, and contradiction (negative) relationships. This resulted in a total of 8,499 S_1 - S_2 pairs. The resulting dataset, named Ko Con NLI (Korea Construction Natural Language Inference), was generated using GPT-4o API with a simple data generation prompt. The entire process of generating all 8,499 pairs through the GPT-4o API incurred a total cost of less than $10.

The final datasets are summarized as follows. The training dataset consists of 8,499 S_1 - S_2 pairs, generated from Korean construction documents, with all data used for training to maximize performance. Validation was performed by observing improvements in Retriever performance based on the embedding model's vector database. Examples of the datasets are provided in Table 1.

3.2 Embedding Model Fine-tuning

To fine-tune the embedding model, this study employed Multiple Negative Ranking Loss (MNRL), which is widely recognized for its effectiveness in retrieval-based tasks.

The MNRL loss function is defined as follows:

$$J_{\text{MNRL}}(\theta) = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{\exp\left(\sigma\left(f_\theta(x_i), f_\theta(y_i)\right)\right)}{\sum_{j=1}^{N}\exp\left(\sigma\left(f_\theta(x_i), f_\theta(y_j)\right)\right)} \tag{1}$$

where N is the batch size, f_θ is the sentence encoder function mapping sentences to the embedding space, and σ is a similarity function, such as cosine similarity. x_i represents the anchor sentence in the i-th pair, y_i is its corresponding positive sentence, and y_j includes all sentences in the batch, including those treated as negative samples.

MNRL optimizes retrieval performance by increasing similarity between positive pairs while minimizing similarity with negative samples, as formulated in Equation 1.

4 Experiment

4.1 Experiment Setup

This study aimed to improve the performance of RAG systems in the Korean construction domain by fine-tuning embedding models and comparing the performance of the fine-tuned models with baseline models. The dataset used for fine-tuning was the **Ko Con NLI dataset**, which was generated by extracting sentences from the Korean Construction Specifications using LLMs, as described earlier. This dataset consists of sentence pairs representing positive, neutral, and negative relationships and was specifically designed to be suitable for embedding model fine-tuning.

The baseline model selected for comparison was OpenAI's "text-embedding-3-large" model, a widely recognized embedding model known for its excellent performance across multiple languages and domains. This model is commercially available. For fine-tuning, the KLUE-RoBERTa-base model[30], pre-trained on the KLUE dataset, was used. Two experimental setups were considered:

1. Using the pre-trained KLUE-RoBERTa-base model without fine-tuning.

2. Fine-tuning the KLUE-RoBERTa-base model using MNRL.

All experiments were conducted on an NVIDIA GeForce RTX 4090 GPU using CUDA 12.2 and PyTorch. The AdamW optimizer was used with a learning rate set to 2×10^{-5}. Model performance was quantitatively measured by creating a vector database (VDB) and evaluating retrieval accuracy.

The retrieval process works by utilizing a test dataset consisting of pairs of documents and corresponding questions based on each document, ensuring that each document has a correct question. All documents are embedded to create a VDB, and when a query is given, it is embedded as well to retrieve the most similar document. Retrieval accuracy is evaluated by checking whether the retrieved document matches the correct document, and several metrics were applied to verify the accuracy of the proposed method.

4.2 Experiment Results

As shown in Table 2, the KLUE-RoBERTa-base model achieved a low Hit Rate @1 of 12.40% without fine-tuning. However, fine-tuning the model using MNRL significantly improved the Hit Rate @1 to 58.65%, demonstrating the effectiveness of the fine-tuning approach.

Notably, this result surpasses OpenAI's "text-embedding-3-large" model, which achieved a Hit Rate @1 of 52.67%, by approximately 6 percentage points, despite being a lightweight model trained on a dataset generated with less than $10 in GPT-4o API costs. This demonstrates that even in low-resource language environments, high-quality domain-specific documents can be leveraged to develop cost-effective and efficient embedding models that rival large-scale commercial models.

Table 2. Retrieval performance comparison of fine-tuned and baseline models.

Models	Learning Method	Hit Rate @1 (%)	nDCG@5	MRR@5
KLUE-RoBERTa-base	None	12.40	0.1983	0.1766
KLUE-RoBERTa-base	MNRL	58.65	0.6904	0.6621
OpenAI	None	52.67	0.6784	0.6349

The effectiveness of this approach stems from two key factors: (1) the structured dataset generation process and (2) the use of MNRL. Unlike manually curated datasets, which often include precise similarity scores or fine-grained semantic relationships between sentences, the dataset generated using GPT only provides categorical labels—such as Entailment (positive), Neutral, and Contradiction (negative)—without explicit numerical similarity scores. MNRL is well-suited for both cases, as it effectively optimizes the ranking of positive and negative samples regardless of whether explicit similarity scores are available. This characteristic makes MNRL particularly effective for training on automatically generated datasets, where manually assigned scores are not available.

This synergy between the generated dataset and MNRL loss function proved to be highly effective, enabling the fine-tuned model to achieve superior retrieval accuracy in the Korean construction domain. These findings suggest that even in specialized technical fields and low-resource languages, it is possible to construct lightweight yet high-performance embedding models using carefully curated domain data and appropriate training techniques.

The improvements in retrieval accuracy directly enhance the quality of final RAG-generated responses. Since RAG relies on retrieving relevant documents before generating answers, higher retrieval accuracy ensures that the language model is provided with more contextually relevant information. This reduces the risk of hallucinations, where the model generates inaccurate or misleading responses due to incorrect retrieval. As a result, the fine-tuned embedding model not only improves document retrieval efficiency but also significantly enhances the reliability and factual consistency of the generated responses in knowledge-intensive applications.

5 Conclusion, Limitations & Future Work

This study proposed a methodology to enhance the retrieval and answer generation performance of RAG systems in low-resource language environments, focusing on the Korean construction domain. By utilizing publicly available construction standards and LLMs, a domain-specific dataset was generated at a minimal cost of less than $10. This dataset enabled the fine-tuning of the KLUE-RoBERTa-base model, significantly improving retrieval performance. Specifically, the Hit Rate @1 of the untrained model, which was only 12.40%, increased to 58.65% after fine-tuning with MNRL. Furthermore, the fine-tuned model demonstrated approximately 6% higher retrieval performance compared to OpenAI's "text-embedding-3-large" model.

Despite its promising results, this study has several limitations, which open avenues for future research. While the study focused on improving the retrieval performance of the RAG pipeline, it did not include a demonstration of final answer generation using LLMs. Although it is widely understood that retrieval performance significantly affects final answer quality, further research is needed to explore the extent of improvement in answer generation performance. Additionally, since the LLM in the RAG pipeline was not fine-tuned, there remain many unexplored areas regarding the level of quality that can be achieved in generating answers based on retrieved documents. Without LLM fine-tuning, the study could not determine the precise impact on answer generation quality.

This study relied on publicly available construction standards to generate high-quality datasets. However, in domains where domain-specific documents are scarce or unavailable, applying this methodology may present challenges. Future research should focus on developing generalized data generation techniques that can adapt to domains lacking structured or standardized documents. Furthermore, while the study focused on the Korean construction domain, the applicability of the proposed approach to other languages and technical fields remains untested.

Beyond the Korean construction domain, the proposed approach can be extended to other low-resource languages and specialized technical fields. Many languages, such as Vietnamese, Thai, and Arabic, face similar challenges in adapting retrieval models due to the lack of high-quality domain-specific training data. Leveraging publicly available documents and generative models enables low-cost fine-tuning for these languages. Furthermore, this methodology can be applied to various technical domains, including healthcare, legal systems, and engineering, where precise retrieval and accurate knowledge synthesis are critical. Future research can investigate domain-specific optimizations and cross-lingual adaptability to improve retrieval-based applications in low-resource settings.

Future research should validate this framework's scal-

Table 3. Retrieval Accuracy Metrics

Metric	Description
Hit Rate @1	Measures whether the correct document appears as the top retrieved result.
nDCG@5	Normalized Discounted Cumulative Gain; evaluates ranking quality by considering both relevance and position in the top 5 results.
MRR@5	Mean Reciprocal Rank; calculates the inverse rank of the first correct document within the top 5 retrieved results.

ability across other low-resource languages and specialized fields, including healthcare, legal, and educational domains.

By addressing these limitations, future research can further enhance the applicability, scalability, and robustness of RAG systems in low-resource environments, ultimately expanding their utility across a broader range of domains and languages.

6 Acknowledgements

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT)(RS-2024-00457308). ※ MSIT: Ministry of Science and ICT
This research was conducted with the support of the "2023 Yonsei University Future-Leading Research Initiative (No. 2023-22-0114)" and the "National R&D Project for Smart Construction Technology (No. RS-2020-KA156488)" funded by the Korea Agency for Infrastructure Technology Advancement under the Ministry of Land, Infrastructure and Transport, and managed by the Korea Expressway Corporation.

References

[1] Yuan Yao, Vivian WY Tam, Jun Wang, Khoa N Le, et al. Automated construction scheduling using deep reinforcement learning with valid action sampling. *Automation in Construction*, 166:105622, 2024.

[2] Lele Zheng, Jingjing Xiao, Yinghui Wang, Wangjie Wu, et al. Deep learning-based intelligent detection of pavement distress. *Automation in Construction*, 168:105772, 2024.

[3] Ning Ding, Yujia Qin, Guang Yang, Fuchao Wei, et al. Parameter-efficient fine-tuning of large-scale pre-trained language models. *Nature Machine Intelligence*, 5(3):220–35, 2023.

[4] Klaus Greff, Francois Belletti, Lucas Beyer, Carl Doersch, et al. Kubric: A scalable dataset generator. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3749–61, 2022.

[5] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33:9459–74, 2020.

[6] Liu Zhuang, Lin Wayne, Shi Ya, and Zhao Jun. A robustly optimized bert pre-training approach with post-training. In *Proceedings of the 20th Chinese National Conference on Computational Linguistics*, pages 1218–27, Huhhot, China, 2021.

[7] Zhilin Yang. Xlnet: Generalized autoregressive pretraining for language understanding. *arXiv preprint arXiv:1906.08237*, 2019.

[8] OpenAI. New embedding models and api updates. URL `https://openai.com/index/new-embedding-models-and-api-updates/`. Accessed: 2024-10-31.

[9] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, et al. Language models are unsupervised multitask learners. *OpenAI Blog*, 1(8):9, 2019.

[10] William Fedus, Barret Zoph, and Noam Shazeer. Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity. *Journal of Machine Learning Research*, 23(120):1–39, 2022.

[11] Joshua Maynez, Shashi Narayan, Bernd Bohnet, and Ryan McDonald. On faithfulness and factuality in abstractive summarization. In *Proceedings of ACL*, pages 1906–19, 2020.

[12] Kelvin Guu, Kenton Lee, Zora Tung, Panupong Pasupat, and Mingwei Chang. Retrieval augmented language model pre-training. In *Proceedings of the International Conference on Machine Learning*, pages 3929–38. PMLR, 2020.

[13] Tu Vu, Mohit Iyyer, Xuezhi Wang, Noah Constant, Jerry Wei, et al. Freshllms: Refreshing large language models with search engine augmentation. In

Findings of the Association for Computational Linguistics: ACL 2024, pages 13697–13720, Bangkok, Thailand, 2024.

[14] Vladimir Karpukhin, Barlas Oguz, Sewon Min, Patrick SH Lewis, et al. Dense passage retrieval for open-domain question answering. In *Proceedings of EMNLP*, pages 6769–81, 2020.

[15] Scott Deerwester, Susan T Dumais, George W Furnas, Thomas K Landauer, et al. Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6):391–407, 1990.

[16] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan):993–1022, 2003.

[17] Jeffrey Pennington, Richard Socher, and Christopher Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–43, Doha, Qatar, 2014.

[18] Ying Sun, Zhaolin Gu, and Sean Bin Yang. Probing vision and language models for construction waste material recognition. *Automation in Construction*, 166:105629, 2024.

[19] Fouad Amer, Yoonhwa Jung, and Mani Golparvar-Fard. Transformer machine learning language model for auto-alignment of long-term and short-term plans in construction. *Automation in Construction*, 132: 103929, 2021.

[20] Yoonhwa Jung, Ikhyun Cho, Shun-Hsiang Hsu, and Mani Golparvar-Fard. Visualsitediary: A detector-free vision-language transformer model for captioning photologs for daily construction reporting and image retrievals. *Automation in Construction*, 165: 105483, 2024.

[21] Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, et al. Deep contextualized word representations. In *Proceedings of NAACL-HLT*, pages 2227–37, New Orleans, Louisiana, 2018.

[22] Xiaorui Xue, Jiansong Zhang, and Yunfeng Chen. Question-answering framework for building codes using fine-tuned and distilled pre-trained transformer models. *Automation in Construction*, 168:105730, 2024.

[23] Tianjie Zhang, Donglei Wang, and Yang Lu. A data-centric strategy to improve performance of automatic pavement defects detection. *Automation in Construction*, 160:105334, 2024.

[24] Yunshun Zhong and Sebastian D Goodfellow. Domain-specific language models pre-trained on construction management systems corpora. *Automation in Construction*, 160:105316, 2024.

[25] Fan Yang, Lei Zhang, Sijia Yu, Danil Prokhorov, et al. Feature pyramid and hierarchical boosting network for pavement crack detection. *IEEE Transactions on Intelligent Transportation Systems*, 21(4): 1525–35, 2019.

[26] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, et al. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–901. Curran Associates, Inc., 2020.

[27] Or Honovich, Thomas Scialom, Omer Levy, and Timo Schick. Unnatural instructions: Tuning language models with (almost) no human labor. In *Proceedings of ACL*, pages 14409–28, Toronto, Canada, 2023.

[28] Elyas Asadi Shamsabadi, Seyed Mohammad Hassan Erfani, Chang Xu, et al. Efficient semi-supervised surface crack segmentation with small datasets based on consistency regularisation and pseudo-labelling. *Automation in Construction*, 158:105181, 2024.

[29] Chunting Zhou, Pengfei Liu, Puxin Xu, Srinivasan Iyer, et al. Lima: Less is more for alignment. In *Advances in Neural Information Processing Systems*, volume 36, pages 55006–21. Curran Associates, Inc., 2023.

[30] Sungjoon Park, Jihyung Moon, Sungdong Kim, Won Ik Cho, et al. Klue: Korean language understanding evaluation, 2021.

Assessment Framework for 3D Concrete Printing Technology in Alignment with Construction 5.0 Criteria

Imtiaz Iqbal[1], Svetlana Besklubova [2], and Tala Kasim[3]

[1,3]Department of Civil Engineering, Aston University, Birmingham, B4 7ET, United Kingdom
[2]Department of Industrial and Manufacturing Systems Engineering, The University of Hong Kong, Pok Fu Lam, Hong Kong
220425973@aston.ac.uk, lanabess@hku.hk, t.kasim@aston.ac.uk

Abstract –

The transition from Construction 4.0 to Construction 5.0 signifies a new era of innovative and advanced practices in the construction industry. This study explores the integration of 3D Concrete Printing (3DCP) technology within the Construction 5.0 paradigm, emphasising its potential to transform conventional practices by aligning with core criteria of Construction 5.0 such as human-centricity, sustainability, resilience, and cost and environmental control. A systematic literature review was conducted to identify the key criteria of Construction 5.0 and determine the critical drivers of 3DCP across each criterion. For each criterion, ten critical drivers were identified that align 3DCP with Construction 5.0 principles. The study employs content analysis to rank and prioritise these drivers, ensuring a robust understanding of their significance, supported by a comprehensive framework for evaluating their alignment with Construction 5.0. Additionally, a conceptual implementation framework for 3DCP was developed, encompassing three phases: knowledge building, the adoption phase, and performance assessment. This framework aims to facilitate the practical application of 3DCP within the construction industry and provides a roadmap for industry and academia to integrate 3DCP into Construction 5.0 practices effectively.

Keywords –

construction 4.0, construction 5.0, 3D concrete printing, human centricity, sustainability, resilience, cost and environmental control

1 Introduction

The construction industry has undergone significant transformations over the years, with the advent of Construction Industry 4.0 marking a pivotal shift through Internet of Things (IoT), artificial intelligence (AI), Cloud Computing (CM), Big data Analytics (DBA), robots and Cyber Physical Systems (CPS), 3DCP, Augmented and Virtual Reality (AR/VR) and blockchain technology to enhance efficiency, automation, and data-driven decision-making [1]. Construction Industry 4.0 primarily focuses on the technological aspects of construction, emphasising automation, digitisation, and data-driven processes, but it often overlooks the human-centric elements of the industry. Despite the potential of these technologies to improve the design, management, operation, and decision-making of construction projects, the industry's ability to fully integrate them remains low [2].

In response, the Construction 5.0 paradigm is the next phase in industrial development that aims to combine the skills of human experts in partnership with efficient and precise machines to achieve production solutions that are resource efficient and preferred by clients [3]. Rather than replacing workers with machines, Construction 5.0 seeks to create a balance where smart technologies such as AI, robotics, 3DCP, Building Information Modelling (BIM), Digital Twin (DT) and IoT, empower workers to become more efficient and safer while fostering innovation [4]. This human-centred approach addresses the evolving demands of the industry, including sustainability, precision, and adaptability. Construction 5.0 builds on advancements from Construction 4.0, prioritising the collaboration between people and technology. It ensures that workers remain central to the process, using technology as a tool to enhance rather than replace their capabilities [3]. Construction 5.0 is neither a replacement nor an alternative to Construction 4.0 but rather an evolution and logical continuation. It centres around four interconnected core values: human-centricity, sustainability, resilience, and cost reduction & environmental control [1].

Among the various promising technologies and applications that Construction 5.0 supports to enhance production and deliver customised products one key technology is 3DCP, which has capacity to completely transform the conventional construction process [5]. The construction 3D Printing process involves using digital

models to layer materials, such as concrete or polymers, to build structures by layering them systematically [6]. The use of this technology provides numerous benefits, including decreased labour needs, lower production expenses, increased productivity due to faster construction processes, cost savings, high efficiency, eco-friendliness, reduced waste generation, and the capability to construct intricate shapes and structures [5]. Thus, 3DCP is pivotal in advancing the construction industry towards the holistic goals of Construction 5.0.

Up to date very limited research studies have investigated the alignment of 3DCP technology with Construction 5.0 goals. Max et al. [7] examined the sustainability benefits of adopting 3DCP in the construction industry, highlighting its potential for significant material savings, reduced carbon footprints compared to traditional methods, increased automation, and addressing the current shortage of skilled labour in the sector. Himanshu et al. [8] highlighted the resilience of 3DCP structures, attributing it to enhanced customisation and faster adaptation to design changes, aligning with Construction 5.0 adaptability goals. Giovanni et al [9] explored the human-centric aspects of 3DCP, demonstrating its potential to create ergonomically safer environments by reducing manual labour. Wang et al. [10] analysed the integration of cloud manufacturing and IoT in 3DCP, highlighting its role in digitalising construction to provide reliable features, enable real-time monitoring, and improve operational efficiency, aligning with Industry 5.0 focus on technology-human collaboration. The above studies collectively underline the significance of 3DCP in fulfilling the core values of Construction 5.0, including human-centricity, sustainability, resilience, and cost reduction & environmental control. They also highlight the need for an integrated assessment framework to optimise its adoption.

A key limitation in the existing body of research is the absence of structured frameworks to evaluate how effectively 3DCP aligns with the criteria of Construction 5.0. This study seeks to fill this critical gap by developing an assessment and implementing framework that systematically evaluates 3DCP technology in alignment with Construction 5.0 criteria.

2 Research Model Development

The research methodology consists of four key steps. The first step includes a literature search. A comprehensive search was conducted across the databases including Scopus, Web of Science, and Google Scholar were utilised to retrieve peer-reviewed journal articles, conference papers, and review articles published in English since 2015. Keywords including “Industry 4.0,” “3D Concrete Printing,” “Industry 5.0,” “Construction Industry,” and “digital Construction” were used to find the articles. This ensured the inclusion of up-to-date and high-quality research addressing the study objectives.

Secondly, a systematic analysis was conducted to explore the alignment of 3DCP technology with Construction 5.0 criteria. The focus was on identifying key criteria for Construction 5.0, which broadly encompass the essential elements necessary for driving innovation, sustainability, and collaboration in modern construction practices. These criteria reflect the integration of advanced technologies with human-centric approaches to create a more adaptive and resilient construction ecosystem. This analysis provided the foundational insights necessary for developing the assessment framework.

Thirdly, a content analysis of the selected literature was conducted to identify the key drivers of 3DCP technology for each of the Construction 5.0 criteria. This is essential because the criteria of Construction 5.0 are inherently abstract and require the identification of specific drivers to enable their practical evaluation. These drivers act as measurable and actionable components that link theoretical concepts to real-world applications. By categorising and highlighting these drivers, the analysis provided valuable insights into how 3DCP technology can align with and support the overarching criteria of Construction 5.0. The findings from this analysis served as an essential foundation for the development of the proposed assessment framework.

Finally, an assessment and conceptual framework for implementing 3DCP in alignment with Construction 5.0 criteria was developed, based on identified key drivers and expert opinions from construction engineering and 3DCP specialists.

2.1 Systematic Literature Review

The systematic literature review was conducted to critically examine existing research and identify the core criteria of Construction 5.0, with a particular focus on their relevance to 3DCP technology. A summary of the systematic literature review is presented in Table 1. The four core dimensions of Construction 5.0 identified through this review are:

Human Centricity (HC): This dimension focuses on prioritising human well-being in construction by combining human ingenuity and craftsmanship with the efficiency and consistency of technology [3].

Sustainability (SU): Additive manufacturing, also known as 3D printing, is a key feature of Construction 5.0, playing a significant role in enhancing the sustainability of manufacturing processes. In Construction 5.0, additive manufacturing focuses on improving customer satisfaction by integrating added value into products and services [3].

Resilience (R): The term 'resilience' refers to the need

to enhance the robustness of industrial production, enabling it to better withstand disruptions and maintain essential infrastructure during crises [1].

Cost and Environmental Efficiency (C&EE): In Construction Industry 5.0, "Cost Reduction and Environmental Control" refers to the strategic approach of minimising expenses while managing and reducing environmental impact [11].

Table 1. Systematic literature review

Author name	Articles POS	Construction 5.0 Criteria				
		HC	SU	R	C&EE	Ref
Morteza et al.	Develop a strategic roadmap outlining how Industry 5.0 achieves its sustainable development goals.	✓	✓	✓	-	[12]
Marinelli.	Explore the integration of Industry 4.0 in construction, emphasising management perspectives and the shift towards Construction 5.0	✓	✓	-	-	[13]
Praveen et al	Provides an overview of Industry 5.0, highlighting its enabling technologies, potential applications, and implementation challenges, while emphasising the importance of human-machine collaboration in advancing manufacturing solutions	✓	✓	✓	-	[14]
Huang et al.	Aims to clarify Industry 5.0 and Society 5.0, highlighting their human-centric, sustainable, and resilient visions, and aligning their development to future operations	✓	✓	✓	-	[15]
Sarioğlu	Aims to explain the shift from Industry 4.0 to Industry 5.0 and the creation of Society 5.0, focusing on human-centricity, resilience, cost and environmental efficiency, and sustainability, while introducing Consumer 5.0 to guide companies in updating marketing strategies	✓	✓	✓	✓	[16]
Humayun	Provides an overview of Industry 5.0, including its concept, applications, the role of cutting-edge technologies (CET), and the associated opportunities and challenges	✓	✓	✓	✓	[11]

3 Mapping 3DCP Drivers to Construction 5.0 Criteria

This section outlines how to identify and shortlist the key drivers of 3DCP to meet the criteria of Construction 5.0 criteria using a content analysis approach. A content analysis of more than 20 academic papers was conducted to identify and evaluate the key drivers enabling 3DCP to meet the core criteria of Construction 5.0: Human-Centricity, Sustainability, Resilience, and Cost Reduction & Environmental Control. A total of 10 drivers were identified for each criterion. Each driver was assigned a unique Code (C) to facilitate clear and structured representation throughout the analysis. The Frequency (F) of each driver's occurrence in the literature was calculated using Equation 1 to quantify its prominence [17]. This frequency evaluation identified the prominence of each driver based on its repetition mention in the literature, providing insight into its significance for aligning 3DCP with the principles of Construction 5.0.

$$\mathrm{F} = \sum \text{Driver Occurance in articles} \quad (1)$$

Next, the Significance (S) of each driver was assessed to quantify its relative importance based on its representation in the reviewed literature. Each driver was classified into one of three qualitative categories: High, Medium, or Low. This qualitative categorisation was determined by observing the emphasis placed on the driver across the reviewed studies. The category with the highest mentions was used to assign the driver's overall significance. To transform the qualitative significance into a quantitative measure for subsequent analysis a Quantitative Score (QS) is assigned to each driver (High as 5, Medium as 3, and Low as 1) and the highest frequency impact is selected for each driver following Rasul et al. [18].

$$Ls = \frac{QS \times F}{A \times N} \quad (2)$$

$$NLS = \frac{LS_{Individual}}{\sum LS} \quad (3)$$

$$\mathrm{CS} = CS_{previous\ driver} + CS_{Current\ driver} \quad (4)$$

The literature score (LS) is calculated as outlined in Equation 2, where A is the highest possible (qualitative) score, N is the total no of papers which were considered for the identification of the challenges, frequency shows the repetition of challenges in papers. The next step was to transform this LS into a Normalised Literature Score (NLS). Normalisation is required to ensure that the data directly related to the database is considered. Further, each data field contains only one data element, which removes redundant (unnecessary) data. It is obtained by dividing each driver individual LS by the total LS ($\sum$LS) as stated in Equation 3. After that, the Cumulative Score (CS) was then computed to provide an aggregated measure of importance. For the first driver, CS was equal to its NLS. For subsequent drivers, the CS was calculated using Equation 4. The drivers were ranked based on their LS and NLS, with the rankings providing a systematic representation of their relative importance [17]. This structured methodology ensured transparency and consistency in the identification, evaluation, and prioritisation of the key drivers for aligning 3DCP with the criteria of Construction 5.0, as presented in Table 2

Table 2. Driver's identification and ranking

Drivers / Human-Centricity Drivers	C	F	S	QS	LS	LNS	CS	Rank	Ref
Design Flexibility	H1	13	High	5	0.650	0.213	0.213	1	[6]
Enhance Precision	H2	8	High	5	0.400	0.131	0.344	2	[6]
Job Creation in New Roles	H3	8	High	5	0.400	0.131	0.475	2	[19]
Worker Fatigue Reduction	H4	3	Medium	3	0.090	0.030	0.505	8	[19]
User-Centric Design	H5	8	Medium	3	0.240	0.079	0.583	5	[19]
Worker Empowerment	H6	4	High	5	0.200	0.066	0.649	6	[20]
Smart Building Integration	H7	5	High	5	0.250	0.082	0.731	4	[21]
Community Involvement	H8	5	Medium	3	0.150	0.049	0.780	7	[21]
Human and Machine Interaction	H9	8	High	5	0.400	0.131	0.911	2	[19]
Reduced Human Error	H10	9	Medium	3	0.270	0.089	1.000	3	[20]
					$\sum$LS = 3.05				
Sustainability Drivers									
Material Efficiency/Consumption	S1	11	High	5	0.550	0.122	0.122	3	[22]
Use of Recycled Materials	S2	13	High	5	0.650	0.144	0.266	2	[22]
Energy Efficiency in Construction	S3	7	High	5	0.350	0.078	0.344	6	[22]
Minimised Waste	S4	15	High	5	0.750	0.167	0.511	1	[23]
Reduced Water Usage	S5	5	Medium	3	0.150	0.033	0.544	8	[23]
Eco-friendly Material	S6	8	High	5	0.400	0.089	0.633	5	[24]
Circular Economy Potential	S7	8	Medium	3	0.550	0.122	0.755	3	[24]
Sustainable Supply Chain	S8	6	High	5	0.300	0.067	0.822	7	[24]
Enhance Durability	S9	10	High	5	0.500	0.111	0.933	4	[24]
Improve Thermal Comfort	S10	6	High	5	0.300	0.067	1.000	7	[24]
					$\sum$LS = 4.5				
Resilience Drivers									
Improve Structural Integrity	R1	7	High	5	0.350	0.119	0.119	2	[25]
Reduce Chemical Hazards	R2	5	High	5	0.250	0.085	0.205	4	[26]
Reduce Ergonomic Hazards	R3	5	High	5	0.250	0.085	0.290	4	[26]
Reduce Physical Hazards	R4	6	High	5	0.300	0.102	0.392	3	[26]
Rapid Deployment	R5	6	Medium	3	0.180	0.061	0.454	5	[27]
Improve Construction Site Safety	R6	9	High	5	0.450	0.154	0.608	1	[26]
Promote Modular Construction	R7	6	High	5	0.300	0.102	0.710	3	[27]
Fire Resistance	R8	5	High	5	0.250	0.085	0.795	4	[27]
Resilience to Natural Disasters	R9	6	High	5	0.300	0.102	0.898	3	[27]
Ease of Repair	R10	6	High	5	0.300	0.102	1.000	3	[26]
					$\sum$LS = 2.93				
Reduce Cost and Environmental Control Drivers									

Reduced Labour Cost	CE1	9	High	5	0.450	0.101	0.101	5	[28]
Increased Productivity	CE2	14	High	5	0.700	0.157	0.258	1	[28]
Reduced Construction Time	CE3	10	High	5	0.500	0.112	0.371	4	[28]
Reduce Construction Cost	CE4	12	High	5	0.600	0.135	0.506	3	[28]
Improve Air Quality	CE5	4	High	5	0.200	0.045	0.551	8	[29]
Reduced CO2 footprint	CE6	13	High	5	0.650	0.146	0.697	2	[28]
Reduced Noise Pollution	CE7	4	High	5	0.200	0.045	0.742	8	[27]
Elimination of Formwork Cost	CE8	8	High	5	0.400	0.090	0.831	6	[27]
Reduce Material Cost	CE9	8	High	5	0.400	0.090	0.921	6	[29]
Reduced Site Disturbance	CE10	7	High	5	0.350	0.101	1.000	7	[24]
					$\sum$LS = 4.450				

4 Assessment Framework for 3DCP in Meeting the Criteria of Construction 5.0

The assessment framework for 3DCP offers a structured methodology to evaluate its alignment with the four core criteria of Construction 5.0: Human-Centricity, Sustainability, Resilience, and Cost and Environmental Control. To effectively illustrate the interconnected causal relationships among the drivers and core criteria, a causal path diagram is employed, enhancing the clarity of the framework and facilitating deeper discussions regarding the outlined causal relationships. Previous literature indicates that causal path diagrams serve as effective tools for visualizing complex processes and sequences of events. Additionally, the use of directional arrows, as suggested by Duncan (1966), elucidates the nature of influence among the various components of the framework [30]. By utilising arrows to indicate the direction of causality, the impact of each driver on its associated criterion is clearly demonstrated [31]. This approach ensures a coherent understanding of how 3DCP aligns with the core criteria of Construction 5.0 by systematically depicting the causal pathways between drivers and their associated criteria.

Human-Centricity is a fundamental pillar of Construction 5.0, and the framework addresses this through ten drivers (H1 to H10) that collectively prioritise the well-being, engagement, and empowerment of individuals involved in construction processes. Drivers such as worker fatigue reduction, enhanced precision, and user-centric design underscore the technology's ability to create safer and more inclusive construction environments. The integration of smart technologies and the reduction of human error further emphasise how 3DCP fosters a human-centred approach to construction.

The criterion of Sustainability is addressed by 10 drivers (S1 to S10) that focus on minimising environmental impact and maximising resource efficiency. Drivers such as material optimisation, waste reduction, and the use of recycled materials collectively reduce the environmental footprint of construction activities. Furthermore, the adoption of eco-friendly materials and circular economy principles aligns C3DP with global sustainability goals, ensuring long-term environmental benefits.

Resilience, another critical criterion, is supported by ten drivers (R1 to R10) that enhance the adaptability, durability, and reliability of 3D printed structures. Rapid deployment capabilities improved structural integrity, and resilience to natural disasters are key features that make 3DCP a transformative technology in dynamic and high-risk environments. These drivers collectively ensure that 3DCP contributes to robust and future-proof construction practices.

Finally, Cost and Environmental Control is comprehensively addressed through ten drivers (CE1 to CE10) that focus on reducing costs while maintaining environmental sustainability. Drivers such as reduced labour costs, faster construction timelines, and the elimination of formwork illustrate how 3DCP achieves cost efficiency. Simultaneously, the reduction of carbon emissions, noise pollution, and improvements in indoor environmental quality demonstrate its alignment with environmental control objectives. By addressing both economic and ecological efficiencies, 3DCP aligns with Construction 5.0 emphasis on sustainable and cost-effective practices.

This assessment framework, as depicted in Figure 1 provides a clear and logical structure to evaluate 3DCP technology to meet Construction 5.0 criteria. The causal path diagram serves as a critical tool for visualising these relationships, facilitating a comprehensive understanding of how 3DCP drivers influence the achievement of each criterion. This structured approach not only underscores the transformative potential of 3DCP but also serves as a basis for further research and practical implementation in advancing construction practices.

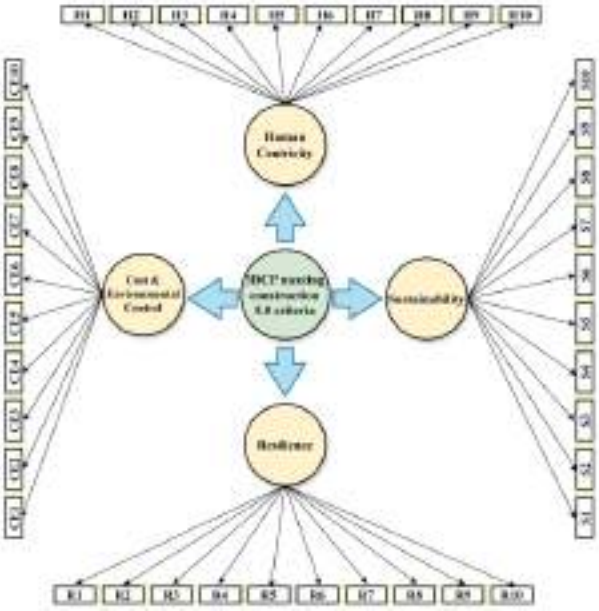


Figure 1. 3DCP assessment framework for Construction 5.0

5 Implementation Framework For 3DCP in Construction Industry

A conceptual framework for implementing a 3DCP in Construction Industry 5.0 was made using an assessment framework and expert opinions of construction engineering and 3DCP specialists, as shown in Figure 2. This conceptual framework comprised three phases: "Knowledge Building Phase", "Implementation phase", and "Evaluation Phase".

5.1 Phase 1: Knowledge Building Phase

The Knowledge Development stage emphasises the need for stakeholder awareness and capacity building through partnerships with professional organisations, academic institutions, and industry bodies. By incorporating 3DCP into training programmes, educational curriculum, and industry case studies, this stage facilitates the dissemination of reliable information and best practices.

5.2 Phase 2: Adoption Phase

The Adoption Phase operationalises the framework through decision making and implementation strategies that address critical barriers, such as high initial costs and a lack of technical expertise. Financial incentives, including subsidies, tax exemptions, and expedited approvals, are positioned as enablers for adoption, and industry stakeholders can use these mechanisms to encourage investment in 3DCP technologies.

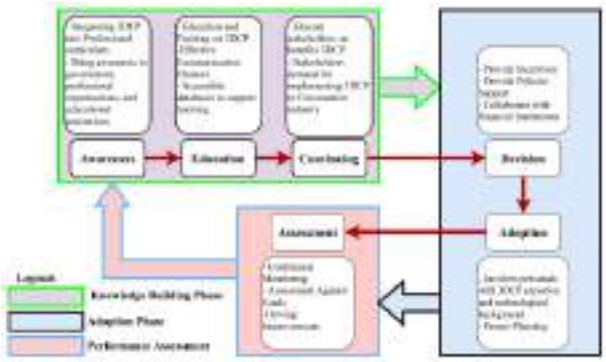


Figure 2. Implementation framework for 3DCP in the construction industry

5.3 Phase 3: Performance Assessment

The Performance Assessment stage provides a mechanism for continuous evaluation and improvement. By monitoring and assessing projects against Construction 5.0 objectives, this stage ensures compliance with human-centric, sustainable, and resilient construction goals.

This framework provides a structured, scalable, and impact-driven approach to integrating 3D concrete printing into Construction 5.0. It enables industry stakeholders to overcome barriers to adoption while delivering sustainable and efficient construction solutions. For academia, it offers a robust foundation for research, collaboration, and knowledge dissemination, advancing the discourse on next-generation construction technologies. By fostering a symbiotic relationship between research and practice, this framework has the potential to catalyse transformative advancements in the built environment, addressing global challenges in sustainability, resilience, human-centric development, and cost and environmental control.

6 Conclusion

This research highlights the transformative role of 3DCP technology in achieving the goals of Construction 5.0. By aligning with the core pillars of Construction 5.0, including human centricity, sustainability, resilience, and cost and environmental control, 3DCP has proven to be a pivotal innovation in modern construction practices. To evaluate the contribution of 3DCP to these objectives, an assessment framework was developed, systematically identifying and prioritising critical drivers across the four dimensions. This framework provided a robust mechanism to quantify the alignment of 3DCP with Construction 5.0 principles, enabling the identification of actionable insights and key performance indicators. By integrating measurable factors, such as design flexibility, material efficiency, and resilience to natural disasters, the

assessment framework establishes a foundation for benchmarking the adoption of 3DCP against Construction 5.0 goals.

In addition to the assessment framework, this research proposed a conceptual implementation framework designed to operationalise the findings and facilitate the practical application of 3DCP within the construction sector. The implementation framework is structured into three phases: knowledge building, adoption, and performance assessment. The knowledge-building phase focuses on raising awareness among stakeholders, promoting capacity development, and establishing partnerships with academic institutions and industry bodies. The adoption phase addresses challenges such as high costs and limited expertise, advocating for strategies like financial incentives and multidisciplinary collaboration. Lastly, the performance assessment phase ensures ongoing evaluation and iterative improvement, aligning implementation efforts with the objectives of Construction 5.0.

Future work will focus on integrating insights from grey literature with empirical validation through surveys and case studies to enhance the practical applicability and robustness of the proposed framework, allowing for a more comprehensive understanding of its real-world implications. These efforts will provide additional insights into the applicability of 3DCP across diverse construction scenarios, ensuring that it remains a transformative technology in the industry. Furthermore, this iterative process will facilitate continuous refinement of the framework, strengthening its role in bridging theoretical principles with practical applications.

References

[1] J. van der Heijden, "Construction 4.0 in a narrow and broad sense: A systematic and comprehensive literature review," *Build. Environ.*, vol. 244, no. August, p. 110788, 2023, doi: 10.1016/j.buildenv.2023.110788.

[2] M. Wang, C. C. Wang, S. Sepasgozar, and S. Zlatanova, "A systematic review of digital technology adoption in off-site construction: Current status and future direction towards industry 4.0," *Buildings*, vol. 10, no. 11, pp. 1–29, 2020, doi: 10.3390/buildings10110204.

[3] M. Marinelli, "From Industry 4.0 to Construction 5.0: Exploring the Path towards Human–Robot Collaboration in Construction," *Systems*, vol. 11, no. 3, pp. 1–23, 2023, doi: 10.3390/systems11030152.

[4] A. E. Ikudayisi, A. P. C. Chan, A. Darko, and Y. M. D. Adedeji, "Integrated practices in the Architecture, Engineering, and Construction industry: Current scope and pathway towards Industry 5.0," *J. Build. Eng.*, vol. 73, no. May, p. 106788, 2023, doi: 10.1016/j.jobe.2023.106788.

[5] F. Craveiro, J. P. Duarte, H. Bartolo, and P. J. Bartolo, "Additive manufacturing as an enabling technology for digital construction: A perspective on Construction 4.0," *Autom. Constr.*, vol. 103, no. March, pp. 251–267, 2019, doi: 10.1016/j.autcon.2019.03.011.

[6] K. Gamage, S. Fawzia, T. Zahra, M. B. F. Teixeira, and N. H. Ramli Sulong, "Advancement in Sustainable 3D Concrete Printing: A Review on Materials, Challenges, and Current Progress in Australia," *Buildings*, vol. 14, no. 2, 2024, doi: 10.3390/buildings14020494.

[7] M. Adaloudis and J. Bonnin Roca, "Sustainability tradeoffs in the adoption of 3D Concrete Printing in the construction industry," *J. Clean. Prod.*, vol. 307, 2021, doi: 10.1016/j.jclepro.2021.127201.

[8] H. Singh *et al.*, "Enhancing structural resilience by using 3D printed complex polymer reinforcement for high damage tolerant structures," *Constr. Build. Mater.*, vol. 425, no. August 2023, 2024, doi: 10.1016/j.conbuildmat.2024.136085.

[9] G. Boschetti, M. Faccio, and I. Granata, "Human-Centered Design for Productivity and Safety in Collaborative Robots Cells: A New Methodological Approach," *Electron.*, vol. 12, no. 1, pp. 1–16, 2023, doi: 10.3390/electronics12010167.

[10] R. Singh *et al.*, "Cloud manufacturing, internet of things-assisted manufacturing and 3D printing technology: Reliable tools for sustainable construction," *Sustain.*, vol. 13, no. 13, 2021, doi: 10.3390/su13137327.

[11] M. Humayun, "Industrial Revolution 5.0 and the Role of Cutting Edge Technologies," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 12, pp. 605–615, 2021, doi: 10.14569/IJACSA.2021.0121276.

[12] M. Ghobakhloo, M. Iranmanesh, M. F. Mubarak, M. Mubarik, A. Rejeb, and M. Nilashi, "Identifying industry 5.0 contributions to sustainable development: A strategy roadmap for delivering sustainability values," *Sustain. Prod. Consum.*, vol. 33, pp. 716–737, 2022, doi: 10.1016/j.spc.2022.08.003.

[13] M. Marinelli, "towards Human – Robot Collaboration in Construction," *Systems*, pp. 1–23, 2023.

[14] P. K. R. Maddikunta *et al.*, "Industry 5.0: A survey on enabling technologies and potential applications," *J. Ind. Inf. Integr.*, vol. 26, no. July

2021, p. 100257, 2022, doi: 10.1016/j.jii.2021.100257.

[15] S. Huang, B. Wang, X. Li, P. Zheng, D. Mourtzis, and L. Wang, "Industry 5.0 and Society 5.0—Comparison, complementation and co-evolution," *J. Manuf. Syst.*, vol. 64, no. June, pp. 424–428, 2022, doi: 10.1016/j.jmsy.2022.07.010.

[16] C. İ. Sarioğlu, "Industry 5.0, Digital Society, and Consumer 5.0," *Handb. Res. Perspect. Soc. Technol. Addict.*, no. July, pp. 11–33, 2023, doi: 10.4018/978-1-6684-8397-8.ch002.

[17] M. Ghufran, K. I. A. Khan, F. Ullah, W. S. Alaloul, and M. A. Musarat, "Key Enablers of Resilient and Sustainable Construction Supply Chains: A Systems Thinking Approach," *Sustain.*, vol. 14, no. 19, 2022, doi: 10.3390/su141911815.

[18] N. Rasul, M. S. A. Malik, B. Bakhtawar, and M. J. Thaheem, "Risk assessment of fast-track projects: a systems-based approach," *Int. J. Constr. Manag.*, vol. 21, no. 11, pp. 1099–1114, 2021, doi: 10.1080/15623599.2019.1602587.

[19] M. A. Hossain, A. Zhumabekova, S. C. Paul, and J. R. Kim, "A review of 3D printing in construction and its impact on the labor market," *Sustain.*, vol. 12, no. 20, pp. 1–21, 2020, doi: 10.3390/su12208492.

[20] H. Rogers, M. Srivastava, M. E. Tsakou, and A. Prashad, "Sustainable supply chain models for 3D printed houses," *Supply Chain Forum*, vol. 00, no. 00, pp. 1–16, 2024, doi: 10.1080/16258312.2023.2298186.

[21] A. K. Al-Tamimi, H. H. Alqamish, A. Khaldoune, H. Alhaidary, and K. Shirvanimoghaddam, "Framework of 3D Concrete Printing Potential and Challenges," *Buildings*, vol. 13, no. 3, 2023, doi: 10.3390/buildings13030827.

[22] S. Bhattacherjee *et al.*, "Sustainable materials for 3D concrete printing," *Cem. Concr. Compos.*, vol. 122, no. December 2020, p. 104156, 2021, doi: 10.1016/j.cemconcomp.2021.104156.

[23] D. M. Hajare and T. S. Gajbhiye, "Additive manufacturing (3D printing): Recent progress on advancement of materials and challenges," *Mater. Today Proc.*, vol. 58, no. 2022, pp. 736–743, 2022, doi: 10.1016/j.matpr.2022.02.391.

[24] M. Fonseca and A. M. Matos, "3D Construction Printing Standing for Sustainability and Circularity: Material-Level Opportunities," *Materials (Basel).*, vol. 16, no. 6, 2023, doi: 10.3390/ma16062458.

[25] H. Singh *et al.*, "Enhancing structural resilience by using 3D printed complex polymer reinforcement for high damage tolerant structures," *Constr. Build. Mater.*, vol. 425, no. April, p. 136085, 2024, doi: 10.1016/j.conbuildmat.2024.136085.

[26] G. A. Filip, F. H. Abanda, and F. A. Fru, "Construction 3D-printing in reducing the incidence of long latency respiratory diseases among construction workers in the UK," *Saf. Extrem. Environ.*, vol. 5, no. 3, pp. 177–197, 2023, doi: 10.1007/s42797-023-00078-4.

[27] M. Bazli, H. Ashrafi, A. Rajabipour, and C. Kutay, "3D printing for remote housing: Benefits and challenges," *Autom. Constr.*, vol. 148, no. January, p. 104772, 2023, doi: 10.1016/j.autcon.2023.104772.

[28] G. H. Ahmed, "A review of '3D concrete printing': Materials and process characterization, economic considerations and environmental sustainability," *J. Build. Eng.*, vol. 66, no. November 2022, 2023, doi: 10.1016/j.jobe.2023.105863.

[29] B. O. Ayegba, K.-J. I. Egbe, A. M. Nazar, M. Huang, and M. A. Hariri-Ardebili, "Resource Efficiency and Thermal Comfort of 3D Printable Concrete Building Envelopes Optimized by Performance Enhancing Insulation: A Numerical Study," *Energies*, vol. 15, no. 3, 2022, doi: 10.3390/en15031069.

[30] S. Besklubova, N. Spicek, R. Y. Zhong, E. Kravchenko, and M. J. Skibniewski, "A framework for the acceptance of 3D printing technology in construction," *Int. J. Appl. Sci. Eng.*, vol. 21, no. 4, 2024, doi: 10.6703/IJASE.202409_21(4).001.

[31] S. Besklubova, M. J. Skibniewski, and X. Zhang, "Factors Affecting 3D Printing Technology Adaptation in Construction," *J. Constr. Eng. Manag.*, vol. 147, no. 5, 2021, doi: 10.1061/(asce)co.1943-7862.0002034.

Mixed Reality-based Digital Twinning of Building Circularity: A Co-Design Approach for Sustainable Buildings

Bo Su[1], Muhammad Fawad[1], Qian Chen[1], and Marek Salamak[2]

[1] Faculty of Applied Sciences, University of British Columbia, Okanagan Campus, Canada
[2] Department of Mechanics and Bridges, Faculty of Civil Engineering, Silesian University of Technology, Poland,
bo.su@ubc.ca, muhammad.fawad@ubc.ca, qian.chen@ubc.ca, marek.salamak@polsl.pl,

Abstract –

Sustainability in construction practices is becoming the need of the day, and the construction sector is getting adoptive to the integration of digital tools to explore the opportunities for enhancing sustainability and circular economy applications, offering significant benefits to both industry and society. To achieve this goal, this article discusses a sustainability framework combining Digital Twin (DT), Mixed Reality (MR), and Life-Cycle Assessment (LCA) to align with circular economy principles in building construction.

A case study of a single-family house in Kelowna, BC, Canada is conducted to demonstrate the potential of this integration for comprehensive LCA of buildings. The LCA analysis of the building is performed using OneClick LCA– an LCA platform. The results of LCA account for the embodied carbon, improved material circularity, life cycle cost efficiency, etc. A DT Dashboard (DTD) of the building's circularity model is also developed, which monitors and optimizes the whole life cycle of the building. The DTD is then deployed to MR hardware – Microsoft HoloLens for enabling onsite circularity analysis of the building. This not only allows for immersive visualization of LCA data but also enhances stakeholders' collaboration and decision-making. This way the study showcases how immersive DT can potentially be a game-changer for sustainable construction and provides an industrially replicable model.

Keywords –
Mixed Reality (MR), Circular Economy (CE), Life-Cycle Assessment (LCA), Digital Twin (DT), Data Fusion.

1 Introduction

Circular Economy (CE) has become prominent in the construction industry as a way of responding to challenges of sustainability through reduced waste and resource extraction. The underlying rationale of this concept is a closed-loop use of materials, which is somewhat contrary to the traditional linear model of "take, make, dispose [1] [2]." On the one hand, it is agreed that using CE tools and practices earlier in the project can reduce the negative environmental impacts constructions can have [1]. On the other hand, it is also highlighted that adopting circular practices during the demolition phase in construction can significantly reduce waste, where most materials are currently discarded due to lack of reuse potential [3]. There are other strategies such as adaptive reuse and selective disassembly that are in line with CE principles and extend the life of building components by maximizing material recovery [4].

Furthermore, a collaboration of CE with Life-Cycle Assessment (LCA) is to be considered while designing an optimum building for sustainability in terms of reuse, recycling, and reduction of any kind of waste in the whole life cycle.

Also, advanced technologies such as Digital Twin (DT) are essential to assist in the use of these strategies, as real-time monitoring and data-driven decision-making capabilities can effectively manage resources. Digital Twin technology empowers several industries to dynamically create replicas of actual running systems. By continuously synchronizing data between the physical and digital worlds, replicas are extremely helpful for performance optimization and predicting upcoming problems [5]. The concept of Digital Twin was first proposed by Michael Grieves in 2002 and was officially named and applied by NASA in 2010 for real-time simulation and monitoring of space missions. It then

developed into a key tool for Industry 4.0 applications [6]. Digital Twin can enhance decision-making and reduce operational inefficiencies by integrating Internet of Things (IoT) sensors and Building Information Modelling (BIM) [7], and can also collect and analyze data continuously, to enable predictive maintenance, optimize the processes, and improve resource management in complex construction projects [6]. Owing to its powerful advantages, Digital Twin is playing a more and more crucial role in the construction industry.

To make digital data more accessible and interactable for stakeholders, Mixed Reality (MR) is another powerful technology to apply in construction projects. MR merges the physical world with the digital one to create an immersive environment. The users can interact with both real and virtual objects at the same time using MR technology. It has shown great potential so far in making the construction processes better by enhancing visualization and interaction with real-time data on-site [8]. MR integrates digital overlays into the physical environment, thus making communication among stakeholders easier and allowing for more intense collaboration and quicker decision-making [9]. For example, Wu et al. (2022) [10] did some research on developing MR-based systems to enhance safety by offering real-time visual warnings about potential hazards on a construction site to workers, thus reducing various accidents and improving safety protocols. In addition, MR can also enhance the construction inspection and monitoring process by allowing inspectors to visualize complex 3D models of hidden building elements and identify discrepancies between design and construction in real-time [11]. However, despite such benefits, the high cost of implementation and user acceptance are the main challenges for more widespread diffusions [9].

The integration of MR, DT, and CE technologies can be transformative in improving sustainability in construction. DT technology has increasingly been recognized as a key tool in the transition of the industry from linear to circular models, where the emphasis is on waste minimization and materials reuse for improved salvage value [12]. MR technologies introduce new ways of engaging stakeholders and building further awareness of the principles underlying CE in construction contexts [13]. The combination of MR and DT enhances real-time decision-making and visualization of the construction process, especially in tasks related to design optimization, life cycle analysis, and demolition planning. This integrated approach enables dynamic monitoring and immersive user interaction [14] [15] [16]. By embedding MR-based DT frameworks into building LCA, the construction sector can better align design and demolition phases with CE objectives, improving efficiency and sustainability throughout a building's lifecycle.

Conclusively, the main goal of this study is the development of an immersive Digital Twin Dashboard (DTD) of the circularity model of a building. This DTD helps to monitor and optimize every step of the entire life cycle of the building and contributes to sustainable construction. Then the DT is deployed into HoloLens, which is Mixed Reality (MR) hardware, for lab-scale testing. Immersive visualization and collaboration will enable stakeholders to make better decisions that improve building performance and sustainability.

2 Description of the case study

The single-family house building is located in Kelowna, British Columbia, Canada, which is designed to be a residential house, is comfortable for the family to stay in. The approximate area of this house is 4,465 square feet, having 5 bedrooms and 4 bathrooms, a spacious living area, a dining area, and a gourmet kitchen. It is situated in the Upper Mission area of Kelowna, featuring scenic views and a desirable residential environment. Kelowna has hot, dry summers and mild winters, and therefore is one of the best cities in terms of outdoor activities to enjoy throughout the year. The location and the size of the house allow the structure to be a representative example for the validation of circular economic principles in residential construction while integrating state-of-the-art technologies such as Digital Twin and Mixed Reality to optimize the life cycle performance. The detailed layout of the building is shown in **Figure 1**.



Figure 1: Plan Views of the House

This case study will go in-depth on the circularity of the building structure and the integration of LCA to analyze the environmental performances of the single-

family house. OneClick LCA, an AI-powered sustainability software platform that helps construction and manufacturing industries calculate, manage, and reduce environmental impacts through automated life-cycle assessments, environmental product declarations, and circularity evaluations [17], is used to gain important insights into the key lifecycle costs, energy efficiencies, and embodied carbon related to the building. These insights include considerations on operational energy consumption, the choice of materials, and possible end-of-life scenarios.

Generic construction materials database – the default database from OneClick LCA – is used for the LCA process [18]. The LCA for this project is conducted following ISO 14040, ISO 14044, and EN 15804 standards, and the TRACI 2.1 method is applied to align with LEED Canada requirements [19] [20]. All Global Warming Potential (GWP) values are calculated using a 100-year time horizon, as required by EN 15804, ISO 21930, and TRACI 2.1 [19] [20]. The functional unit is m^2, and the assessment is performed from cradle to grave, covering all stages from raw material extraction and manufacturing to construction, use, and end-of-life disposal, in accordance with EN 15804 and ISO 14040/44.

The following section will discuss the implementation of the principles of the circular economy in the design and how the LCA model quantifies the environmental impact.

3 Circularity of the building structure and LCA model of case-study

3.1 Introduction to Circularity and LCA

The major objective of circularity in building construction is waste and resource consumption reduction by means of design for material reuse, recycling, and adaptability during its lifetime. This also meets the wider concept of the circular economy: principles of the circular economy support the closed-loop use of materials with limited environmental impacts [1]. LCA is the structured method for assessing the environmental footprint of a building throughout its whole life cycle, from material sourcing to construction, operation, and end-of-life. The following section demonstrates how such principles can be applied in practice to a single-family house by evaluating its environmental performance based on data provided by OneClick LCA.

3.2 Circularity of the Building Structure

The building was designed with circularity principles, using materials that at the end of their life will be either reusable or recyclable. The structure is designed to allow the dismantling and reputation of materials. Materials and finishes, as evidenced shows in the results from OneClick LCA, consume 55% of the embodied carbon; therefore, its thoughtful choice is extremely important to reduce the ecological footprint. Besides, for the sake of adaptation, the building design allowed partial deconstruction with recoverable elements. This flexibility will make components reusable or recyclable both in renovations or demolishing and extend their life span by fully meeting the goals of a circular economy.

3.3 LCA Model

The life cycle assessment of the single-family house involves environmental performance at every stage from material sourcing to end-of-life. Results from OneClick LCA show that a large part of the embodied carbon in the building arises during the production and construction phases. This, therefore, calls for a greater need to employ sustainable sourcing practices and energy-efficient construction techniques. It will also have substantial energy use during the operational phase, another big contributor to its environmental footprint.

The end-of-life phase of the building is designed to maximize recovery for reuse and recycling. A design for disassembly allows the structure to recover materials such as timber, steel, and concrete with high efficiency, reducing waste and conserving virgin resources. The LCA results quantify these strategies and prove that they are effective in reducing the overall environmental impact of the building.

3.4 Integration of Circularity and LCA

Merging circularity principles into an LCA model is one of the effective ways toward sustainability. The strategies under circular design – material reuse, adaptability, and efficient construction practices – reduce the embodied carbon and operational energy demands and hence contribute to improved LCA outcomes directly. OneClick LCA data also presented economic benefits by depicting possible cost savings resulting from material waste reduction, reduced energy usage, and minimized end-of-life disposal costs [21].

This project creates a holistic framework for the evaluation and enhancement of sustainability in residential construction by linking circularity with LCA. Such synergy not only reduced the environmental impact on the building but also proved to be a replicable model to integrate these principles into future projects, balancing functionality with economic efficiency and ecological responsibility. The OneClick LCA dashboard with results is shown in **Figure 2**. It presents the value of equivalent carbon dioxide emissions per square meter ($\mathrm{kg\ CO_2e/m^2}$) for the project is 733, and this falls in the

G rating level as per Canadian standards. In addition, the result of the tool Building Circularity v2 shows that 29.59% of the building materials can be reused, recycled, or repurposed after demolition [22].



Figure 2: OneClick LCA Dashboard with Results and Benchmarking

4 Development of the immersive DT of the circularity model

The development of the immersive DT starts with the creation of the detailed 3D model (LoD 350 BIM model) using the Autodesk Revit environment. To effectively use this BIM model in the 3D game engine either the Industry Foundation Class (IFC) file or the FBX file can be imported from Revit to UNITY 3D. The only problem with using the IFC is that all the native properties of each structural element make the 3D model heavy and inefficient thus some extra manual work will be required to import all the design data of the building. Therefore, direct import using the FBX file format is preferred and recommended in such cases. Upon successful import of the BIM model to the UNITY platform, the DTD is created from scratch that works while switching to the game mode of the 3D model. The DTD workflow is based on the User Interface (UI) modules such as canvases that are used to mesh the 3D model (virtual counterpart of the DT model). These meshing elements enable the automation function of DTD and help to integrate with the physical model of the building. In the case of this research, the physical model is replicated with the online web platform (OneClick LCA) that accounts for the circularity of the building. In order to ensure the two-way communication between the DTD and OneClick LCA a C# code is embedded in the canvas of the DTD which calls upon the circularity model of the building in OneClick LCA once the game mode is switched on in the UNITY 3D.

The calculation of the circularity parameters i.e., Embodied carbon benchmark, Embodied carbon by life-cycle stage, and Embodied carbon by structure are automatically linked to the DTD directly from the OneClick LCA interface. This way the DTD performs the LCA of the building while ensuring the basic principle of circular design such as material reuse, adaptability, and efficient construction practices resultantly reducing the embodied carbon and operational energy demands while contributing to improved LCA outcomes automatically. The DTD with the LCA data is shown in **Figure 3**.

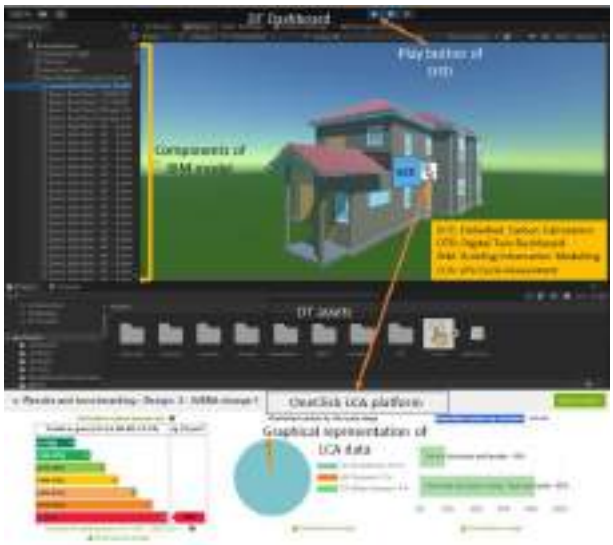


Figure 3: Digital Twin Dashboard of Integrated BIM-LCA system in 3D game engine

5 Deployment of DT to MR hardware and lab-scale testing of the DT-based circularity model

To enhance collaboration between the stakeholders and enable better visualization, the DT model, along with the LCA system, is deployed into Microsoft HoloLens – an MR hardware. This includes embedding the LCA system into the DT model, adjusting the direction and location of the model to better fit the camera, building the application in Unity, and finally deploying the application into HoloLens via Visual Studio.

For starters, the LCA system of the building is added to the DT model by attaching a button with a web link to the wall of the building, so that the user can click on the button and access the OneClick LCA webpage to see the LCA results of the building. Detailed result components will be discussed in the next section.

To deploy the app to HoloLens, the Mixed Reality Tool Kit (MRTK) is used for this project. The main camera of the Mixed Reality Play Space is adjusted to the right location to provide a better view of the building and the LCA button. It is expected that once the users open the app in HoloLens, they can see the whole building in a proper size, and can see the LCA button clearly, guiding them to click on the button to see the LCA results.

As for building the application, the Universal Windows Platform (UWP) in Unity is used for the purpose of developing and running the application on a wide range of Windows devices, including HoloLens, PCs, etc. In this way, the application can be tested not only in HoloLens but also on PCs.

The application is built on the Remote Device (Via Windows Device Portal). HoloLens is connected to the same Wi-Fi network as the PC, so the PC can be paired with HoloLens via Windows Device Portal. The application is built in Unity and then deployed to HoloLens in Visual Studio using the IP address HoloLens connected to, the username and password of the Portal. After deploying, the application can run on HoloLens.

While running the application, the users can see the entire building DT model and the LCA button attached to it. By clicking the button, that embeds the web link, the users can see a webpage jumping out, as shown in **Figure 4**. The webpage shows the LCA results, including graphs and diagrams of Embodied Carbon, Life Cycle Cost, etc.



Figure 4: Immersive DT of the building circularity system

6 Discussion

The LCA results in OneClick LCA mainly explain embodied carbon, life cycle costs, and building circularity. Following is the breakdown of each part.

6.1 Embodied Carbon Results

The analysis provides very important insights into the embodied carbon of the design. Almost all emissions lie within A1-A3 (Materials), and only a small amount lies within A4 (Transport) and C2 (Waste Transport). Those materials comprise almost the entirety of the total carbon footprint in the building project. Horizontal structures such as beams, floors, and roofs contribute the most. The results have shown strong evidence for the consideration of material alternatives in the design stage so as to reduce environmental impact.

Benchmarking also puts the performance of the design in a wider context, such as building types in Canada or the MEP benchmarks. By doing so, one gains the ability to perform a relative comparison, showing areas that are excellent or fall behind in the design. These indeed are valuable insights from where the design team can thereafter work towards making targeted improvements to meet and hopefully even better the benchmarked targets in sustainability.

6.2 Building Circularity

Circularity is important because it reduces waste, saves resources, and supports the closed-loop system for long-term sustainability [1]. The report shows that the Building Circularity is around 29.59%. This means that although there is a considerable portion of materials that can be recycled or renewed at the end of the building's life cycle, most materials may still be linear and would be disposed of in the end. To improve circularity, we changed these materials to recyclable materials, and the LCA results are shown in **Figure 5**. In this case, the Building Circularity is 100%, which means all the materials we use in this building can be reused at the end of the building's life cycle, and therefore the embodied carbon intensity is decreased from 733 kg CO_2 e/m^2 to 229 kg CO_2 e/m^2 .



Figure 5: LCA Results after Changing Materials

6.3 Life Cycle Cost Analysis

The LCC results show a detailed breakdown of the financial impacts across various building elements and life-cycle stages. Construction costs between A0 to A5 come out as the largest cost contributor. These are primarily determined by material and labor costs. The design team may reduce the initial cost through such cost-effective and sustainable methods as pre-fabrication or sourcing materials locally.

The costs of operational energy use B6 are also high, indicating that energy efficiency will be very important

for the reduction of long-term expenses. The end-of-life costs, C1 to C4, are relatively low but can be further reduced by specifying the materials and systems for reuse or recycling, thereby avoiding demolition and waste management costs.

6.4 Combined Analysis of Life-Cycle Stages and Elements

An integrated view of life cycle stages and elements gives an insightful understanding of how the costs evolve over the life of the building. The analysis will enable stakeholders to prioritize interventions based on their life-cycle stage and element-specific impact. For example, it would have immediate cost and environmental benefits if high-impact materials were focused on during the construction phase.

7 Conclusions

The integration ofDT and MR technology with the CE framework is a great improvement in the building sustainability analysis. This study has developed a novel DTD which is further deployed to MR headset and allows for immersive monitoring and optimization of the whole life cycle of the building. It enhances the efficiency of resource management and allows for the potential of collaborations between stakeholders.

The case study of the Kelowna single-family house is considered in this research. LCA of the building is performed by incorporating circularity principles – such as using reusable materials – that significantly reduce embodied carbon emission and improve building sustainability performance. After changing to more eco-friendly materials, the embodied carbon intensity decreased from 733 kg CO_2 e/m^2 to 229 kg CO_2 e/m^2, and the building circularity ratio is increased from 29.59% to 100%.

This study emphasizes the importance of embedding LCA and CE principles early in the design process to mitigate environmental impacts effectively and by integrating cutting-edge technologies such as DT and MR, this analysis process can be improved significantly.

However, there are also limitations and challenges for the integration of MR, DT, and LCA. Construction workers and project teams may need targeted training to develop the skills required to use these tools effectively, which could delay widespread implementation. Furthermore, how well these technologies perform could differ depending on project size and building type. While they may be suitable for straightforward projects, more extensive or highly complex developments could experience challenges related to data integration, system compatibility, and overall performance, which would benefit from further investigation. Future work can focus on how to implement this framework in different construction types, reducing costs and increasing user acceptance to promote wider sustainable practices. In addition, conducting a cost-benefit analysis that compares advanced tools, such as OneClick LCA software, with traditional LCA methods could provide valuable insights. This analysis would help stakeholders make more informed decisions by evaluating both the financial investment required and the potential benefits, such as improved accuracy, time savings, and enhanced sustainability outcomes.

References

[1] G. L. F. Benachio, M. do C. D. Freitas, and S. F. Tavares, “Circular economy in the construction industry: A systematic literature review,” *Journal of Cleaner Production*, vol. 260, p. 121046, Jul. 2020, doi: 10.1016/j.jclepro.2020.121046.

[2] B. Sanchez, C. Rausch, C. Haas, and R. Saari, “A selective disassembly multi-objective optimization approach for adaptive reuse of building components,” *Resources, Conservation and Recycling*, vol. 154, p. 104605, Mar. 2020, doi: 10.1016/j.resconrec.2019.104605.

[3] F. Rasmussen, M. Birkved, and H. Birgisdóttir, “Upcycling and Design for Disassembly – LCA of buildings employing circular design strategies,” *IOP Conference Series: Earth and Environmental Science*, vol. 225, 2019, doi: 10.1088/1755-1315/225/1/012040.

[4] B. Sanchez and C. Haas, “A novel selective disassembly sequence planning method for adaptive reuse of buildings,” *Journal of Cleaner Production*, vol. 183, pp. 998–1010, May 2018, doi: 10.1016/j.jclepro.2018.02.201.

[5] M. Liu, S. Fang, H. Dong, and C. Xu, “Review of digital twin about concepts, technologies, and industrial applications,” *Journal of Manufacturing Systems*, vol. 58, pp. 346–361, Jan. 2021, doi: 10.1016/j.jmsy.2020.06.017.

[6] Y. Pan and L. Zhang, “A BIM-data mining integrated digital twin framework for advanced project management,” *Automation in Construction*, vol. 124, p. 103564, Apr. 2021, doi: 10.1016/j.autcon.2021.103564.

[7] S. Cascone, G. Parisi, and R. Caponetto, “BIM-Based Strategies for the Revitalization and Automated Management of Buildings: A Case Study,” *Sustainability*, vol. 16, no. 16, p. 6720, Aug. 2024, doi: 10.3390/su16166720.

[8] J. Chalhoub and S. K. Ayer, “Using Mixed Reality for electrical construction design communication,” *Automation in Construction*, vol. 86, pp. 1–10, Feb. 2018, doi: 10.1016/j.autcon.2017.10.028.

[9] H. F. Moore and M. Gheisari, "A Review of Virtual and Mixed Reality Applications in Construction Safety Literature," *Safety*, vol. 5, no. 3, p. 51, Aug. 2019, doi: 10.3390/safety5030051.

[10] S. Wu, L. Hou, G. (Kevin) Zhang, and H. Chen, "Real-time mixed reality-based visual warning for construction workforce safety," *Automation in Construction*, vol. 139, p. 104252, Jul. 2022, doi: 10.1016/j.autcon.2022.104252.

[11] R. M. I. Anwar and S. Azhar, "Mixed Reality in Building Construction Inspection and Monitoring: A Systematic Review," *Virtual Worlds*, vol. 3, no. 3, pp. 319–332, Aug. 2024, doi: 10.3390/virtualworlds3030017.

[12] X. Meng, S. Das, and J. Meng, "Integration of Digital Twin and Circular Economy in the Construction Industry," *Sustainability*, vol. 15, no. 17, p. 13186, Sep. 2023, doi: 10.3390/su151713186.

[13] Y.-C. Lee and F. Leite, "Mixed Reality Promoting Circular Economy in Urban Water Systems," in *Computing in Civil Engineering 2023*, Corvallis, Oregon: American Society of Civil Engineers, Jan. 2024, pp. 102–109. doi: 10.1061/9780784485248.013.

[14] S. H. Choi *et al.*, "An integrated mixed reality system for safety-aware human-robot collaboration using deep learning and digital twin generation," *Robotics and Computer-Integrated Manufacturing*, vol. 73, p. 102258, Feb. 2022, doi: 10.1016/j.rcim.2021.102258.

[15] S. Huang, Y. Mao, Z. Peng, and X. Hu, "Mixed reality-based digital twin implementation approach for flexible manufacturing system design," Jan. 24, 2022, *In Review*. doi: 10.21203/rs.3.rs-1265753/v1.

[16] M. Fawad *et al.*, "Development of immersive bridge digital twin platform to facilitate bridge damage assessment and asset model updates," *Computers in Industry*, vol. 164, p. 104189, Jan. 2025, doi: 10.1016/j.compind.2024.104189.

[17] M. Fawad et al., "Integration of Bridge Health Monitoring System With Augmented Reality Application Developed Using 3D Game Engine–Case Study," in IEEE Access, vol. 12, pp. 16963-16974, 2024, doi: 10.1109/ACCESS.2024.3358843.

[18] "OneClickLCA - About Us." Accessed: Mar. 01, 2025. [Online]. Available: https://oneclicklca.com/company/about-us

[19] "Global LCA Database – One Click LCA." Accessed: Mar. 02, 2025. [Online]. Available: https://oneclicklca.com/why-us/capabilities/global-lca-data

[20] "One Click LCA," SolidCAD. Accessed: Mar. 02, 2025. [Online]. Available: https://www.solidcad.ca/one-click-lca/

[21] "One Click LCA EPD Generator | Trusted by leading manufacturers." Accessed: Mar. 02, 2025. [Online]. Available: https://oneclicklca.com/en-us/software/manufacturing/epd-generator

[22] R. Minunno, T. O'Grady, G. M. Morrison, R. L. Gruner, and M. Colling, "Strategies for Applying the Circular Economy to Prefabricated Buildings," *Buildings*, vol. 8, no. 9, Art. no. 9, Sep. 2018, doi: 10.3390/buildings8090125.

[23] "Building Circularity V2," One Click LCA Help Centre. Accessed: Mar. 02, 2025. [Online]. Available: https://oneclicklca.zendesk.com/hc/en-us/articles/15602310529820-Building-Circularity-V2

Knowledge-Based Systems in the Era of Large Language Models: A Case Study in Fire Safety Management

Dilan Durmus [1], Shabtai Isaac[2], Alessandro Carbonari[1], Alberto Giretti[1]

[1] Polytechnic University of Marche, DICEA Department, Ancona, Italy

[2] Ben-Gurion University of the Negev, Department of Civil and Environmental Engineering, Beer-Sheva, Israel

d.durmus@pm.univpm.it,isaacsh@bgu.ac.il, alessandro.carbonari@staff.univpm.it, a.giretti@staff.univpm.it

Abstract –

The construction industry faces complex decision-making challenges that require the integration of extensive domain knowledge and data analysis. Traditional methods are inadequate to effectively address these challenges due to the dynamic nature of construction projects. In the fast-evolving landscape of construction technology, artificial intelligence (AI) plays a crucial role in enhancing decision-making and prediction processes. Therefore, the main purpose of this research is to examine how AI, particularly Large Language Models (LLMs), can be integrated in the analysis and management of emergency scenarios to improve decision-making, prediction, and optimization through knowledge-based systems. This research follows the CommonKADS method for knowledge engineering and embeds sub-symbolic and symbolic AI techniques. By integrating LLMs, this research conceptualizes and operationalizes tacit construction knowledge into structured knowledge systems. This approach has been applied in a case study analysing National Fire Protection Association (NFPA) fire incidents to test possible interactions to extract related data for the knowledge-based systems. The findings inspire future research on the scalability of AI-integrated systems across different segments of the construction industry, their potential in regulatory compliance scenarios, and the development of automated knowledge-based systems in construction industry.

Keywords –

Knowledge-based systems; Large Language Models; Fire Safety Management; Artificial Intelligence; Construction Industry

1 Introduction

The construction industry plays an important role in shaping the built environment but is facing challenges arising from the complexity and dynamic nature of its processes. These challenges are multifaceted, involving the coordination of multiple stakeholders, adapting to changing environmental conditions and complying with complex regulatory requirements especially in case of emergency management.

Traditional decision-making tools are often insufficient to meet the demands, as they struggle to proceed with large volumes of data and fail to pace up with rapid changes on construction projects. For example, Zhu et al. identifies challenges facing decision support systems in construction in the following aspects: impreciseness or uncertainty of information, dynamic and unstable situations, high volume, high velocity and high variety of the information.[1]

Algorithmic approaches to decision support show systemic limitations that push away the possibility of achieving optimal solutions for effective real time management in the era of big data. This creates a need for innovative approaches that are able to cope with result accuracy, that are computational, efficient, responsive and effective in the fast-paced world of the construction industry.

Recent developments in artificial intelligence (AI) technologies have produced significant advances to knowledge technologies, bringing renewed attention to the role of knowledge in solving complex problems. The introduction of knowledge-based systems (i.e. expert systems) in the 1980s marked a significant milestone in addressing the automation of knowledge-based problem solving. These systems used deep domain knowledge and heuristic reasoning to tackle complex, real-world problems. Their declarative nature allowed reasoning under uncertainty and provided explainable rational for their decisions [2], [3].

However, knowledge-based systems are not without limitations, such as the labor-intensive process of knowledge acquisition, scalability challenges, and an inability to provide creative, adaptive and context-specific responses[2]. Moreover, the cost of developing and maintaining such systems remains a significant barrier.

Knowledge engineering (KE) emerged to address these challenges by systematically structuring and

organizing knowledge. KE methodologies, such as CommonKADS [4], provide robust frameworks for modeling tasks, agents, and knowledge systems. These frameworks emphasize conceptualization, operationalization, and formulation, in other words, knowledge coding, which ensure the systematic transformation of domain knowledge into structured, machine-readable and reusable systems [4]. While conceptualization involves identifying and organizing domain knowledge; operationalization translates abstract concepts into actionable procedures; and formalization encodes these procedures into formal representations that can be implemented in computational rule-based systems[4], [5]. As an example, CLIPS (C-Language Integrated Production System), a rule-based tool, integrates AI outputs into operational systems by providing components like a fact-list, knowledge base, and inference engine, enabling the application of domain-specific rules as per CommonKADS methodology [6].

Despite advances, traditional KE faces obstacles in integrating tacit knowledge (informal, experiential knowledge) into structured formats. This includes the significant resources required for knowledge extraction and representation, making it labor-intensive and costly. Additionally, adapting existing models to new circumstances is challenging, as traditional methodologies often lack the flexibility needed to accommodate evolving knowledge and dynamic environments effectively. The rise of artificial intelligence (AI), in particular the emergence of Large Language Models (LLMs), offers transformative opportunities to overcome these limitations.

LLMs such as GPT-4 are adept at processing unstructured data, identifying patterns, and generating human-like responses [7], [8], thereby operationalizing tacit knowledge and transforming natural language into structured representations. These capabilities, when combined with methodologies like CommonKADS, appears as a promising approach to automate and scale KE processes, creating adaptive, dynamic systems that bridge the gap between symbolic and sub-symbolic AI.

However, LLM systems show some severe limitations, due to their black-box architecture and statistical nature of the computing, which are ill-suited to the requirements of reliability and traceability of reasoning. A reliable and explainable reasoning process is in fact required in construction decision-making, because the responsibilities assumed by the decision maker are a determining factor in the decision-making process. Hence hallucination and the lack of explainable reasoning are a barrier to the straightforward application of AI in decision-making [9], [10], [11].

The idea claimed in this paper is therefore to use AI to improve the fragility of domain models and the high cost of building the knowledge base of symbolic systems on the one hand, and to define a procedure to mitigate the hallucination and explainability problems of these systems on the other.

Building upon our previous article presented at ISARC 2024 [12], where we introduced the AI-in-the-loop model as a novel approach for the construction industry, this paper serves as a direct follow-up. While the previous one is focusing on detecting and addressing the issue of hallucinations in AI outputs, this study further explores the integration of AI, particularly LLMs, into established knowledge engineering frameworks dynamically to mitigate such hallucinations more effectively.

Technically, this study aims to structure and operationalize tacit construction knowledge into dynamic, scalable systems by following the guidelines of CommonKADS methodology. The case study is centered on National Fire Protection Association (NFPA) fire incidents because fire safety management, a critical area in construction, provides a rich testing ground for these technologies, requiring fast, responsible decision-making under uncertainty [12], [13]. The research demonstrates how knowledge modelling strategy is done by an expert manually and then how LLMs could contribute to automate or speed up some steps by replacing the manual work done by the domain expert by extracting, structuring, and categorizing the elements in a ontological framework for improved safety management decisions.

This combines structured methodologies with the adaptive capabilities of LLMs to set the foundation for future research into scalable and reliable AI-driven systems across the construction industry.

The following sections will look into the integration of Large Language Models in knowledge-based systems. Section 2 will elaborate on the essential forms of knowledge and knowledge modelling, Section 3 will explain the integration of LLMs into KE. This is followed by a case study on fire safety management in Section 4 as a practical application of LLMs into KE processes. This paper will conclude with a short discussion.

2 Knowledge Modelling

Knowledge is a multi-dimensional asset that is critical to human reasoning and decision-making. In the field of knowledge engineering (KE), it is essential to understand and benefit from forms of knowledge to develop systems that simulate expert problem solving. Knowledge is commonly classified into explicit and tacit types, as well as declarative and procedural forms, each presenting unique opportunities and challenges for formalization and utilization in computational systems [4], [14].

Explicit knowledge refers to knowledge that can be

expressed, codified, and shared, such as manuals, standards, and databases. It is relatively straightforward to capture and encode in knowledge-based systems (KBS) using tools like ontologies and knowledge graphs [5], [4]. In contrast, *tacit knowledge* is personal, context-specific, and often unspoken. This includes skills, intuitions, and experiences accumulated over time. This type of knowledge is more challenging to formalize, requiring advanced methodologies and tools for extraction and representation.

Declarative knowledge, or "know-what," captures phenomenological information and relationships, often suitable for logical reasoning and explicit modelling. *Procedural knowledge*, or "know-how," involves the steps or methods for performing tasks, making it valuable in systems that simulate decision-making processes [2], [4].

Methodologies such as CommonKADS provide structured frameworks for transforming tacit and procedural knowledge into explicit, actionable formats. This process consists of three phases: conceptualization (identifying and organizing knowledge), operationalization (structuring knowledge into workflows or rules), and formulation (formalizing it into computational representations) [4], [5].

Knowledge modelling is a cornerstone of knowledge engineering (KE), serving as the process by which knowledge is identified, structured, and formalized into models that are usable by computational systems. The strategy for knowledge modelling is essential for enabling effective decision-making, reasoning, and problem-solving in knowledge-based systems. At its core, knowledge modelling aims to bridge the gap between human expertise and computational implementation by systematically capturing and representing the subtleties of domain knowledge.

Knowledge models are abstract representations of the knowledge embedded in a domain. These models provide a structured framework to capture relationships, dependencies, and tasks, ensuring that knowledge is both understandable and applicable. The CommonKADS methodology outlines a series of models that captures the organizational, task, and agent aspects of a system, including a specialized knowledge model for domain-specific information and reasoning processes [4].

Following the CommonKADS approach, in this research, the knowledge base will be created in three stages:

1. **Conceptualization:** The initial phase involves identifying domain concepts and relationships. This stage captures both tacit and explicit knowledge through interviews, reports, and observations with domain experts. The conceptualization process prepares the ground for formal representation by structuring informal knowledge into a coherent mental model [4].
2. **Operationalization:** After conceptualization, the abstract model is translated into actionable workflows, algorithms, or rules that the system can execute. This step allows the system to mimic expert reasoning and decision-making processes by including input-output relationships, logical dependencies and procedural workflows [2], [5].
3. **Formalization:** The formalization phase is a straightforward translation of the operational knowledge arrangement into the language of the inference engine. This phase involves encoding the operationalized knowledge into formal computational languages, for example codes, or XML representations, such as ontologies, rule-based systems, or knowledge graphs. The ontological alignment can be considered as a complementary and optional phase linking the conceptualization performed with the ISO standards. This aims to enable reasoning and inference while ensuring consistency [4], [5].

However, it should be noted that one of the critical challenges in knowledge modelling is the representation and analysis of tacit knowledge. Experiential and context-specific tacit knowledge is difficult to formalize and integrate into structured systems. This requires intensive and strong expert knowledge, adaptability and serious time dedication.

3 Integrating Large Language Models into Knowledge Engineering

The integration of Large Language Models (LLMs) into knowledge engineering (KE) indicates a transformative shift in the field. LLMs, like GPT-4, are general-purpose technologies that highly useful in processing unstructured natural language, transforming it into formal, structured knowledge representations. This capability contributes one of the most challenging aspects of KE: the acquisition and formalization of tacit and unstructured knowledge. This ability allows them to formalize that tacit knowledge into explicit forms. Their ability to pre-process data complements traditional Symbolic AI, which relies on structured methods such as logical rules. While symbolic AI provides precision and explainability, LLMs contribute with adaptability and scalability.

The integration of LLMs into KE process is achieved by using the logical-linguistic capabilities of LLMs in each step of the CommonKADS protocol as a tool

available to the knowledge engineer to speed up the work and have more control over the soundness and completeness of the elicitation process. The objective of the study is therefore to qualify the various steps of KE processes in terms of the structure and quality of the input knowledge and of the output knowledge. Furthermore, for each step the knowledge engineer's interaction protocols with the LLM should be defined. This interaction, in turn, will be studied through a prompt engineering process examining diverse prompt techniques such as few-shot prompting or chain-of-thought-prompting [15].

This research contribution presents a phase of development the AI-in-the-loop approach that was presented in Durmus et al. [12], demonstrated the integration of LLMs such as GPT-4 in the development of decision support systems in fire safety planning. These developments benefit from the capabilities of AI to adaptively manage and integrate domain knowledge in dynamic building environments, thus improving operational reliability and decision accuracy.

In the following section annotated examples of the main steps will be provided. The proposed results are the outcomes of a work in progress.

4 The Use of LLMs in the Practice of Knowledge Engineering: Case Study (NFPA)

In this section, we will exemplify the proposed methodology using a case study from the NFPA Case Study: Nightclub Fires report [16]. This manual contains a series of reports of catastrophic fire events described by domain experts. These reports represent a good approximation of what an interview with a domain expert might be. The selection of the NFPA nightclub fire incidents, in particular, was based on the availability of concise and comprehensive documentation that fits well with the LLM's ability to process and analyze texts effectively in terms of text length and complexity. The discussion will be limited to one case study and to the first two phases of the KE process as they share a conceptual level of analysis. The formalization phase, which, on the contrary, involves syntactic aspects, requires specific attention and will be addressed in future works of this research.

As a case study for this research, a scenario from NFPA nightclub fire reports has been selected (Figure 1). Each step for the both phases was firstly done by an expert by manual reading and extracting, then we submitted the same fire event narrative to GPT-4 to do same task and assess its responses.

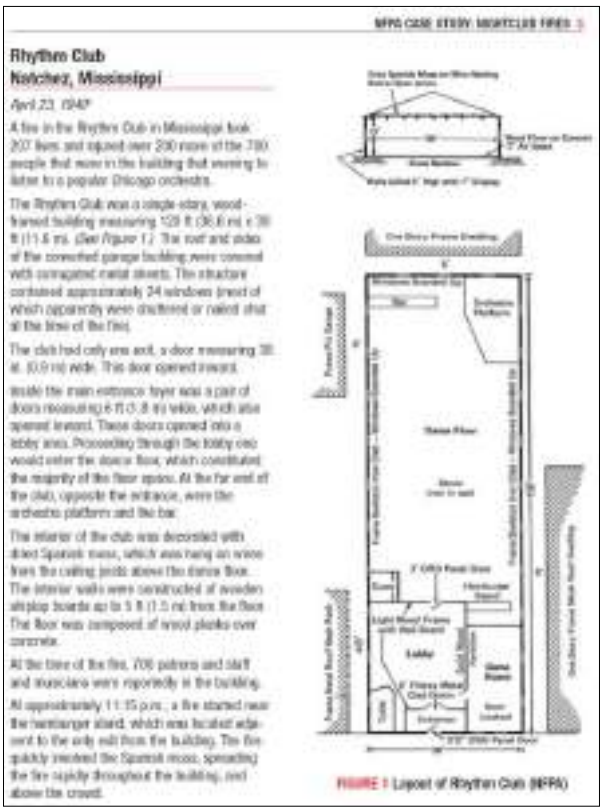


Figure 1: Selected case study from NFPA report

4.1 Conceptualization

The conceptualization phase is the foundation of knowledge engineering and focuses on understanding the domain and structuring knowledge in an informal, human-readable format. This phase increases the clarity and accessibility of domain knowledge through natural language to effectively differentiate between static and dynamic aspects of fire events. This type of structuring helps facilitate the detailed analysis of complex scenarios such as a fire emergency. The goal at this stage is to identify and categorize the critical elements of the scenario and identify key issues. In practice, as shown in the Rhythm Club fire case study, information is organized into two structured categories: context arguments (Table 1) and event arguments (Table 2). This categorization abstracts static building characteristics and dynamic aspects of the fire event and thus provides a framework for systematic analysis. Below, a detailed breakdown presents a clear comparison between the insights extracted by a human expert and GPT-4 and demonstrates the practical application and effectiveness of this approach in real-world scenarios.

a) *Context Arguments*: capture the static, descriptive aspects of the scenario, such as the building's structure, layout, and materials.

b) *Event Arguments:* describe the dynamic elements of the fire incident, including the sequence of events, fire propagation, and occupant behaviour during the emergency.

Table 1: Comparison of the context arguments extracted by human expert vs GPT-4

a) Context Arguments	
HUMAN EXPERT	**GPT-4**
	Prompt: *please extract the arguments in following groups: 1.context arguments, 2. event arguments. List the arguments in bullet points.*
The Rhythm Club was a single-story, wood-framed building measuring 120 ft (36.6 m) x 38 ft (11.6 m).	The Rhythm Club was a single-story, wood-framed building measuring 120 ft by 38 ft.
The structure contained approximately 24 windows (most of which apparently were shuttered or nailed shut at the time of the fire).	It had approximately 24 windows, most of which were shuttered or nailed shut at the time of the fire.
The club had only one exit, a door measuring 38 in. (0.9 m) wide.	There was only one exit from the club, a door measuring 38 inches wide, which opened inward.
This door opened inward.	
Inside the main entrance foyer was a pair of doors measuring 6 ft (1.8 m) wide, which also opened inward.	Inside the main entrance foyer, a pair of doors measuring 6 ft wide also opened inward, leading into a lobby area.
These doors opened into a lobby area.	
The floor was composed of wood planks over concrete.	The floor was composed of wood planks over concrete.
(the list goes on…)	*(the list goes on…)**

*In this phase of context argument extraction, domain experts extracted total 12 arguments and ChatGPT-4 extracted 10 arguments. The accuracy of these results was evaluated using the F1 scoring method [17], and the outcome was 0.90, indicating a highly successful performance.

Table 2: Comparison of the event arguments extracted by human expert vs GPT-4

b) Event Arguments	
At the time of the fire, 700 patrons and staff and musicians were reportedly in the building.	On the night of the fire, 700 patrons and staff, including musicians, were in the building.
At approximately 11:15 p.m., a fire started near the hamburger stand, which was located adjacent to the only exit from the building.	The fire started near a hamburger stand located adjacent to the only exit from the building at approximately 11:15 p.m.
The fire quickly involved the Spanish moss, spreading the fire rapidly throughout the building, and above the crowd.	The fire quickly involved the Spanish moss, spreading rapidly throughout the building and above the crowd.
Patrons in the vicinity of the entrance were able to escape.	Patrons near the entrance managed to escape, but those beyond the lobby area were trapped by the fire.
(the list goes on…)	*(the list goes on…)**

*In this phase of event argument extraction, domain experts extracted total 10 arguments and ChatGPT-4 extracted 9 arguments. The accuracy of these results was evaluated using the F1 scoring method, and the outcome was 0.94, indicating a highly successful performance.

4.2 Operationalization

The operationalization phase focuses on converting conceptual knowledge into structured elements, such as triples, to form a preliminary event ontology structure. This structure includes annotated classes, objects, and relationships in order to build a functional ontology suitable to solve problems.

This phase is conceptually complex because the expert, in order to formulate effective problem-solving strategies by means of rule sets, uses a significant amount of general domain knowledge and introduces considerations and conceptualizations that are not present in the description of the initial event.

4.2.1. Extracting Triples and Ontology Elements

By using previously extracted arguments and the 2D layout of the nightclub, triples were extracted to represent objects, relationships, and events.

Table 3: Comparison of the triples extracted by human expert vs GPT-4

HUMAN EXPERT	**GPT-4**
a) Triples From Arguments	
	Prompt: *From these arguments, please extract triples in subject predicate object form. examples: 1. (Rhythm_Club instance-of SINGLE_STORY_BUILDING) 2. (rc_windows instance-of SET) 3. (rc_exit instance-of DOOR) 4.(rc_entrance_foyer instance-of SPACE) 5. (Rhythm_Club structure wood-framed)*
(*Rhythm_Club* **instance-of** SINGLE_STORY_BUILDING)	(Rhythm_Club instance-of SINGLE_STORY_BUILDING)
(*rc_windows* **instance-of** SET)	(rc_roof_cover material corrugated_metal_sheets)
(rc_entrance_foyer **instance-of** SPACE)	(rc_walls_cover material corrugated_metal_sheets)
(rc_windows **instance-of** WINDOW)	(rc_windows instance-of SET)
(rc_windows **number-of** 24)	(rc_windows quantity 24)
(rc_windows ***status*** shuttered)	(rc_windows status shuttered)
(rc_windows ***status*** nailed-shut)	(rc_windows status nailed_shut)
(rc_exit **instance-of** DOOR)	(rc_exit instance-of DOOR)
(List goes on…)	*(List goes on…)**

*In this phase of triples extraction from the previous arguments, domain experts extracted 45 related triples and ChatGPT-4 extracted 23 triples. The accuracy of these results was evaluated using the F1 scoring method and the outcome was 0.53, indicating a satisfactory performance.

Table 4: Comparison of the triples extracted from 2D layout by human expert vs GPT-4

b) Triples From 2D Layout (Figure 1)	
	Prompt: *From the given layout, extract structured data in the form of triples (subject, predicate, object). Examples:* *1. (rc_bar adjacent rc_orchestra_platform)* *2. (rc_centre instance-of POSITION)* *3. (rc_dance_floor is-at rc_centre)*
(rc_bar **adjacent** rc_orchestra_platform)	(rc_bar adjacent rc_orchestra_platform)
(rc_centre **instance-of** POSITION)	(rc_lobby_area instance-of SPACE)
(rc_dance_floor **is-at** rc_centre)	(rc_dance_floor is-at rc_centre)
(rc_entrance **lead-to** lobby)	(rc_entrance lead-to rc_lobby_area)
(rc_coats_room is-near rc_entrance)	(rc_coat_area is-close-to rc_entrance)
(rc_metal_clad_doors instance-of DOOR)	(rc_entrance instance-of DOOR)
(rc_stove is-at dance_floor_centre)	(rc_dance_floor contains stove)
(List goes on…)	*(List goes on…)**

*In this phase of triples extraction from the 2D layout image, domain experts extracted 32 related triples and ChatGPT-4 extracted 21 triples. The accuracy of these results was evaluated using the F1 scoring method, the outcome was 0.67, indicating a satisfactory performance.

Table 5: Comparison of the objects and classes from the previously extracted triples by human expert vs GPT-4

HUMAN EXPERT	**GPT-4**
c) Objects and Classes	
	Prompt: *From the triples you extracted above, extract 1- classes and objects, 2- relationships as following example:* *1-Objects and Classes* *BUILDING TYPE: single-story-building* *COMPONENT: Windows, Door, Floor, Ceiling joists,....* *2- Relations* *above, access, adjacent, component, connect, connected-to,...*
BUILDING TYPE: single-story-building COMPONENT: Windows, Door, Floor, Ceiling joists, Interior walls, Wooden shiplap boards, Layout, Stove MATERIAL: Wood planks, Concrete, Corrugated metal sheets, Metal clad, Spanish moss STATUS: Flimsy, Dried, Trapped SPACE: Entrance, Foyer, Dance floor, Lobby, Bar, Hamburger stand, Orchestra platform, Exit, Coat area, Game room, Toilet POSITION: corner, center OCCUPANT: Patrons, Staff, Musicians	BUILDING TYPE: Single-story-building COMPONENT: Windows, Door, Floor, Ceiling joists, Stove, Lobby area doors MATERIAL: Wood planks, Concrete, Corrugated metal sheets SPACE: Dance floor, Lobby area, Entrance, Coat area, Game room, Toilet, Orchestra platform, Bar POSITION: Centre, Far end of building, Front FURNISHING: Stove (not in use), Doors (flimsy metal clad) STATUS: Not in use, Locked, Boarded up*
d) Relations	
above, access, adjacent, component, connect, connected-to, decorated, dimension, dimension-wide, hung-on, instance-of, is-a, is-at, is-close-to, is-isolated-from, is-near, is-next-to, is-opposite, layed-over, lead-to, made-of, number-of, open, part-of, status, structure	adjacent, is-at, lead-to, is-close-to, contains, located-in, located-at, number-of, dimension-wide, type, status**

*In this phase of extraction of the ontological elements from previous triples, domain experts extracted 6 classes and ChatGPT-4 extracted 7 categories. F1 score is resulted 0,87. **For the extraction of relations, domain expert listed 26 relations while GPT listed 11, this resulted as 0,59 in F1 scoring.

4.2.2. Developing System Goals

The goals of the system are derived from the Rhythm Club case study by analyzing the emergency dynamics, supported by OSHA (Occupational Safety and Health Administration) guidelines for fire safety. These goals are designed to guide the system's monitoring, planning, and action modules effectively. For example, the catastrophic consequences of a single exit, the rapid spread of fire through flammable materials, and the need for coordinated evacuation directly informed Goal 01 (monitoring safety data) and Goal 05 (leading occupants to safety). By breaking down the emergency into manageable objectives, the goals align the system's actions with real-world needs and ensure adaptive, effective responses. Goals represent strategies to minimize the impact of emergencies. Derived from case dynamics and aligned with OSHA guidelines, these rules include:

1. **Goal 01:** Continuously collect fire-safety information.
2. **Goal 02:** Collect fire-emergency information.
3. **Goal 03:** Alert rescue teams, occupants, and fire departments.
4. **Goal 04:** Mitigate fire propagation through suppression and isolation.
5. **Goal 05:** Lead occupants out of the building quickly.
6. **Goal 06:** Ensure occupant safety during evacuation.

4.2.3. Monitoring Rules to achieve defined goals

Monitoring involves tracking system states using sensor data and triggering actions based on significant changes. These rules interact with the defined goals by ensuring real-time system responsiveness. This interaction ensures that the system adapts dynamically to evolving conditions during emergencies.

The use of the LLM allows the rule-setting phase to be approached in an iterative form, in which the context is examined as a whole, and all relevant aspects are identified and examined and investigated in parallel. At this stage, it is essential to control the semantics of the language in a precise manner. A pseudo language based on triples was used for this purpose. The LLMs was asked to: "write a rule set that monitor the distribution of the fire on a fire safety management system, by asserting the triple: (fire <position> <intensity>), concerning the rhythm_club dance floor" using the whole set of triples defined so far in the conceptualization phase as the query context." LLMs was required to provide the result in: "if <conditions> the <actions> format using triples to express statements"

The LLM provided a structured outcome, tagging each rule with a statement about the issue of the rule is concerned on. The following set of issues has been identified by GPT:

1. Fire Detection at Specific Position
2. Fire Intensity Threshold
3. Fire Spreading to Opposite Spaces
4. …

…

11. Fire Alert for Boarded Windows
12. Decorative Material Hazard

A partial example of the rule set extracted by GPT follows:

1. **Fire Detection at Specific Position**
   If (fire <position> <intensity>) and (rc_dance_floor is-at <position>).
   Then (alert fire_detected_at rc_dance_floor).
2. **Fire Intensity Threshold**
   If (fire rc_dance_floor <intensity>) and (threshold exceeded_by <intensity> 5).
   Then (alert high_fire_intensity rc_dance_floor) and (dispatch fire_team to rc_dance_floor).
3. **Fire Spreading to Adjacent Spaces**
   If (fire rc_dance_floor <intensity>) and (threshold exceeded_by <intensity> 3).
4. …

…

11. **Fire Alert for Boarded Windows**
   If (fire rc_dance_floor <intensity>) and (rc_windows status boarded_up).
   Then (alert ventilation_restricted) and (alert escape_routes_blocked).
12. **Decorative Material Hazard**
   If (fire rc_dance_floor <intensity>) and (threshold exceeded_by <intensity> 2).
   Then (alert high_flammability_material rc_dried_spanish_moss).

5 Results and Discussion

The integration of GPT-4 into the knowledge engineering process revealed both its advantages and limitations. F1 scoring has applied to quantitatively assess GPT-4's performance, with experts manually extracting lists from the same texts queried to the model. These expert-curated lists as presented on the tables served as ground truth, against which GPT-4's outputs were compared. In the conceptualization phase, GPT-4 demonstrated high accuracy in extracting context and event arguments, achieving F1 scores of 0.90 and 0.94, respectively. However, its ability to identify nuanced relationships remained limited, emphasizing the importance of domain expert involvement. During operationalization, GPT-4 efficiently extracted triples and ontology elements, but its performance was moderate in capturing complex relationships, with F1 scores of 0.53 for narrative-based triples and 0.67 for layout-based triples. The tested capabilities showed a potential for automating rules development and was close in content, structure, and case specificity, which indicated alignment with what the human experts could produce. Furthermore, this reveals that GPT-4 provides contextual understanding for generating contextually logical and relevant rules from the scenarios representing complex events such as a fire emergency. Moreover, this is achieved without domain-specific, pre-trained embeddings, matching expert-made guidelines of OSHA, demonstrating strong generalization capabilities of the LLM. Despite these strengths, LLMs' reliance on probabilistic methods remain as a limitation, requiring expert oversight to address gaps; this points back AI-in-the-loop model (Durmus et al., 2024) for a hybrid human-ai collaboration for such complex tasks as fire safety management.

6 Conclusion

This study explores the integration of Large Language Models (LLMs) into the knowledge engineering (KE) process, focusing on their application on fire safety management through a hybrid approach that combines human expertise with LLM capabilities. It presents an innovative methodology for modelling both formal and tacit knowledge for decision-making purposes, demonstrating a promising and original approach to integrating structured methodologies like CommonKADS, with the adaptive and scalable potential of LLMs.

The case study on NFPA fire incidents highlighted GPT-4's potential to automate labor-intensive tasks, such as data extraction and rule generation, while complementing expert knowledge. The outcomes are twofold: first, it makes the knowledge engineering

process more scalable by providing a framework for structuring and operationalizing large volumes of tacit knowledge in a short time frame; second, it offers a tool to address and adapt to unpredictable aspects, such as emergent patterns in dynamic scenarios. While GPT-4 shows strong potential to contribute many steps of the process, its reliance on probabilistic reasoning and LLMs' lack of explainability imply it cannot yet automate the entire workflow. Expert oversight remains crucial to ensure the accountability and reliability of outputs, particularly in high-risk applications like fire safety management.

Future research should focus on automating the formalization phase and its practical applications, includes applying the methodologies and rules developed in this work using tools such as CLIPS to validate the practical implementation of LLMs into hybrid knowledge engineering systems. Expanding case studies across different construction scenarios would evaluate the scalability and generalizability of AI-integrated knowledge systems. These efforts aim to develop robust and efficient decision-support systems for complex and dynamic environments for construction industry.

7 References

[1] X. Zhu, X. Meng, and M. Zhang, 'Application of multiple criteria decision making methods in construction: a systematic literature review', *Journal of Civil Engineering and Management*, vol. 27, no. 6, pp. 372–403, Jul. 2021, doi: 10.3846/jcem.2021.15260.

[2] D. Ashlock, 'Modeling Frameworks for Knowledge Engineering Approaches', *JCNS*, pp. 15–20, Jan. 2022, doi: 10.53759/181X/JCNS202202003.

[3] M. Stefik, *Introduction to Knowledge Systems*. Elsevier, 2014.

[4] G. Schreiber, *Knowledge Engineering and Management: The CommonKADS Methodology*. MIT Press, 2000.

[5] L. F. Lai, 'A knowledge engineering approach to knowledge management', *Information Sciences*, vol. 177, no. 19, pp. 4072–4094, Oct. 2007, doi: 10.1016/j.ins.2007.02.028.

[6] J. C. Giarratano and G. Riley, *Expert Systems: Principles and Programming*. USA: Brooks/Cole Publishing Co., 1989.

[7] B. P. Allen, L. Stork, and P. Groth, 'Knowledge Engineering using Large Language Models', Oct. 01, 2023, *arXiv*: arXiv:2310.00637. doi: 10.48550/arXiv.2310.00637.

[8] S. Islam *et al.*, 'A comprehensive survey on applications of transformers for deep learning tasks', *Expert Systems with Applications*, vol. 241, p. 122666, May 2024, doi: 10.1016/j.eswa.2023.122666.

[9] M. Chelli *et al.*, 'Hallucination Rates and Reference Accuracy of ChatGPT and Bard for Systematic Reviews: Comparative Analysis', *Journal of Medical Internet Research*, vol. 26, no. 1, p. e53164, May 2024, doi: 10.2196/53164.

[10] Z. Xie, 'Order Matters in Hallucination: Reasoning Order as Benchmark and Reflexive Prompting for Large-Language-Models', Dec. 30, 2024, *arXiv*: arXiv:2408.05093. doi: 10.48550/arXiv.2408.05093.

[11] P. J. Phillips *et al.*, 'Four principles of explainable artificial intelligence', National Institute of Standards and Technology (U.S.), Gaithersburg, MD, NIST IR 8312, Sep. 2021. doi: 10.6028/NIST.IR.8312.

[12] D. Durmus, A. Giretti, O. Ashkenazi, A. Carbonari, and S. Isaac, 'The Role of Large Language Models for Decision Support in Fire Safety Planning', International Association on Automation and Robotics in Construction, 2024. doi: 10.22260/ISARC2024/0045.

[13] O. Ashkenazi, S. Isaac, A. Giretti, A. Carbonari, and D. Durmus, 'Transforming Building Industry Knowledge Management: A Study on the Role of Large Language Models in Fire Safety Planning', in *Proceedings e report*, P. Capone, V. Getuli, F. Pour Rahimian, N. Dawood, A. Bruttini, and T. Sorbi, Eds., Firenze University Press, 2023. doi: 10.36253/979-12-215-0289-3.73.

[14] A. Ławrynowicz, 'A Knowledge Engineering Primer', Mar. 25, 2024, *arXiv*: arXiv:2305.17196. doi: 10.48550/arXiv.2305.17196.

[15] P. Sahoo, A. K. Singh, S. Saha, V. Jain, S. Mondal, and A. Chadha, 'A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications', Feb. 05, 2024, *arXiv*: arXiv:2402.07927. doi: 10.48550/arXiv.2402.07927.

[16] R. F. Duval, 'NFPA Case Study: nightclub fires'. National Fire Protection Association, Fire Investigations Department, 2006.

[17] D. Durmus, J. Rožanec, A. Carbonari, A. Giretti, Ž. Turk, and M. Grobelnik, 'Potential of Large Language Models for Ontology Development in Construction Domain', in *Proceedings of the Creative Construction Conference 2024*, Praha, Czech Republic: Budapest University of Technology and Economics, 2024, p. null-null. doi: 10.3311/CCC2024-062.

Proposal of a Benchmark Indicator for Automated Construction Technology

Genki Yamauchi[1], Taro Abe[1], Daisuke Endo[1], Takeshi Hashimoto[1] and Keiji Nagatani[2]

[1]Public Works Research Institute, Japan
[2]University of Tsukuba, Japan

yamauchi-g573bs@pwri.go.jp, abe-t575co@pwri.go.jp, endou-d177cl@pwri.go.jp, t-hashimoto@pwri.go.jp, nagatani@cs.tsukuba.ac.jp

Abstract -

In recent years, the construction industry has urgently needed to address labor shortages caused by an aging workforce and population decline, drawing increasing attention to automated construction technologies. In this paper, focusing on earthworks, we propose a benchmark indicator composed of two levels: an 'automated construction level', which expresses the degree of technical sophistication following the example of autonomous driving levels in automobiles, and a 'social implementation level', which represents the degree of dissemination of technology through stages from research and development to market adoption. By introducing this indicator, we aim to assess both the automation level of construction machinery operations and the extent of their widespread use.

Keywords -

Autonomous Construction; Benchmark; Earthworks

1 Introduction

The construction sector is confronting a confluence of challenges: demographic shifts have led to a dwindling, aging workforce, tight labor markets push up costs, and the complexity and scale of projects continue to increase [1]. As traditional workflows prove insufficient to meet rising demands for quality, speed, and cost-effectiveness, automation emerges as a critical enabler for enhancing productivity, improving safety, and ensuring operational continuity. Prominent initiatives, such as Japan's i-Construction2.0 [2], aim to boost productivity by integrating advanced digital technologies, automated equipment, and streamlined data exchange mechanisms, aspiring to a 30% reduction in workforce demands by 2040.

Yet merely deploying automated machines is not enough. The industry needs robust frameworks to evaluate the maturity and broader impacts of emerging technologies. An Automated Construction Level (ACL), inspired by autonomous driving classifications defined by the Society of Automotive Engineers, has been proposed to map out the progression from manual operations (Level 0) to fully autonomous site management (Level 5) [3]. However, focusing solely on technical sophistication neglects the equally important dimension of societal acceptance and diffusion. For instance, even a highly advanced technology might remain confined to experimental testbeds if it fails to address market needs, instill user trust, or integrate smoothly with existing practices.

To address this gap, we introduce a benchmark that integrates the ACL with a Social Implementation Level (SIL), inspired by Moore's Chasm Theory [4], Rogers' Diffusion of Innovations [5], and complementary concepts such as Adoption Readiness Levels and Technology Readiness Levels [6]. By defining a two-dimensional matrix, we can situate technologies in terms of both their technical automation maturity and their stage in the social adoption lifecycle. This dual perspective informs developers about the importance of standardizing core planning processes before full automation, encourages investors to calibrate risk based on market penetration rates, assists policymakers in identifying where strategic support might be most effective, and enables end-users to choose tools aligned with their operational readiness.

ARL (Adoption Readiness Level), in particular, focuses on the commercial adoption process and social risks that cannot be fully captured by technological maturity alone. While this broad scope offers a multifaceted perspective, it can make the stage of technology adoption less intuitive to grasp. Meanwhile, evaluating automation in construction requires a comprehensive approach that considers not only the automation level of the machinery itself but also the automation of construction management processes. Although TRL (Technology Readiness Level) is widely used as a general framework for assessing technological maturity, it remains challenging to comprehensively evaluate automation under the diverse conditions unique to construction sites without adopting the SAE-based concepts employed in automotive autonomous driving—namely, DDT (Dynamic Driving Task) and ODD (Operational Design Domain). By aligning our new framework with insights from ARL and TRL, we aim to capture the interplay between technical progression and societal uptake

more holistically, while remaining sufficiently tailored to the realities of the construction industry.

In the following sections, we first elaborate on the concept of Automated Construction Levels (ACL), then propose a parallel Social Implementation Level (SIL), and finally integrate both to form a two-dimensional benchmark. We illustrate this approach with references to current examples of construction automation—such as partially and fully automated machines—and discuss how the combined ACL–SIL matrix can guide various stakeholders in decision-making and policy formulation.

2 Automated Construction Level: Beyond Machine Autonomy to Systemic Intelligence

The ACL framework to date has ranged from:

- **Level 0:** Manual operation with no automation.
- **Level 1:** Operator assistance (e.g., machine guidance) that augments human control while leaving most decision-making to the operator.
- **Level 2:** Partial automation of specific machine functions (e.g., automated blade control), reducing the cognitive load and skill barrier.
- **Level 2.9:** Full automation of a single machine's essential tasks, effectively removing the need for direct operator input in core operations.
- **Level 3:** Coordinated automation across multiple machines on a single site, yet still reliant on site-specific human planning.
- **Level 4:** Automated planning and task allocation, leveraging site data and design schematics to direct multiple machines autonomously within defined operational parameters.
- **Level 5:** Full autonomy across machines and tasks, unrestricted by predefined Operational Design Domains (ODD).

We propose an additional stage: **Level 3.5**, focusing on the generalization and systematization of the construction planning process—known in Japanese practice as "dandori." While Level 3 demonstrates that automation can handle a particular site's complexity, scaling this capability to diverse conditions demands codifying planning logic into a machine-readable, standardized framework. By achieving Level 3.5, developers set the stage for more readily attainable Level 4 and 5 solutions.

Examples from the industry underscore the necessity of this intermediate level. Current site-level automation cases, such as automated dam and tunnel construction using coordinated fleets of machinery [7], rely heavily on expert planners and site managers. Without codifying these human planning heuristics, scaling up remains prohibitively expensive and complex. Level 3.5 thus represents a strategic stepping-stone: it encourages a focus on knowledge representation, standardized planning protocols, and interoperable data models.

3 Social Implementation Level: Aligning Market Penetration with Trust and Adoption

Technological capability alone does not guarantee acceptance. Innovations must navigate the chasm between early enthusiasts and the mass market. Drawing on Moore's Chasm Theory and diffusion models, we categorize social implementation into:

- **Level 1:** Laboratory-scale R&D, prototypes, and pilot projects with negligible market presence.
- **Level 2:** Limited field trials and early demonstrations, potentially with small-scale commercial pilots.
- **Level 3:** Initial commercialization with constrained market penetration, adopted by specialized or niche segments.
- **Level 4:** Approaching the "chasm", technology gains footholds in multiple segments, nearing the early market's 16% threshold. Customer testimonials and early references start to build trust.
- **Level 5:** Stable adoption exceeding approximately 20% market share, suggesting a shift from novelty-driven acceptance to widespread trust and integration into industry norms.

The precise definition of market share depends on selecting an appropriate reference population. Construction markets are heterogeneous: a technology suitable for large-scale earthworks in infrastructure projects might not translate directly to residential or specialty construction. Thus, determining the relevant "market universe" is essential. Efforts to standardize reporting—potentially through industry associations or regulatory bodies—could help normalize these metrics.

4 Integrating the Two Axes: A Two-Dimensional Benchmark

By plotting ACL on one axis and SIL on the other, we create a two-dimensional matrix (Figure 1) that situates technologies within a nuanced landscape. For instance:

- **Teleoperation (ACL 0, SIL 3):** Remote controlled machines have been proven in disaster recovery since the 1990s [8]. Although technologically simple (human-operated), their wide acceptance in niche

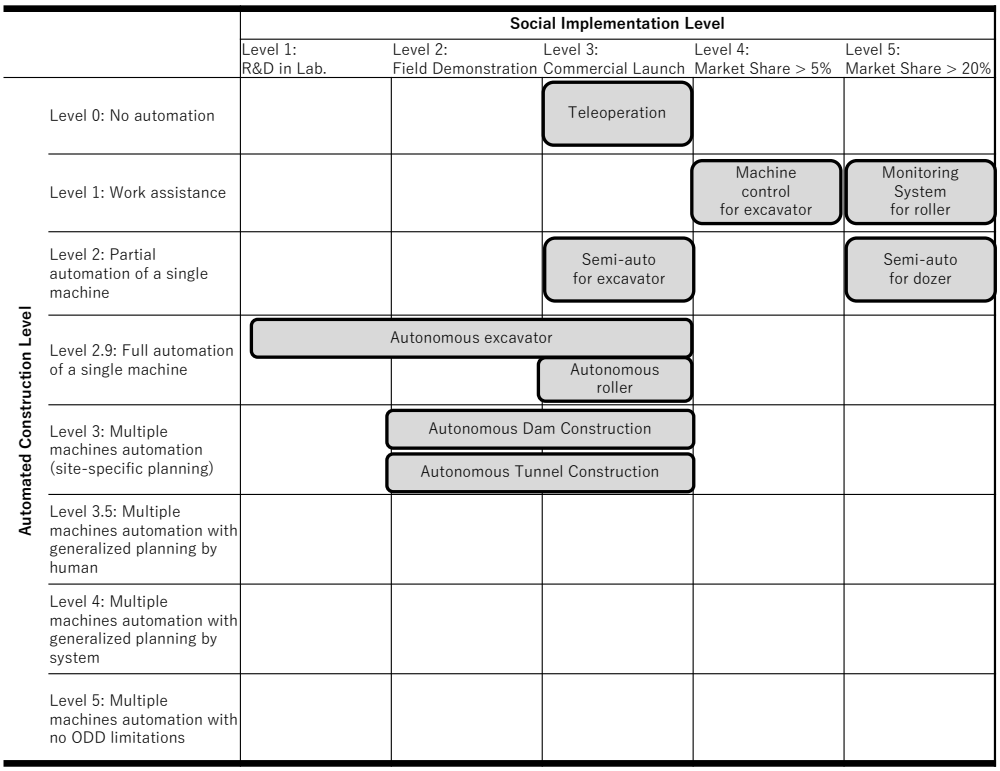


Figure 1. Evaluation of Automated Construction Technology Using Benchmark Indicators

scenarios places them around Social Implementation Level 3.

- **Machine Control System (ACL 1, SIL 4-5):** . In Japan, it is called machine guidance. Excavator machine control systems using GNSS, laser-based sensors, and 3D design data are widely available. Multiple manufacturers offer them and, in some regions, more than half of excavators can incorporate guidance features, pushing them into SIL 4 or 5.
- **Semi-auto system (ACL 2, Various SIL):** In Japan, it is called machine control. Automated blade control for bulldozers or partial autonomy in hydraulic excavators can boast significant market penetration, potentially reaching SIL 5 for certain machine categories [9].
- **Fully autonomous machine (ACL2.9, SIL 1-3):** Most fully autonomous machines remain trapped at R&D or prototype levels (SIL 1-2) [10, 11, 12, 13]. However, some autonomous excavator and roller have already been launched [14, 15].
- **Site-Level Automation (ACL 3, SIL 2-3):** Projects such as automated dam and tunnel construction [7] exemplify site-wide automation. However, these remain bespoke, heavily engineered solutions, with minimal standardization and limited replication across diverse projects.

Crucially, we find no examples at ACL 3.5, 4, or 5, nor do we see technologies in widespread, stable mainstream use at a corresponding SIL for these higher ACLs. This observation reveals a strategic opportunity: bridging the technical and social dimensions is not simply about incremental improvement, but about systemic change, collaborative data standards, and knowledge exchange.

5 Discussion and Future Directions

Our benchmark reveals that while some technologies, such as machine guidance, are well integrated and widely trusted, others, such as fully autonomous excavators or site-level systems, remain at an early stage. For researchers and developers, ACL 3.5 stands out as a priority: codifying planning processes to reduce site-specific engineering costs. Investors, meanwhile, can better calibrate risk by noting whether a technology aligns with established mainstream practices (SIL 5) or remains experimental (SIL 1–2). Since there are still few real-world cases in ACL 3 or higher, further development of these technologies should be accompanied by large-scale empirical studies and rigorous statistical analyzes to thoroughly validate and refine the proposed benchmark.

Policymakers can use these metrics to allocate resources and encourage standardization. Supporting common data protocols, testbeds, and training could streamline the transition from isolated successes to broad adoption. Prac-

titioners benefit by selecting tools that match their operational readiness and understanding whether a particular solution is proven or still exploratory.

Future refinements may involve adding sub-levels, incorporating sustainability metrics, or considering regional market nuances. Aligning this benchmark with efforts in related industries—such as automotive or mining—could further enhance its clarity and relevance. As the construction automation ecosystem evolves, continuous feedback will help refine both ACL and SIL scales, ensuring they remain practical and reflective of real-world conditions.

6 Conclusion

This paper introduced a benchmark framework that unites technical maturity (Automated Construction Level) with societal diffusion (Social Implementation Level) to provide a holistic perspective on automated construction technologies. By incorporating a novel ACL 3.5 stage that emphasizes the formalization of planning logic, and linking market penetration milestones to known thresholds of consumer acceptance, the benchmark reveals strategic development pathways and potential gaps.

In practice, this framework offers value to multiple stakeholders: researchers gain clearer R&D targets, developers appreciate the importance of systematic planning processes, investors gauge market readiness more accurately, policymakers identify priority support areas, and practitioners discern which technologies fit their current needs. Although still in a formative stage, our benchmark invites further input, adaptation, and validation. By continually refining these metrics, the construction industry can move toward more transparent, data-driven decision-making and accelerate the meaningful adoption of automation on construction sites worldwide.

Acknowledgement

This work was supported by Council for Science, Technology and Innovation(CSTI), Cross-ministerial Strategic Innovation Promotion Program (SIP), the 3rd period of SIP "Smart Infrastructure Management System" Grant Number JPJ012187 (Funding agency: Public Works Research Institute).

References

[1] McKinsey Global Institute. Reinventing construction: A route to higher productivity. On-line: https://www.mckinsey.com/~/media/mckinsey/business%20functions/operations/our%20insights/reinventing%20construction%20through%20a%20productivity%20revolution/mgi-reinventing-construction-a-route-to-higher-productivity-full-report.pdf, Accessed: 18/12/2024.

[2] Transport Ministry of Land, Infrastructure and Tourism. i-construction 2.0. On-line: https://www.mlit.go.jp/tec/constplan/content/001738240.pdf, Accessed: 18/12/2024.

[3] Hirokuni Morikawa and Takashi Otsuki. Study on the level concept of autonomous construction in mechanized construction. In *Proceedings of the 37th International Symposium on Automation and Robotics in Construction (ISARC)*, pages 1114–1121, 2020.

[4] G. A. Moore. *Crossing the Chasm*. HarperBusiness, 1991.

[5] E. M. Rogers. *Diffusion of Innovations*. 2003.

[6] Adoption Readiness Levels (ARL): A Complement to TRL. On-line: https://www.energy.gov/technologytransitions/adoption-readiness-levels-arl-complement-trl, Accessed: 18/12/2024.

[7] Kajima Corporation. A^4CSEL Next-Generation Construction System. On-line: https://www.kajima.co.jp/english/tech/c_a4csel, Accessed: 18/12/2024.

[8] Kazuhiro Chayama et al. Technology of unmanned construction system in japan. *Journal of Robotics and Mechatronics*, 26(4):403–417, 2014.

[9] KOMTATSU. Intelligent machine control. On-line: https://www.komatsu.com/en/technology/smart-construction/intelligent-machine-control/, Accessed: 18/12/2024.

[10] Rauno Heikkilä et al. Development of an earthmoving machinery autonomous excavator development platform. In *Proceedings of the 36th International Symposium on Automation and Robotics in Construction (ISARC)*, pages 1005–1010, 2019.

[11] ETH Zurich. HEAP Project. On-line: https://rsl.ethz.ch/robots-media/heap.html, Accessed: 18/12/2024.

[12] SafeAI. Autonomous construction and mining. On-line: https://www.safeai.ai/, Accessed: 18/12/2024.

[13] DeepX. Automating any machine and innovating global industries. On-line: https://www.deepx.co.jp/en, Accessed: 18/12/2024.

[14] BuiltRobotics. Trenching. On-line: https://www.builtrobotics.com/solutions/trenching, Accessed: 18/12/2024.

[15] Sakai Heavy Industries. Autonomous roller ARMs. On-line: https://www.sakainet.co.jp/products/ict/autonomous_roller.html, Accessed: 18/12/2024.

BIM-Based Public Participation: Insights and Lessons Learned from a Case Study

Jonathan Matthei[1], Maximilian Friedhelm Heinrich Christ[1], Sven Mackenbach [1], Katharina Klemt-Albert[1]

[1]RWTH Aachen University, Germany
matthei@icom.rwth-aachen.de, m.christ@icom.rwth-aachen.de, mackenbach@icom.rwth-aachen.de, klemt-albert@icom.rwth-aachen.de

Abstract

Public participation is critical to the success of construction projects as it enables citizens to influence and engage in planning processes. Traditional participation methods often rely on two-dimensional technical plans, which are frequently perceived as difficult to understand, particularly by non-experts. In recent years, digital visualization techniques, especially 3D models, have emerged as promising alternatives. These approaches promise to enhance communication among stakeholders and deepen engagement by offering clearer and more comprehensible representations of construction projects.

This paper explores how public participation based on 3D BIM model data can be implemented and examines the user experiences associated with such an approach. The concept is built on a 3D BIM model-based participation platform integrated with a Common Data Environment (CDE), which serves as the source for three different model variants.

A prototype of the participation platform was tested in the context of public involvement for the redesign of a mobility space in a German city. Citizens were therefore invited to interact with the various model variants and actively participate in the process.

Findings from this initiative demonstrate that guided textual information with 3D visuals enhances citizens' comprehension of the project. Users particularly valued the 3D models for enhancing their understanding, while expressing a preference for rather realistic models, even if they are more complex. In summary, the results suggest that BIM models can significantly enhance public participation by improving comprehension and engagement, ultimately contributing to social sustainability.

Keywords

Building Information Modeling (BIM), public participation, 3D visualization, e-participation, participatory planning, Scrollytelling, Common Data Environment (CDE), social sustainability

1 Introduction

Public participation plays a crucial role in the success of construction projects [1], as it enables citizens to exert influence and contribute to the decision-making processes of projects planned within their communities [2]. Inadequate integration of the public in planning processes can have significant consequences, including protests, legal disputes, and delays in project timelines. These issues often result in construction projects failing to meet budgetary and scheduling targets [3].

Understandable visualizations are considered fundamental for successful public participation [4]. According to King et al. (1989), visualizations are the only common language that all stakeholders, both technical and non-technical, can refer to [5].

Currently, primarily two-dimensional technical plans, floor plans, sections, and views are provided as supporting visualizations of the project to the public [3]. Consequently, technical drawings represent the most commonly used method for visualizing construction projects [2].

However, there is a consensus that these are generally difficult to understand, especially for stakeholders with little or no experience in interpreting maps [1; 3; 6]. Consequently, such forms of representation can be considered outdated and ineffective in the context of public participation.

With the increasing digitization in recent years, methods have increasingly been employed that promise improved communication and representation of information [2]. These novel visualization methods are often closely linked to digital participation formats, collectively referred to as e-participation measures. At the core of these measures are Information and Communication Technologies (ICT), which aim to facilitate and make the exchange between stakeholders more efficient [7].

Numerous studies can be found that examine the role of digital visualization formats and their impact on public understanding and participation.

The study by Skaaland and Pitera (2021) investigated the use of digital visualization techniques in the context of public participation in infrastructure projects in Norway. A survey was conducted among professional users in the Norwegian infrastructure sector, comparing various visualization techniques to examine their impact on public understanding. The study identified that digital visualization models led to a statistically significant increase in public understanding and engagement compared to traditional visualization models [2].

In another study by Onyimbu et al. (2018), the use of 3D visualizations in the context of digital public participation in Kenya was examined. Questionnaires were sent to planners to assess the user-friendliness of 3D visualizations. The study concluded that the use of 3D visualizations has the potential to improve both the quality and quantity of public participation. Consequently, 3D visualizations are able to promote the involvement of stakeholders and enhance their understanding of the proposed plans [1].

This finding aligns with the statement by Milosz, who further emphasizes that digital 3D models allow non-experts to have more control over what they want to see, as opposed to what planners want to show them [8].

The research project VISB+ examined the contribution of visualizations to improved communication of construction plans. Various visualization approaches were analyzed within the study, with interactive visualizations being particularly highlighted for their added value. Furthermore, the study emphasizes the potential benefits of applying Building Information Modeling (BIM) in the context of public participation [4].

BIM represents a methodology that has seen increasing adoption in the construction industry since the early 21st century. Fundamentally, this approach relies on the use of digital building models and the semantic information within these models throughout the entire lifecycle of a building [9]. Buildings are represented as three-dimensional models, with each component being detailed and digitally depicted [4].

By using BIM models, different planning variants can be visualized and made tangible in the real environment [3]. Furthermore, these models allow for the interactive integration of additional layers of information, such as noise propagation or shading effects. This could significantly simplify communication in the context of public participation and contribute to greater transparency, as well as enhance the social sustainability of construction projects [3].

Despite this potential, the VISB+ study concludes that the use and research related to the application of BIM models for public participation is still limited. In particular, there is a lack of comprehensive studies on how these models can be effectively integrated into participatory processes to foster broad acceptance and understanding among non-technical stakeholders. Therefore, further research and development efforts are necessary to explore the application of BIM in the context of public participation.

The aim of this study is to examine the use of BIM model data in the context of public participation and to collect user experiences in this regard.

The study is part of the research project BIM4People, funded by the Federal Ministry for Digital and Transport in Germany. As part of this project, a digital participation platform is being developed, which presents BIM model data and will be tested and optimized through several real-world use cases.

This study provides a comprehensive overview of the approach and findings from a case study of a BIM-based participation platform. The case study is situated in a German city with approximately 100,000 inhabitants, representing a voluntary initiative to explore innovative means of involving the public in urban development projects.

The underlying concept of BIM-based participation is detailed in the following second chapter. Additionally, the study area, the model variants, and the opportunities for public participation are described. In the third chapter, the empirical study that accompanied this case study is presented. The aim of the empirical study is to identify the user experience of participants in relation to the BIM-based participation platform. The study aims to gather insights into how digital 3D models can be effectively integrated into public participation processes to enhance transparency, foster acceptance of construction projects, and explore public perceptions of these models. This chapter includes the specific research topics as well as the corresponding findings. The final chapter of this study provides a discussion and conclusion regarding the case study of public participation based on BIM model data.

2 The concept of BIM-based public participation

At the core of BIM-based public participation are two key software products. On one hand, there is a Common Data Environment (CDE), which serves as the central information source. A CDE is used to collect, manage, and disseminate relevant information related to a specific project [10].

In the study discussed below, three model variants were developed collaboratively by the city and the planning office as part of a BIM-based planning process. These models were created using authoring software and subsequently uploaded to the CDE in Industry Foundation Classes (IFC) format. This approach facilitated a centralized storage system, enabling collaboration and data exchange among the different

stakeholders.

In this use case, the IFC model was subsequently integrated into the city model developed during the urban planning concept phase. The city model itself was also available as an IFC file.

Since a CDE is primarily designed as a planning tool for professionals, a second software product, a digital participation platform, is developed and subsequently used for a case study. This platform enables the public to view the model variants through an IFC viewer and access the associated information, allowing them to actively participate in the planning process.

In order to display the IFC models on the platform, xeokit, an IFC viewer which allows open-source 3D graphics formats to be displayed in the browser [11], was used.

The study described below incorporates a Google Maps layer into the BIM-based platform as an additional feature for spatial orientation. This integration highlights the geographical context of the project, displaying its location in relation to the existing urban structures. The familiar Google Maps interface is intended to enhance users' comprehension of the project's spatial context by offering a clear and intuitive visualization of the area in which the project is situated.

The concept of BIM-based participation, which is based on the research findings of Matthei et al. (2024) [12], is summarized in the following figure.

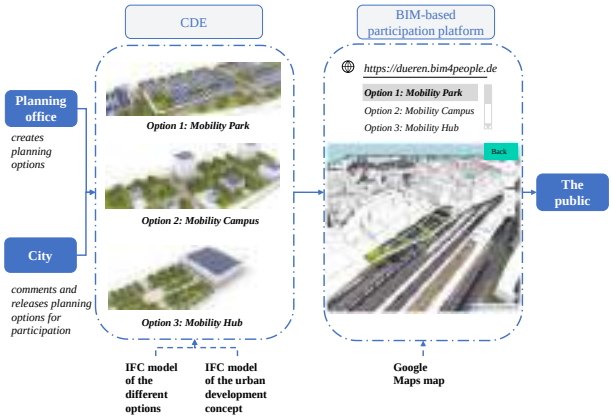


Figure 1. Concept of the BIM-based public participation

2.1 Study area

The study area for the BIM-based public participation concept is located in Düren, a city of approximately 100,000 residents, situated between Aachen and Cologne. Due to its location in a densely populated state and between two major cities, Düren has a high number of commuters. The city is also significantly impacted by the structural transformation from the coal phase-out, with job losses, weakening supply industries, and declining retail purchasing power [13].

An urban development project is planned along Düren's main train station, where a new 150,000 m² district will be created. A specific area within this project will focus on developing and evaluating innovative mobility solutions. In this area, the BIM-based public participation using three model variants is located.

2.2 Model Variants and Participation Opportunities

Three model variants were developed early in the BIM-based planning process to be assessed during the public participation phase. These are preliminary drafts, not detailed plans, aimed at gathering public input on needs and preferences. Option 1 (Mobility Park) emphasizes mobility services, open spaces, parking, green areas, and solar roofs. Option 2 (Mobility Campus) features modern parking towers with automated parking and sustainable materials. Option 3 (Mobility Hub) focuses on a compact building with integrated uses.

Each option is supplemented with brief textual descriptions on the BIM-based participation platform, providing citizens with a clearer overview of the content. The presentation of the variants is done in the format of "scrollytelling," which serves as a promising tool for conveying complex information in an understandable way [14]. In scrollytelling, participants navigate from one story point to the next by scrolling. Each story point represents specific locations or situations within the model that have been pre-selected to guide users through the respective variant in a structured manner. These story points are accompanied by explanatory texts that provide additional information about the displayed scenes, deepening the users' understanding of the content.

The model is automatically aligned with the location described in the text, establishing a strong connection between the visual content and the accompanying texts. This methodical linkage aims to enhance the understanding of the complex spatial and urban planning relationships. In addition to the guided presentation, citizens have the opportunity to freely explore the model variants. Participants can switch between views, zoom in, and rotate the 3D model, allowing for a more personalized and comprehensive impression of the planning variants. Public participation takes place by allowing citizens to view the model variants, create their own contributions, and to read contributions from other users, which they can also agree with. When submitting a contribution, users have the option to associate it with one of the three model variants and place a marker at a specific location within the model (see figure 2). This allows concerns to be linked to a precise spatial allocation, facilitating clearer understanding of the issues raised. Additionally, contributions can be categorized by topic to facilitate targeted processing of concerns by the city administration.



Figure 2. Participation opportunity on the BIM-based participation platform

3 Empirical study

3.1 Proceedings

As part of the BIM-based public participation process, 600 flyers were distributed within the study area, supplemented by announcements in local newspapers, on the city's website, and through social media channels. A QR code and a direct link to the participation platform were provided to facilitate access and encourage public engagement. The public participation period was open from October 24, 2024, to November 17, 2024.

To capture detailed user experiences, a survey was integrated directly into the BIM-based participation platform. Participants were invited to take the survey on the final page of the platform after reviewing the different model variants. Additionally, users who submitted a contribution were also prompted to complete the survey. This methodology aimed to gather a broad and representative sample by encouraging both passive and active users to participate.

The survey is divided into two sections. The first part collects demographic and socio-economic information about the participants, including gender and age. In addition, participants are asked about their relationship to the particular city, e.g. whether they are residents or commuters. The section also gathers information on their previous experiences with public participation in general.

The second part of the survey aims to gain detailed insights into the user experience of participants with the BIM-based participation platform. It focuses on experiences related to navigation and spatial orientation within the 3D models. Additionally, the usefulness of various platform functionalities is examined, including the georeferenced placement of contributions, the guided presentation of content through scrolling on the platform, and the categorization of contributions. Another key focus is on how the 3D representations help participants better understand the project and how useful the ability to compare different project variants using a 3D model is. The second part of the survey is conducted through multiple-choice questions on a 5-point Likert scale.

3.2 Findings

The following subchapter present the key findings from the conducted BIM-based public participation process. First, general observations are presented, followed by user experiences based on the insights gathered from the survey placed on the platform.

3.2.1 General findings

The BIM-based participation platform had 593 unique users, with an average visit duration of 3 minutes and 39 seconds. Of these, 364 users accessed the platform via mobile devices (61%), spending an average of 2 minutes and 12 seconds, while 229 users accessed it via desktop, spending an average of 4 minutes and 40 seconds.

Among the 593 visitors, 76 individuals were returning users, indicating that they accessed the platform multiple times. A total of 24 contributions were submitted on the platform, 16 of which included a direct link to the respective variant referenced in the comment. The first variant was linked six times, making it the most frequently referenced option. Furthermore, in three instances, contributions were linked to specific locations within the model.

An analysis of the thematic categorization of contributions revealed that in 71% of cases, the assigned topic accurately matched the content of the submission.

3.2.2 User experience

A total of 23 individuals participated in the survey aimed at identifying user experiences. Of these, 21 respondents completed at least 80% of the survey. The following section exclusively presents the findings derived from these 21 participants.

The participant background reveals that 86% had never participated in public engagement activities before. Regarding gender, 48% identified as male, 43% as female, and two participants chose not to disclose their gender. In terms of their relationship to Düren, 41% lived there, 18% worked in the city, and 27% commuted for work. The age distribution was balanced across 18-64 years, with a slight prominence in the 25–34 age group. No participants were under 18 or over 64, and one participant did not provide their age (see table 1).

Table 1 Age Distribution of Survey Participants

Age	Number
<18	0
18-24	4
25-34	5
35-44	4
45-54	3
55-64	4
>64	0
n/a	1

The evaluation of user experiences with spatial orientation and navigation in the 3D model shows that 33.33% of respondents found orientation to be "neither easy nor complicated," while 20% experienced some difficulty, and 15% found it "very easy." Regarding navigation, 38.10% rated it as "very easy," 23.81% found it "somewhat difficult," 14.29% considered it "somewhat easy," and another 14.29% viewed it as "neither easy nor difficult." One participant, aged 55–64, found navigation "very difficult." A detailed representation of these results is provided in the following figure.

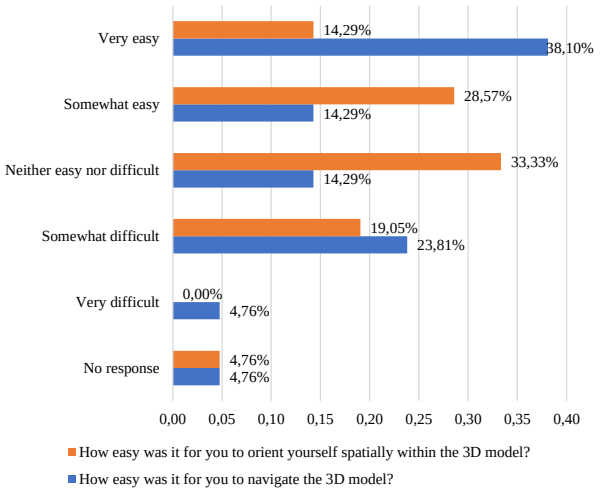


Figure 3. User experiences regarding orientation and navigation in relation to the 3D model

The investigation of functions on the BIM-based participation platform revealed that the guided presentation of information (scrollytelling) with textual and 3D model elements was seen as helpful by most participants. Over two-thirds rated it as "very helpful" (47.62%) or "somewhat helpful" (19.05%), with 14.29% neutral. The ability to assign contributions to specific locations was considered "somewhat helpful" by 35% of respondents, with an equal number providing no answer. About 15% found it "very helpful," and 10% neutral. Categorizing contributions by theme was mostly rated as "somewhat helpful" (30%) or "very helpful" (25%), while 15% were neutral and 15% found it "somewhat unhelpful".

Regarding 3D visualization, 90% of participants said it somewhat improved their understanding, with 45% rating it "very helpful" and 45% "somewhat helpful." The comparison function of different variants using the 3D model was rated positively, with 50% finding it "very helpful" and 33.33% "somewhat helpful." A detailed overview of the evaluation is shown in the following figure.

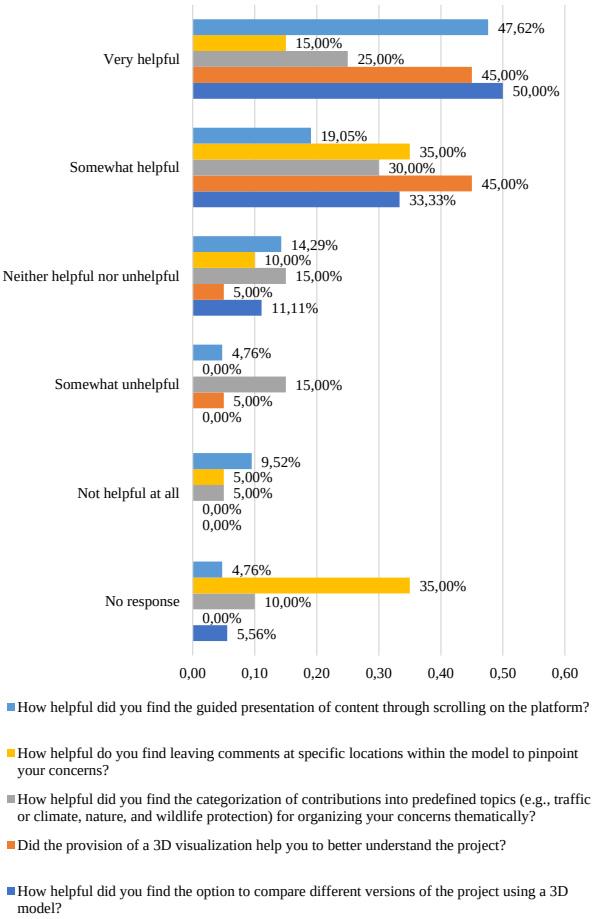


Figure 4. Usefulness of different functions

The majority of respondents preferred more realistic 3D models, even at the cost of increased complexity. About 17% favored a very realistic model, despite potential complexity, while 39% preferred a realistic model that was somewhat complex. One-third sought a balance between realism and simplicity, and only 5.56% desired a very simple, minimalistic model with little realism. No participants preferred a simple, unrealistic model. Detailed results are shown in the following figure.

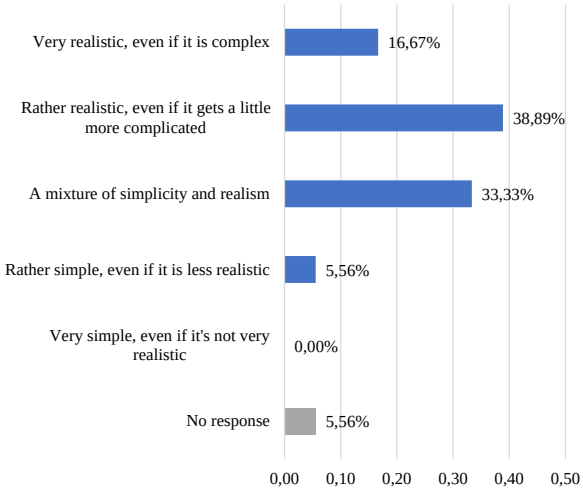


Figure 5. Respondents' 3D model detail and realism preferences

4 Discussion and Conclusion

4.1 Discussion

This case study of BIM-based public participation shows that 63% of users accessed the platform via mobile devices. However, desktop users spent more than twice as long on the platform, suggesting that while mobile devices are convenient, desktops offer a better experience for in-depth engagement, likely due to larger screens and easier navigation of complex 3D content.

Although the platform recorded a total of 593 different visitors and was available to the public for 24 days, only 24 contributions were made. This low level of participation may be due to the need to register by email when creating contributions and the mandatory public visibility of these. Low-threshold access could therefore significantly increase the willingness to participate in the future. However, the possibility of low-threshold participation harbors the risk that people will participate without having a deeper connection to the topic and without being directly affected by the measures under discussion [15]. The phenomenon of “slacktivism” or “clicktivism” can occur in this context. This term describes a form of digital participation in which people get involved through simple, often situational actions such as liking, sharing or commenting, often without deep engagement or direct personal impact [16]. The result would be a distorted participation result. Another reason for the low participation level may be due to the high information density of the platform, which could have overwhelmed users and hindered active engagement.

The active linking of contributions to the model variants was carried out for the majority of posts (16 out of 24). This indicates that the majority of users consider the option of linking their concerns directly to specific model variants to be valuable and useful. In three cases, contributions were linked to specific model locations, suggesting the function was useful to some participants.

A common problem in public participation is the lack of representation: The same, particularly committed citizens often take part, which can lead to a distorted picture of the opinions and interests of the population. Only a specific, actively engaged group participates regularly, while the majority hardly uses the opportunities for participation. As a result, the population structure often remains incompletely represented and the results of participation do not necessarily reflect the entirety of those affected [17].

The results of this study, however, present a different picture: A large majority of participants (86%) reported that they had never previously participated in a public engagement process. This high proportion of first-time participants suggests that the digital participation platform was successful in reaching individuals who would otherwise not engage in traditional participation opportunities. Additionally, the age data from the survey indicates a broad distribution of participants. The age groups between 18 and 64 years are almost evenly represented, with a slight concentration in the 25–34-year-old category. This suggests that the platform also appeals to younger adults, a group that is often underrepresented in traditional forms of public participation [18].

The balanced gender distribution and the presence of both residents and commuters among participants suggest that the platform successfully appeals to a diverse user group. These findings indicate that digital participation platforms can reach a broader and more varied audience than traditional formats.

The results regarding the analysis of user experiences of spatial orientation and navigation in the 3D model show mixed experiences. In connection with orientation in the 3D model, around a third of respondents felt that orientation in the model was neutral, which indicates that the 3D model is functional in terms of orientation, but not particularly intuitive for many. This leaves room for optimization in connection with orientation support within 3D models. If necessary, additional introductions or tutorials could help to promote orientation within 3D models for citizens.

The majority of respondents found navigation within the 3D model "very easy," suggesting that the interactive controls are well-implemented and intuitive. This may be due to the increasing accessibility of immersive technologies, familiarizing more people with 3D models [19]. However, about a quarter of participants found navigation "rather difficult," indicating potential challenges related to varying technical skills or devices.

The conclusions drawn from the analysis of the various functions of the BIM-based participation

platform show that the platform is perceived as helpful and user-friendly overall. The guided presentation of information (scrollytelling), which consists of textual and 3D model elements, was rated particularly positively. Over two thirds of participants found this function helpful, which indicates that clear and structured information transfer in conjunction with interactive visual elements promotes user understanding.

A large proportion of respondents also found it helpful to locate contributions at specific points in the model, which underlines the importance of the spatial allocation of a contribution in the context of participation. Nevertheless, there was a considerable number of participants (35%) who did not provide any information on this, which could indicate that this option was not sufficiently pointed out or that it is still considered less relevant at such an early planning stage. The categorization of contributions according to thematic categories was predominantly seen as helpful. This shows that structuring contributions thematically allows for a better overview and targeted discussion of specific aspects of a project. However, some respondents found this function less helpful, which may indicate that an even finer adjustment of the categories or additional filter options could be useful.The provision of 3D visualization was rated as helpful by almost all participants, confirming that visual representations of complex projects significantly improve understanding. In particular, the ability to compare different variants of the project in the 3D model was found to be extremely helpful by the majority of users. This shows that the visual comparison of project variants, in addition to the textual comparison, is considered valuable by users and can contribute to a better understanding of the project.

Several conclusions can be drawn from the respondents' preferences regarding the level of detail and realism of 3D models: Firstly, there is a high demand for realistic representations. The majority of respondents prefer realistic 3D models, even if this is associated with increased complexity. This indicates that a detailed and realistic visualization is perceived as more important for the understanding and acceptance of the project than a simplified representation. Furthermore, there is an acceptance of complexity for better information content. The willingness to prefer more complex models shows that users value detailed information and are prepared to accept greater complexity in favor of better understanding. This speaks in favor of using models that reproduce details as accurately as possible. Regardless of this, a need for balance can be seen among some users. Around a third of respondents prefer a balance between realism and complexity, which indicates different user needs and abilities. In addition, there is a low demand for simple models. This is reflected in the very low preference for simple or highly abstracted models.

4.2 Conclusion

Digital transformation, particularly through Building Information Modeling (BIM), is enhancing infrastructure planning and public participation. BIM enables three-dimensional representations, improving transparency and comprehensibility compared to traditional visualizations, potentially fostering more intensive citizen engagement.

As part of this study, an approach for BIM-based public participation was developed that makes it possible to present BIM model variants located in a CDE to the public using a digital participation platform. As part of a case study in the course of redesigning a mobility space in a German city, initial empirical values were collected in order to evaluate the applicability and user-friendliness of a participation platform based on 3D BIM model data.

The study aimed to understand user perceptions of the platform and identify challenges and opportunities in interacting with 3D visualizations and other features.

The BIM-based public participation process has shown that guided information transfer (scrollytelling) in conjunction with interactive visual elements can be considered helpful. Citizens consider the possibility of seeing the project in 3D and comparing variants in this way to be helpful for their understanding of the project. The option of linking contributions with different model variants was frequently used in this context.

The evaluation of navigation in the 3D model is predominantly positive, which suggests that it is generally easy to use. Orientation in the 3D model also tended to be rated positively, although some users perceived this as “rather difficult”. This indicates a need for optimization in order to improve spatial orientation when using 3D models.

The background information on the participants suggests that a digital participation platform is a suitable medium for reaching a broader and more diverse spectrum of the population than is possible in traditional formats. In this context, digital public participation represents an important step towards inclusive participation, a key element of social sustainability.

With regard to the level of detail of 3D models, this study has shown that realistic and detailed 3D models are considered more useful for public participation. However, in order to take different user needs into account, a platform that allows flexible adjustment of the level of detail - from detailed to simplified representations - could be useful.

This study has demonstrated how public participation can be facilitated using BIM model data. The results of this case study suggest that a guided presentation of information, combined with interactive 3D visualization, can significantly enhance citizens' understanding of planned projects. This structured approach appears to be effective in making complex content more accessible and comprehensible, potentially improving the quality of

public participation. However, it is important to acknowledge the relatively small number of participants who actively engaged with the platform beyond merely accessing it. While 593 users visited the site, only 23 individuals participated in the user experience survey. This raises the question of whether some users were deterred by the platform's technological aspects, particularly the 3D models, which could challenge the assumption that BIM-based visualization inherently fosters greater participation. This paradox warrants further investigation.

Future studies should expand the sample size and use detailed models for deeper insights into user experience and acceptance. Comparing 3D visualizations with 2D plans can clarify their advantages in participation, comprehension, and decision-making. Furthermore, integrating citizen feedback into BIM-based planning remains essential.

Acknowledgment

This paper was written as part of the BIM4People project, funded by Germany's Federal Ministry for Digital and Transport. The model variants were created by Carpus + Partner AG, and the BIM-based participation platform by Interactive Pioneers GmbH.

References

[1] Onyimbi, J. R.; Koeva, M.; Flacke, J. (2018) *Public Participation Using 3D Web-Based City Models: Opportunities for E-Participation in Kisumu, Kenya* in: ISPRS International Journal of Geo-Information, 7 (12). https://doi.org/10.3390/ijgi7120454

[2] Skaaland, E.; Pitera, K. (2021) *Investigating the use of visualization to improve public participation in infrastructure projects: how are digital approaches used and what value do they bring?* in: Urban, Planning and Transport Research, 9 (1), pages 171–185. https://doi.org/10.1080/21650020.2021.1887757

[3] vbw-Vereinigung der Bayrischen Wirtschaft e.V. (2018) *Digitales Planen und Bauen-BIM*

[4] *VisB+ Leitfaden BWStiftung_2017_10_25 Fraunhofer IAO_Uni Hohenheim*.

[5] King, Stanley, Conley, Merinda, Latimer, Bill, Ferrari, Drew (1989) *Co-design : a process of design participation* [online]. https://cir.nii.ac.jp/crid/1130282270751550976 Accessed: 06/12/2024.

[6] Lane, M. B. (2005) *Public Participation in Planning: an intellectual history* in: Australian Geographer 36 (3), pages 283–299. https://doi.org/10.1080/00049180500325694

[7] Sanford, C.; Rose, J. (2007) *Characterizing eParticipation* in: International Journal of Information Management 27 (6,) pages 406–421. https://doi.org/10.1016/j.ijinfomgt.2007.08.002

[8] Milosz, M.; van Lammeren, R.; Hoogerwerf, T. (2007) *Usability of 3D geo-visualisation for spatial orientation* in: van den Brink, A. [Ed.] *Imaging the future: Geo-visualisation for participatory spatial planning in Europe*. Wageningen Academic Publishers, pages 119–126.

[9] Borrmann, A. et al. (2021) *Die BIM-Methode im Überblick* in: Borrmann, A. et al. [Ed.] *Building Information Modeling*. Wiesbaden: Springer Fachmedien Wiesbaden, pages 1–34.

[10] DIN EN ISO 19650-1_ (2019) *Organisation und Digitalisierung von Informationen zu Bauwerken und Ingenieurleistungen einschließlich BIM – Teil 1 Begriffe und Grundsätze*.

[11] xeokit *Web Programming Toolkit for AEC Graphics* [online]. https://xeokit.io/

[12] Matthei, J.; Rausch, S.; Klemt-Albert, K. (2024) *BIM-basierte Planfeststellung für Infrastrukturgroßprojekte* in: Bautechnik. https://doi.org/10.1002/bate.202400016

[13] Stadt Düren (2024) *Strukturwandel* [online]. https://www.dueren.de/wirtschaft-handel/strukturwandel. Accessed: 18/10/2024.

[14] Seyser, D.; Zeiller, M. (2018) *Scrollytelling – An Analysis of Visual Storytelling in Online Journalism* in: *2018 22nd International Conference Information Visualisation*. Fisciano, Italy. IEEE, pages 401–406.

[15] Leitner, M. (2018) *Digitale Bürgerbeteiligung: Chance und Herausforderung* in: Leitner, M. [Ed.] *Digitale Bürgerbeteiligung*. Wiesbaden: Springer Fachmedien Wiesbaden, pages 1–10.

[16] Morozov, E. (2009) *Foreign Policy: Brave New World Of Slacktivism* [online]. https://www.npr.org/2009/05/19/104302141/foreign-policy-brave-new-world-of-slacktivism Accessed: 10/11/2024.

[17] Klages, H.; Vetter, A. (2013) *Bürgerbeteiligung auf kommunaler Ebene-Perspektiven für eine systematische und verstetigte Gestaltung*. Baden-Baden: Nomos Verlagsgesellschaft mbH&Co.KG.

[18] Fernández Guzmán Grassi, E.; Portos, M.; Felicetti, A. (2024) *Young People's Attitudes towards Democracy and Political Participation: Evidence from a Cross-European Study* in: Government and Opposition 59 (2), pages 582–604. https://doi.org/10.1017/gov.2023.16

[19] Partarakis, N.; Zabulis, X. (2024) *A Review of Immersive Technologies, Knowledge Representation, and AI for Human-Centered Digital Experiences* in: Electronics 13 (2), https://doi.org/10.3390/electronics13020269

Web Based Security Assessment of Open-Source CAD tools using ZED Attack Proxy (ZAP) scanner

Abdul Samad Shakir[1], and Bharadwaj R. K. Mantha[1*]

[1]Department of Civil and Environmental Engineering, College of Engineering, University of Sharjah, Sharjah, United Arab Emirates.
shakirabdul31@gmail.com, rmantha@sharjah.ac.ae

Abstract –

The post-pandemic rise in Computer-Aided Design (CAD) tool adoption has driven a shift toward accessible, cost-effective open-source and "Do-it-yourself" (DIY) CAD tools, increasingly favored by individuals, small and medium sized enterprises (SMEs), and entrepreneurs over proprietary software. Despite their growing popularity, none of the existing studies investigate the security of these open-source CAD tools, particularly when users download and install them on personal devices. This study addresses this key research gap by identifying vulnerabilities of open-source CAD tools used specifically in the Additive Manufacturing (AM) process using the Zed Attack Proxy (ZAP) scanner. A web-based security assessment of five widely used CAD tools was conducted, where 25 unique web-based vulnerabilities were identified. This study provides valuable insights for SMEs operating in the architecture, engineering, construction, and operations (AECO) industry by highlighting the potential risks associated with using open-source CAD tools. These risks include the corruption of critical information on user machines, manipulation and loss of CAD data, and unavailability of CAD tools to operate. Such issues directly impact the core principles of cyber security i.e. confidentiality (C), integrity (I), and availability (A) highlighting the critical need to address these vulnerabilities and ensure secure use of open-source CAD tools.

Keywords –

Common weakness enumeration (CWE); Cyber-attacks; AECO; CAD model; Application security testing; Web based attacks.

1 Introduction

Computer aided design (CAD) tools are essential in the additive manufacturing (AM)/Three-Dimensional (3D) printing process, serving as the primary tools for designing and verifying models across various industries. They are used to create, modify, and analyse designs, ensuring that the final product meets specific requirements. Recently, there has been an increasing use of CAD tools across various sectors or industries—whether in large organizations within industries like construction, heavy manufacturing, or aerospace, or by smaller businesses and startups looking to design on a smaller scale. CAD tools are also widely used by civil engineering students, general and sub-contractors, architects, small business owners, and freelancers working on projects or prototypes [1].

However, with the growing reliance on CAD tools, a critical question arises about their cybersecurity. Since CAD tools are among the first steps in the additive manufacturing (AM) process, it is important to assess how secure these tools are. According to [2] and [3],, the AM process can be divided into five key phases: CAD, Finite Element Analysis (FEA), slicing, printing, and assembly. While much research has been conducted on the security concerns in later stages of the AM process (such as slicing and printing), there is a significant gap in the literature when it comes to cybersecurity vulnerabilities in the CAD phase, particularly during the tool download and installation process. This study focuses on the initial phase evaluating the security of CAD tools. Each of these phases' present unique cybersecurity risks and understanding the vulnerabilities at the CAD phase is crucial for ensuring the overall security of the AM process. The risks associated with downloading and using CAD tools have not been fully explored, even though this phase is often the first line of attack. The problem lies in the insufficient research and lack of attention to cybersecurity risks, particularly during the CAD tool download and installation phase.

There are several common types of potential cybersecurity attacks that users may encounter during the usage of CAD tools. As per [4], these attack], these attacks can exploit vulnerabilities in both the software and the data it generates or handles. Some of the frequently encountered or potential threats when using CAD tools include: 1) Malware Insertion during CAD tool download; 2) Denial of Service (DoS) attacks when users brose related web pages, making them unavailable

to legitimate users; 3) Social Engineering attacks, where attackers trick users to reveal their credentials and designs; 4) Ransomware attacks targeting CAD tools, where users are locked out or their files are encrypted; 5) Vulnerabilities in the CAD tool itself, which may be exploited by attackers[5].

As mentioned in [6], open-source CAD tools open source CAD tools are often more vulnerable to cyber-attacks for several reasons which includes: 1) Visibility of source code in the public domain, allowing hackers to inspect, identify, and exploit vulnerabilities; 2) Inconsistent updates and maintenance, despite community support and upkeep; 3) Limited human resources for security management; 4) Fewer security standards compared to paid CAD tools. Despite this, the security concerns related to the initial CAD tool download phase remain under-researched, particularly in the context of open-source CAD tools. The process of a cyberattack typically follows a series of stages. It begins with reconnaissance, where attackers gather information about the target. This is followed by exploitation, where vulnerabilities are leveraged, such as phishing or malware deployment. After gaining access, the attackers may escalate privileges, steal data, or disrupt operations, while attempting to hide their presence [2].

Additionally, a recent study addressed in [3] and [7] indicates a significant increase of 63% in the number of downloads of open-source CAD tools in 2022, further highlighting the growing trend. However, this study focuses specifically on the initial phase of the CAD tool usage process—when users download the tool, as all CAD tools require a download. This paper aims to fill this gap by examining the vulnerabilities during the download and installation phase of CAD tools, which is crucial to securing the overall AM process. The following section describes the journey of a typical user who browses through the internet and visits the website of the CAD tool. It also depicts the journey of an attacker who at the same time could perform malicious activities on the internet.

2 Literature Review

In our study, we conducted a structured literature review by searching two academic databases, i.e., IEEE Xplore and Google Scholar. We used specific keywords such as CAD tools, security assessment, open-source CAD vulnerabilities, security in AM, etc., to identify relevant studies. Additionally, we applied inclusion and exclusion criteria to filter relevant papers based on their relevance to web-based security assessments of CAD tools. Upon reviewing selected papers and analyzing their abstracts, we drew inspiration from those that discuss and categorize various taxonomies of cyberattacks in the AM industry, ensuring a comprehensive understanding of the security landscape. In addition, as described above, in the interest of adhering to the page limit, not all references were included in the study**.** To understand the specific gaps in the most recent literature, it is imperative to comprehend the generic process of any user's navigation journey through typical CAD tools. In an objective to make this clear, section 2.1 outlines the overview of the user journey, section 2.2 delves into the specifics of user interaction with the CAD website, and section 2.3 details the attacker's interaction with the CAD website. Finally, section 2.4 provides a summary of the most relevant studies in the domain and identifies the specific research gaps that will be addressed in this study

2.1 Overview of the User Journey

Based on the understanding and inferences from the literature review [8], the following elements have been identified as relevant and suitable for our study (as shown in Figure 1). Where 1) user (blue icon) represents a legitimate user engaging with the website of a CAD application, and 2) attacker (black-hat icon) represents a malicious actor attempting to exploit web-based vulnerabilities for them.

2.2 User Interaction with the CAD website

First, the website exploration phase, where the user visits the CAD website, typically to explore, read, and download the CAD tool. At this phase, the website can be potentially exposed to countless potential web based cyber threats. Second, CAD tool download phase, where the user downloads the CAD application from the website. This step carries risks such as the possibility of downloading compromised or malicious software. Finally, desktop application phase, where the downloaded tool transitions into a desktop-based application. However, this phase is beyond the scope of our study and is not addressed in our analysis.

2.3 Attackers Interaction with the CAD website

The attacker can target two out of 3 phases in the user journey, i.e., the website exploration phase and the CAD tool download phase. These phases present opportunities for exploiting vulnerabilities. Figure 1 highlights several common attack vectors as mentioned by [9]: 1) DDoS (Distributed Denial-of-Service): overloads the website, making it inaccessible to legitimate users, 2) injection attacks: injects malicious code into the website or the download processes 3) phishing: tricks users into providing sensitive information or downloading harmful files, and 4) ransomware: encrypts or locks the user's data or application, demanding payment to restore access.



Figure 1: User-Attacker Interaction in CAD environment

2.4 Relevant Studies

Several existing studies emphasize the significance of cyber-attacks during different phases of the AM process. For example, in [10], [11], and [12], the authors have highlighted and discussed the taxonomy of potential cyber-attacks in the AM industry. In summary, the following five cyber-critical phases were identified: CAD, Finite Element Analysis (FEA), slicing, printing, testing, and assembly. These phases represent the primary attack vectors highlighted in the literature, providing a general overview of the current state of cybersecurity concerns in the AM industry. All three papers have identified and analyzed the attack phases; however, none of these studies specifically focus on the first attack phase, which is the CAD phase. To elaborate, study [4] examines cyber-attacks in the FEA phase, slicing phase, and printing phase, while Papers [5] and [6] primarily focus on cyber-attacks in the slicing and printing phases.

In addition to the mentioned sources, [13] and [14] emphasize the cybersecurity challenges within the AM process, particularly as it pertains to the design and simulation stages. These stages, which rely heavily on CAD tools, are crucial for ensuring the accuracy and reliability of the final product. Since CAD tools are the first point of contact in the AM process, securing them is essential for protecting the integrity of the entire workflow.

Moreover, as the AM industry continues to embrace digitalization and the integration of advanced technologies such as Artificial Intelligence (AI) and the Internet of Things (IoT), the attack surface for cyber threats expands, especially during the early stages of the design process. The evolution of these technologies brings new challenges in securing CAD tools, which, if compromised, can lead to broader risks across the entire manufacturing chain.

This paper, while drawing on these studies, emphasizes the importance of understanding and mitigating the cybersecurity risks during the CAD phase. Given the increasing reliance on digital tools in the AM industry, ensuring the security of CAD tools is paramount to maintaining the confidentiality, integrity, and availability of the data and models that are central to the manufacturing process.

3 Methodology

Figure 2 shows all the steps in the proposed methodology. This methodology outlines a nine-step process along with two decision boxes for conducting a web-based security assessment of CAD tool.

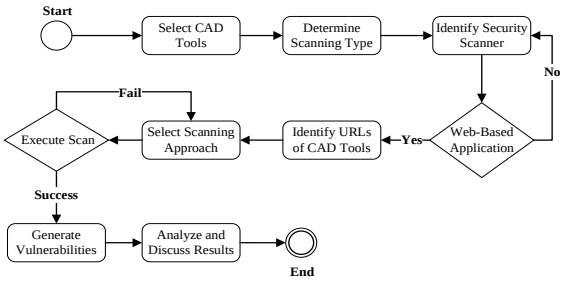


Figure 2: Overview of the proposed methodology

3.1 Select CAD Tools

The selection of appropriate CAD tools was the first step in the methodology. According to [15] and [16], CAD systems are classified based on two main criteria: 1) Licensing and 2) Type of computer application. Licensing of the CAD systems refer to their categorization based on commercial and non-commercial use. This study focuses on non-commercial use, which includes freeware and open-source tools. In contrast, commercial CAD systems fall under the closed-source category. These closed-source systems are high-performance platforms equipped with advanced features. According to the same study, the largest providers of commercial CAD systems are Autodesk, CATIA, and solid edge. The second classification is based on the type of application, which can be further divided into desktop-based, web-based, and mobile-based CAD tools. For our study, we have chosen freeware tools, and the specific type of CAD application selected is web-based**.** Therefore, selection of the right CAD tool is crucial, and it is essential to clearly identify the specific type of CAD system that this study intends to evaluate and hence of relevance in the further steps of the methodology.

3.2 Determine Scanning Type

It is crucial to understand the three main types of testing that should be conducted when an application goes live. Whether the application is intended for commercial or public use, ensuring its protection against cyberattacks is a top priority. According to [11] and [18], there are three most common types of testing namely static code analysis testing (SAST), dynamic application security testing (DAST), and interactive application security testing (IAST). For our study, we have chosen to

use Dynamic Application Security Testing (DAST). This approach is preferred as it allows for testing the application while it is running, simulating real-world cyberattacks and identifying vulnerabilities that may not be apparent during the development phase. A basic definition of each is provided in the Table 1 for easier comparison.

Table 1: Three most common scanning types

SAST	DAST	IAST
White box-based testing/ Source code-based testing	Black box testing. Conducted while the application is running	Code is analysed for security vulnerabilities while an application is running

3.3 Identify Security Scanner

Security scanners exist in various forms, each designed to identify several types of vulnerabilities, threats, and weaknesses within an environment. There are several other types of security scanners, including: 1) Network Security Scanners, in which they scan for vulnerabilities related to open ports and network configurations.2) Vulnerability Scanners, in which vulnerabilities are identified in systems or networks by comparing them against databases of common weaknesses. 3) Web Application Security Scanners: These specifically detect vulnerabilities within web applications, such as SQL injection or cross-site scripting (XSS). 4) Static Application Security Testing (SAST): These scanners analyze source code or binaries to identify security flaws before the application is run. 5) Dynamic Application Security Testing (DAST) Scanners: These scanners find vulnerabilities by testing applications during runtime, simulating real-world attacks.6) Cloud Security Scanners: These scanners identify vulnerabilities in cloud environments, such as misconfigurations or security gaps in cloud-based services.

Given that the scope of this study is focused on scanning vulnerabilities in web applications, it is important to note that, as stated in [8], web application security scanners simulate the actions of a penetration tester to thoroughly inspect and assess the target application. Among the web application security scanners used in this study, we have selected OWASP ZAP (Zed Attack Proxy), a widely recognized and popular scanner within the security community. There are two primary reasons for choosing this tool which includes: 1) Community Support and Relevance: ZAP is developed, managed, and maintained by the Open Web Application Security Project (OWASP), a prominent organization that also publishes the OWASP Top 10 list of the most critical security risks faced by web applications. According to [9], it is important to note that the OWASP Top 10 serves as a baseline for information security professionals when assessing the security posture of any web application running on the internet. 2) Customizability and Features: The ZAP scanner offers a wide range of customizable features essential for conducting a comprehensive security assessment. These include various scanning techniques, flexible reporting options, and automation capabilities, making it an ideal choice for thorough security evaluations.

3.4 Web based application

As the focus of this study is on scanning web applications, it is imperative to utilize tools and methodologies specifically designed for this purpose. The current step acts as a decision box and supports as discussed in the following sections. However, if the application under examination is not web-based, we would need to reassess our approach and move to Step 3.3, where we would identify an appropriate security scanner for non-web applications

3.5 Identify URLs of CAD Tools

Once the target application is identified, a crucial step is to determine the URLs associated with the application. This is a key step highlighted in [9], where identifying the URLs (i.e., web pages) to be scanned is considered one of the foundational tasks in conducting a security assessment. Furthermore, the paper describes the general architecture of web application security scanners, noting that the URLs serve as the primary input for the testing process. The list of the main web pages of the five open-source CAD tools that were part of our study is as follows: FreeCAD (https://www.freecad.org/),OpenScad (https://openscad.org/),Blender(https://www.blender.org/), LibreCAD (https://solvespace.com/), and Solverspace (https://librecad.org/).

3.6 Select Scanning Approach

There are two primary scanning modes: Active scanning and Passive scanning, both of which are available in the web application security scanner, OWASP ZAP. According to [20], active scanning is considered the more aggressive form of scanning, as it involves actively attacking the application to identify vulnerabilities. During this process, there is a high likelihood that the scan will be visible to the target application team, who may monitor the traffic and observe the different attacks being performed. Typically, active scanning is conducted over private web connections and can be carried out internally. In contrast, passive scanning does not modify the traffic between the scanner and the target application, nor does it execute any

malicious attacks. This mode simply observes the application's behavior without interacting with or altering the system. It is crucial to note that explicit authorization from the application owner is required to conduct active scanning, as it can potentially harm the production environment. For this study's purposes, the ZAP tool has been configured to use passive scanning, and throughout this study, only passive scanning has been performed.

3.7 Execute the Scan

Sections 3.7.1 and 3.7.2 outline the prerequisites for using the scanner, along with the appropriate environmental conditions required for its successful operation.

3.7.1 Pre-requisites for ZAP proxy 2.15.9

As mentioned in [9], there are 2 prerequisites required to install the tool: 1) Supported operating systems, i.e., Windows, Linux, and macOS. 2) Ensure that the Java Runtime Environment (JRE) is fully patched and maintained.

3.7.2 Environment setup

There are three main steps that include: 1) Cloning a Virtual Machine: The primary host machine to conduct this study was completed on Windows VM. 2) Download ZAP Proxy**:** Visit the OWASP ZAP website to download the latest version compatible with your operating system. 3) Install ZAP Proxy**:** Run the downloaded installer on the Windows VM and follow the installation prompts.

3.8 Generate Vulnerabilities

The active scanning technique performed on all five URLs was successful in identifying vulnerabilities across all five CAD tools examined. The results are displayed in Table 2. High (H), which represents critical vulnerabilities that require immediate attention; Medium (M), indicating moderate risks that should be addressed promptly; Low (L), referring to less severe issues that may not immediately impact security; and Informational (I), which are minor details that do not pose a direct threat. These levels are inherited from the scanner. Additionally, the ZAP Proxy scanner provides a breakdown of vulnerabilities based on their severity, using data from its own database. The vulnerabilities from the ZAP proxy are categorized into four levels of severity in decreasing order (as shown Table 2): High (H), Medium (M), Low (L), and Informational (I) which are inherited from the scanner.

It is important to note that OWASP ZAP utilizes its own vulnerability assessment database, meaning that the scanning algorithm and rules are defined within the ZAP scanner itself. The scanner identifies vulnerabilities based on predefined patterns. Similarly, the scanner references classifications from the Open Web Application Security Project (OWASP) and the Common Weakness Enumeration (CWE), which is maintained by MITRE.org. These are community-driven resources that catalog common software and hardware weaknesses. According to CWE.mitre.org, which is a not-for-profit organization that addresses challenges in cybersecurity, develops comprehensive cybersecurity frameworks, and provides resources to help organizations identify, defend, and respond to cyberattacks. A weakness is defined as a condition in software, firmware, hardware, or a service that, under certain conditions, could lead to vulnerabilities. These vulnerabilities were detected by the ZAP scanner, each of which is associated with a CWE ID, as explained in Section 4 and 4.1

Table 2: Breakdown of Vulnerabilities by Severity

Tool Name	H	M	L	I	Sub Total
Free Cad	1	5	2	5	13
OpenScad	0	6	4	5	15
Blender	0	2	3	5	10
LibreCad	0	3	1	3	10
SolverSpace	0	4	5	0	9

3.8.1 Risk Vs Confidence matrix

According to [9] and [10], the ZAP Proxy scanner further categorizes detected vulnerabilities using two parameters: Risk and Confidence. Risk can be defined as the potential consequences of data damage caused by a cyberattack. As noted by [21], risk is calculated in Equation 1.

$$\text{Risk} = \text{Threat} + \text{Vulnerability} \quad (1)$$

The level of Confidence is determined by the maliciousness of the Indicator of Compromise (IOC), which may include elements such as Internet Protocol (IP) addresses, email addresses, domains, URLs. The ZAP Proxy scanner provides both risk and confidence levels when listing vulnerabilities [11]. Based on this data, a heat map (as shown in Figure 4) is generated for the five CAD tools analyzed in this study. In the heat map, Confidence is represented along the X-axis, and Risk is represented along the Y-axis. The risk levels are categorized as Informational, Low, Medium, and High, in ascending order of severity. The X-axis is divided into three categories corresponding to low, medium, and high confidence levels, arranged in increasing order of confidence.

3.9 Analyze and Discuss Results

The scan was conducted independently for each of the five CAD tools. From these independent scans, a total

of 54 vulnerabilities were identified, with only 25 vulnerabilities being unique. The list of unique vulnerabilities are as follows: V1) Cross-site scripting, V2) Header not set, V3) Hidden file found, V4) Missing anti-clickjacking header, V5).htmaccess information leak, V6) Absence of anti-csrf tokens, V7) Strict-transport security header not set, V8) X-content options header missing, V9) Modern web application, V10) Re-examine cache control directives, V11) Security policy header not set, V12) User agent fuzzer, V13) User controllable HTML element (XSS), V14) Wildcard Directive, V15) style-src unsafe-inline, V16) Cross-domain misconfiguration, V17) Vulnerable JS library, V18) Cross-domain JavaScript source file inclusion, V19) Timestamp disclosure – Unix, 20) CSP: Header & Meta, V21) Information disclosure – suspicious comments, V22) Retrieved from Cache, 23) Charset Mismatch, V24) Application error disclosure, V25) Information disclosure – debug error messages. Figure 3 below displays a 5-set Venn diagram illustrating the unique vulnerabilities identified across the five CAD tools. Notably, only one vulnerability (V2) was found in all five CAD tools, while two vulnerabilities (V4 and V7) were found in four tools, and four vulnerabilities (V9, V10, V21, and V19) were found in three tools. This highlights the significant overlap among these CAD tools. In contrast, the study also identified distinct vulnerabilities unique to certain CAD tools: 8 distinct vulnerabilities for Free Cad, 4 distinct vulnerabilities for Solverspace, 3 distinct vulnerabilities for OpenScad, 1 distinct vulnerability for Blender, and none for Libre CAD. According to Figure 3, the most frequent vulnerabilities that were repeated are shown in Table 3.

The analysis of the results was conducted for the following scenarios: 1) The detected vulnerability with the highest severity, and 2) The significance of the most frequent vulnerabilities.

3.9.1 High Risk Vulnerabilities

The Free CAD tool was identified with a high-severity vulnerability classified as CWE-79, which pertains to Improper Neutralization of Input During Web Page Generation (Cross-Site Scripting). This particular vulnerability was detected exclusively within this tool. It typically arises when untrusted data is allowed to enter a web application, often via a web request. According to [12] and [13], during the process of page generation, the victim's web browser executes the injected malicious code within the context of the web server's domain, thus violating the web browser's same-origin policy. This policy is designed to prevent scripts from one domain from accessing or modifying data in another domain. According to [22], exploitation of this vulnerability could enable an attacker to perform a range of malicious activities, including, but not limited to, the theft of sensitive application data such as session cookies and other confidential information.2) Identification of Input Vectors: Understanding all potential areas where untrusted input may enter the application, such as parameters, arguments, cookies, network data, environment variables, reverse DNS lookups, query results, request headers, URL components, email data, files, filenames, databases, and external systems providing data to the application. This is a critical aspect of overall architectural design.3) Server-Side Validation: Ensuring that all input validation checks are also duplicated on the server side, providing an additional layer of protection.4) Use of Security Appliances: Implementing a Layer 7 firewall and a web application firewall (WAF) capable of detecting and mitigating attacks targeting this vulnerability [15] [16].

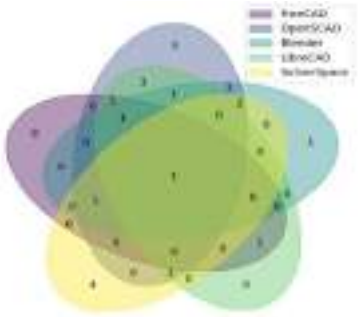


Figure 3: Vulnerability across CAD tools



Figure 4: HeatMap for FREECAD tool

Table 3: Most Frequent Vulnerabilities

Name of Vulnerability	Frequency
Header not set	5
Missing anti-clickjacking header	4
Strict transport security header not set	4
Re-examine cache control directives	3
Information disclosure – suspicious comments	3
Modern web application	3
Timestamp disclosure – Unix	3

3.9.2 Most Frequent Vulnerabilities

According to Table 3, the "Header not set" vulnerability (CWE 693) is present in all 5 CAD tools. Additionally, the "Missing anti-clickjacking header" (CWE 1021) is present in 4 CAD tools, and the "Strict Transport Security header not set" (CWE 319) is also present in 4 out of 5 CAD tools. This vulnerability represents a subset of the broader Content Security Policy (CSP) framework, which functions as an additional layer of security designed to detect and mitigate critical threats, particularly Cross-Site Scripting (XSS) and data injection attacks. By injecting CSP headers from the server, the browser becomes equipped to protect the user from malicious dynamic content that may be loaded onto the page they are currently visiting [9]. It is important to recognize that this vulnerability is also a component of XSS, and we can conclude that all five of these CAD tools are highly susceptible to Cross-Site Scripting attacks. The vulnerability related to the absence of an anti-clickjacking header has been detected across all CAD tools, except Blender [17]. Clickjacking is a type of cyberattack wherein users are deceptively manipulated into clicking on elements they did not intend to interact with. This is typically achieved by superimposing a transparent layer over legitimate content, causing the user's click to activate hidden or malicious elements controlled by the attacker [22]. One effective countermeasure against such attacks is the implementation of an anti-clickjacking header in the HTTP response. The HTTP Strict Transport Security (HSTS) vulnerability is another significant issue identified in all CAD tools, except Libre CAD. In these tools, sensitive or security-critical data is transmitted in cleartext over communication channels, making it susceptible to interception by unauthorized actors. According to [18], this compromise violates the *Confidentiality* component of the CIA triad, as anyone gaining access to the network channel can read the transmitted information. To mitigate this type of attack, it is essential to encrypt the data using robust protocols prior to transmission, ensuring SSL/TLS encryption is implemented throughout the entire session, from login to logout.

4 Conclusions and Future work

This study contributes to the web-based security evaluation of five widely used CAD tools (i.e., Free Cad, OpenScad, Blender, Libre Cad, and Solverspace) and identifies 25 unique vulnerabilities (numbered from V1 to V25) with the help of ZAP proxy scanner. The proposed methodology however is generic and applicable to examine other web-based open-source CAD tools as well. The vulnerabilities observed through the analysis in this study range in severity, from highly risky to informational, and likelihood of occurrence, from the most probable to the least probable. The most critical vulnerability, V1, was detected in the Free Cad tool and is considered a high-risk issue. That is CWE 79, which pertains to Improper Neutralization of Input During Web Page Generation, Cross-Site Scripting, common in web-based attacks.

It is crucial to address these vulnerabilities, as they compromise the tools' adherence to fundamental security principles and expose end-users to a wide range of threats, from data theft to sophisticated attacks such as clickjacking and session hijacking. These findings are particularly relevant for AECO industry professionals, who increasingly rely on CAD tools for design and collaboration, as addressing these vulnerabilities ensures the security of sensitive project data and reduces risks to increasingly prevalent digital project workflows.
Additionally, future work also seeks to draw broader, more applicable inferences about the vulnerabilities and corresponding countermeasures on specifically web-based open-source CAD tools in general, which is the focus of this study. As part of this effort, AI-driven security tools will be explored to enhance the detection, classification, and mitigation of vulnerabilities in these CAD applications.

One of the key limitations of this study is that the results were generated using the ZAP proxy scanner. Potential variations can be observed if other scanners are utilized on these CAD tools, the proposed methodology and the results observed still largely will remain applicable. That is, the authors cannot guarantee that similar results would be obtained under different conditions, such as: 1) using a different web scanner, 2) applying a different scanning method, or 3) conducting the study at a different time. Vulnerabilities may be addressed, or new ones may emerge, potentially altering the results even if the same scanner is used due to the dynamic nature of website traffic [23]. As part of an ongoing effort, the authors are working to address the identified limitations of this study. Future work aims to generate more comprehensive results that provide deeper insights into the selected CAD tools. Additionally, future work also seeks to draw broader, more applicable inferences about the vulnerabilities and corresponding counter measures on specifically web based open-source CAD tools in general which is the focus of this study.

5 References

[1] Sun, Y., Liu, K., Li, Y., & Zhang, D. (2022). Towards an Open-Source Industry CAD: A Review of System Development Methods. *Tehnički vjesnik*, *29*(6), 2127-2136.
[2] Belikovetsky, S., Yampolskiy, M., Toh,
[3] J., Gatlin, J., & Elovici, Y. (2017). dr0wned–

{Cyber-Physical} attack with additive manufacturing. In *11th USENIX workshop on offensive technologies (WOOT 17)*.

[4] Junk, S., & Kuen, C. (2016). Review of open source and freeware CAD systems for use with 3D-printing. *Procedia CIRP*, *50*, f30-435. Review of Open Source and Freeware CAD Systems for Use with 3D-Printing

[5] L. Wang, R. L. Mosher and P. Duett, "Cyberattacks and Cybersecurity in Additive Manufacturing," SoutheastCon 2024, Atlanta, GA, USA, 2024, pp. 1040-1045, doi: 10.1109/SoutheastCon52093.2024.10500033.

[6] M. Ekstedt, P. Johnson, R. Lagerström, D. Gorton, J. Nydrén and K. Shahzad, "Securi CAD by Foreseeti: A CAD Tool for Enterprise Cyber Security Management," 2015 IEEE 19th International Enterprise Distributed Object Computing Workshop, Adelaide, SA, Australia, 2015, pp. 152-155, doi: 10.1109/EDOCW.2015.40.

[7] J. Straub, "Automated testing and quality assurance of 3D printing/3D printed hardware: Assessment for quality assurance and cybersecurity purposes," *2016 IEEE AUTOTESTCON*, Anaheim, CA, USA, 2016, pp. 1-5, doi: 10.1109/AUTEST.2016.7589613.

[8] Moritani, H., Yogo, S., Morita, T., Kojima, M., Watanabe, K., Sun, J., ... & Hashimoto, Y. (2014, October). Development of cad for zone dividing of process control networks to improve cyber security. In *2014 14th International Conference on Control, Automation and Systems (ICCAS 2014)* (pp. 1311-1316). IEEE.

[9] Muhammad Syarif, H., Hoga, S., & Aedah, A. R. (2019). Web vulnerability assessment and maturity model analysis on Indonesia higher education. *Procedia Computer Science*, *16*(1), 1165-1172.

[10] Razzaq, A., Anwar, Z., Ahmad, H. F., Latif, K., & Munir, F. (2014). Ontology for attack detection: An intelligent approach to web application security. *computers & security*, *45*, 124-146.

[11] Sriram, G., & Rambabu, S. (2024). IMPORTANCE OF CYBERSECURITY IN ADDITIVE MANUFACTURING IN THE ERA OF INDUSTRY 4.0. *Int. J. Manuf. Mater. Mech.(IJMMME)*, *2940*, 6442.

[12] Ekstedt, M., Johnson, P., Lagerström, R., Gorton, D., Nydrén, J., & Shahzad, K. (2015, September). Securi cad by foreseeti: A cad tool for enterprise cyber security management. In *2015 IEEE 19th international enterprise distributed object computing workshop* (pp. 152-155). IEEE.

[13] Chen, F. (2019). *Design-Based Security Strategies for The Additive Manufacturing Cyber-Physical Systems* (Doctoral dissertation, New York University Tandon School of Engineering).

[14] Castro, H., Putnik, G., Castro, A., & Fontana, R. D. B. (2019). Open design initiatives: an evaluation of CAD Open Source Software. *Procedia CIRP*, *84*, 1116-1119. Open Design initiatives: an evaluation of CAD Open Source Software - ScienceDirect

[15] Chang, H., Kim, K. K., & Kim, Y. (2008). The development of security system for sharing CAD drawings in U-environment. *Computing and Informatics*, *27*(5), 731-741.

[16] Dencheva, L. (2022). *Comparative analysis of Static application security testing (SAST) and Dynamic application security testing (DAST) by using open-source web application penetration testing tools* (Doctoral dissertation, Dublin, National College of Ireland). lyubkadencheva.pdf

[17] Seng, L. K., Ithnin, N., & Said, S. Z. M. (2018). The approaches to quantify web application security scanners quality: a review. *International Journal of Advanced Computer Research*, *8*(38), 285-312.

[18] Ramos Flores E. (2023). ZAP Proxy and OWASP Top10. Computer Science; PUPR_CEAH_SJU_WI23_MCS_Eduard_Ramos Flores_Article.pdf

[19] Achuthan, K., SudhaRavi, S., Kumar, R., & Raman, R. (2014, May). Security vulnerabilities in open-source projects: an India perspective. In *2014 2nd International Conference on Information and Communication Technology (ICoICT)* (pp. 18-23). IEEE.

[20] Plate, H., Ponta, S. E., & Sabetta, A. (2015, September). Impact assessment for vulnerabilities in open-source software libraries. In *2015 IEEE International Conference on Software Maintenance and Evolution (ICSME)* (pp. 411-420). IEEE.

[21] Zhao, Y., Liang, R., Chen, X., & Zou, J(2021).Evaluation indicators for open-source software: a review. *Cybersecurity*, *4*, 1-24.

[22] Walkowski, M., Oko, J., & Sujecki, S. (2021). Vulnerability management models using a common vulnerability scoring system. *Applied Sciences*, *11*(18), 8735.

Huynh, T., & Miller, J. (2010). An empirical investigation into open-source web applications' implementation vulnerabilities. *Empirical Software Engineering*, *15*, 556-576.